

*Edited by Daniele Ielmini
and Rainer Waser*

Resistive Switching

Related Titles

Waser, R. (ed.)

Nanoelectronics and Information Technology

Advanced Electronic Materials and Novel Devices

Third Edition

2012

Print ISBN: 978-3-527-40927-3

Raveau, B., Seikh, M.

Cobalt Oxides

From Crystal Chemistry to Physics

2012

Print ISBN: 978-3-527-33147-5

Feringa, B.L., Browne, W.R. (eds.)

Molecular Switches

Second Edition

2011

Print ISBN: 978-3-527-31365-5

Piramanayagam, S.N., Chong, T.C. (eds.)

Developments in Data Storage

Materials Perspective

2011

Print ISBN: 978-0-470-50100-9

Waser, R. (ed.)

Nanotechnology

Volume 4: Information Technology II

2008

Print ISBN: 978-3-527-31737-0

Waser, R. (ed.)

Nanotechnology

Volume 3: Information Technology I

2008

Print ISBN: 978-3-527-31738-7

Edited by Daniele Ielmini and Rainer Waser

Resistive Switching

From Fundamentals of Nanoionic Redox Processes to
Memristive Device Applications

WILEY-VCH
Verlag GmbH & Co. KGaA

Editors**Prof. Daniele Ielmini**

Politecnico di Milano
Dipartimento di Elettronica, Informazione
e Bioingegneria
Piazza L. da Vinci 32
20133 Milano
Italy

Prof. Rainer Waser

RWTH Aachen
Institut für Werkstoffe der Elektrotechnik
Sommerfeldstr. 24
52074 Aachen
Germany

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

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

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2016 Wiley-VCH Verlag GmbH & Co. KGaA, Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photostriking, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Print ISBN: 978-3-527-33417-9

ePDF ISBN: 978-3-527-68094-8

ePub ISBN: 978-3-527-68093-1

Mobi ISBN: 978-3-527-68090-0

oBook ISBN: 978-3-527-68087-0

Cover Design Formgeber, Mannheim,
Germany

Typesetting SPi Global, Chennai, India
Printing and Binding

Printed on acid-free paper

Contents

Preface *XIX*

List of Contributors *XXI*

1	Introduction to Nanoionic Elements for Information Technology	1
	<i>Rainer Waser, Daniele Ielmini, Hiro Akinaga, Hisashi Shima, H.-S. Philip Wong, Joshua J. Yang, and Simon Yu</i>	
1.1	Concept of Two-Terminal Memristive Elements	1
1.1.1	Classifications Based on Behavior, Mechanisms, and Operation Modes	1
1.1.2	Scope of the Book	6
1.1.3	History	9
1.2	Memory Applications	12
1.2.1	Performance Requirements and Parameter Windows	12
1.2.2	Device Isolation in Crossbar Arrays	16
1.2.3	3-D Technology	19
1.2.4	Memory Hierarchy	20
1.3	Logic Circuits	21
1.4	Prospects and Challenges	24
	Acknowledgments	25
	References	25
2	ReRAM Cells in the Framework of Two-Terminal Devices	31
	<i>E. Linn, M. Di Ventra, and Y. V. Pershin</i>	
2.1	Introduction	31
2.2	Two-Terminal Device Models	32
2.2.1	Lumped Elements	32
2.2.2	Ideal Circuit Element Approach	32
2.2.3	Dynamical Systems Approach	33
2.2.3.1	Memristive Systems	33
2.2.3.2	Memristor	34
2.2.4	Significance of the Initial Memristor and Memristive System Definitions in the Light of Physics	34
2.2.4.1	Limitations of Ideal Memristor Models	35
2.2.5	Memristive, Memcapacitive, and Meminductive Systems	35

2.2.6	ReRAM: Combination of Elements, Combination of Memory Features, and Consideration of Inherent Battery Effects	36
2.3	Fundamental Description of Electronic Devices with Memory	38
2.4	Device Engineer's View on ReRAM Devices as Two-Terminal Elements	40
2.4.1	Modeling of Electrochemical Metallization (ECM) Devices	41
2.4.2	Modeling of Valence Change Mechanism (VCM) Devices	43
2.5	Conclusions	46
	Acknowledgment	47
	References	47
3	Atomic and Electronic Structure of Oxides	49
	<i>Tobias Zacherle, Peter C. Schmidt, and Manfred Martin</i>	
3.1	Introduction	49
3.2	Crystal Structures	50
3.3	Electronic Structure	54
3.3.1	From Free Atoms to the Solid State	55
3.3.2	Electrons in Crystals	58
3.3.2.1	Free Electron Model (Sommerfeld Model)	58
3.3.2.2	Band Structure Model	60
3.3.2.3	Density of States (DOS) and Partial DOS	62
3.3.2.4	Crystal Field Splitting	64
3.3.2.5	Exchange and Correlation	65
3.3.2.6	Computational Details	66
3.4	Material Classes and Characterization of the Electronic States	67
3.4.1	Metals	67
3.4.2	Semiconductors	68
3.4.3	Insulators	71
3.4.4	Point Defect States	72
3.4.5	Surface States	73
3.4.6	Amorphous States	75
3.5	Electronic Structure of Selected Oxides	76
3.5.1	Nontransition Metal Oxides	76
3.5.1.1	Al_2O_3	76
3.5.1.2	SrO	77
3.5.1.3	ZnO	77
3.5.2	Titanates	79
3.5.2.1	TiO	79
3.5.2.2	Ti_2O_3	79
3.5.2.3	TiO_2	81
3.5.2.4	SrTiO_3	82
3.5.3	Magnetic Insulators	82
3.5.3.1	NiO	84
3.5.3.2	MnO	85
3.5.4	M^{VB} Metal Oxides	86

3.5.4.1	Metal–Insulator Transitions: NbO ₂ , VO ₂ , and V ₂ O ₃	86
3.5.4.2	Tantalum Oxides TaO _x	87
3.6	Ellingham Diagram for Binary Oxides	90
	Acknowledgments	91
	References	91
4	Defect Structure of Metal Oxides	95
	<i>Giuliano Gregori</i>	
4.1	Definition of Defects	95
4.1.1	Zero-Dimensional Defects	95
4.1.2	One-Dimensional Defects	95
4.1.3	Two-Dimensional Defects	97
4.1.4	Three-Dimensional Defects	97
4.2	General Considerations on the Equilibrium Thermodynamics of Point Defects	98
4.3	Definition of Point Defects	99
4.3.1	Intrinsic Defects	99
4.3.1.1	Frenkel Defects	99
4.3.1.2	Anti-Frenkel Defects	99
4.3.1.3	Schottky Defects	100
4.3.1.4	Anti-Schottky Defects	100
4.3.1.5	Electron Band–Band Transfer	100
4.3.2	Extrinsic Defects	100
4.3.2.1	Reactions with the Environment	100
4.3.2.2	The Brouwer Diagram	101
4.3.2.3	Impurities and Dopants	102
4.4	Space-Charge Effects	103
4.4.1	Mott–Schottky Situation	104
4.4.2	Gouy–Chapman Situation	105
4.5	Case Studies	106
4.5.1	Titanium Oxide (Rutile)	106
4.5.1.1	Nominally Pure TiO ₂	107
4.5.1.2	Acceptor-Doped TiO ₂	108
4.5.1.3	Donor-Doped TiO ₂	108
4.5.1.4	The Role of Dislocations	109
4.5.2	Strontium Titanate	110
4.5.2.1	Acceptor-Doped SrTiO ₃	110
4.5.2.2	Donor-Doped SrTiO ₃	111
4.5.2.3	Grain Boundaries in SrTiO ₃	111
4.5.3	Zirconium and Hafnium Oxide	113
4.5.3.1	Zirconium Oxide	113
4.5.3.2	The Role of Grain Boundaries and Dislocations	115
4.5.3.3	Hafnium Oxide	116
4.5.4	Aluminum Oxide	116
4.5.4.1	Acceptor-Doped Alumina	117

4.5.4.2	Donor-Doped Alumina	118
4.5.5	Tantalum Oxide	119
	References	121
5	Ion Transport in Metal Oxides	125
	<i>Roger A. De Souza</i>	
5.1	Introduction	125
5.2	Macroscopic Definition	126
5.2.1	Two Solutions of the Diffusion Equation	127
5.2.2	Dependence of the Diffusion Coefficient on Characteristic Thermodynamic Parameters	128
5.3	Microscopic Definition	129
5.3.1	Mechanisms of Diffusion	130
5.3.2	Diffusion Coefficients of Defects and Ions	131
5.3.3	The Activation Barrier for Migration	132
5.4	Types of Diffusion Experiments	134
5.4.1	Chemical Diffusion	135
5.4.2	Tracer Diffusion	137
5.4.3	Conductivity	139
5.5	Mass Transport along and across Extended Defects	141
5.5.1	Accelerated Transport <i>along</i> Extended Defects	143
5.5.2	Hindered Transport <i>across</i> Extended Defects	145
5.6	Case Studies	145
5.6.1	Strontium Titanate	147
5.6.2	Yttria-Stabilized Zirconia (YSZ)	150
5.6.3	Alumina	153
5.6.4	Tantalum Pentoxide	155
	Acknowledgments	156
	References	157
6	Electrical Transport in Transition Metal Oxides	165
	<i>Franklin J. Wong and Shriram Ramanathan</i>	
6.1	Overview	165
6.2	Structure of Transition Metal Oxides	166
6.2.1	Crystal Structures of Oxides	166
6.2.2	Bonding and Electronic Structure	167
6.3	Models of Electrical Transport	168
6.3.1	Band Transport of Carriers	168
6.3.2	Electronic Bandwidth	169
6.3.3	Small Polaron Formation	169
6.3.4	Small Polaron Transport	171
6.3.5	Thermopower (Seebeck Coefficient)	172
6.3.6	Hopping Transport via Defect States	172
6.3.7	Bad Metallic Behavior	174
6.4	Band Insulators	175

6.4.1	SnO_2 : 3d ¹⁰ System	175
6.4.2	TiO_2 : 3d ⁰ System	176
6.5	Half-Filled Mott Insulators	177
6.5.1	Correlations and the Hubbard U	177
6.5.2	MnO : 3d ⁵ System	179
6.5.3	NiO : 3d ⁸ System	179
6.5.4	$\alpha\text{-Fe}_2\text{O}_3$: 3d ⁵ System	182
6.5.5	Summary	184
6.6	Temperature-Induced Metal–Insulator Transitions in Oxides	184
6.6.1	Orbitals and Metal–Insulator Transitions	184
6.6.2	VO_2 : 3d ¹ System	186
6.6.3	Ti_2O_3 : 3d ¹ System	187
6.6.4	V_2O_3 : 3d ² System	189
6.6.5	Fe_3O_4 : Mixed-Valent System	190
6.6.6	Limitations	191
6.6.7	Summary	192
	References	193
7	Quantum Point Contact Conduction	197
	<i>Jan van Ruitenbeek, Monica Morales Masis, and Enrique Miranda</i>	
7.1	Introduction	197
7.2	Conductance Quantization in Metallic Nanowires	197
7.3	Conductance Quantization in Electrochemical Metallization Cells	204
7.3.1	Current–Voltage Characteristics and Definition of Initial Device Resistance	206
7.3.2	Stepwise Conductance Changes in Metallic Filaments	207
7.4	Filamentary Conduction and Quantization Effects in Binary Oxides	210
7.5	Conclusion and Outlook	218
	References	218
8	Dielectric Breakdown Processes	225
	<i>Jordi Suñé, Nagarajan RagHAVAN, and K. L. PEY</i>	
8.1	Introduction	225
8.2	Basics of Dielectric Breakdown	226
8.3	Physics of Defect Generation	231
8.3.1	Thermochemical Model of Defect Generation	232
8.3.2	Anode Hydrogen Release Model of Defect Generation	233
8.4	Breakdown and Oxide Failure Statistics	235
8.5	Implications of Breakdown Statistics for ReRAM	237
8.6	Chemistry of the Breakdown Path and Inference on Filament Formation	241
8.7	Summary and Conclusions	246
	References	247

9	Physics and Chemistry of Nanoionic Cells	253
	<i>Ilia Valov and Rainer Waser</i>	
9.1	Introduction	253
9.2	Basic Thermodynamics and Heterogeneous Equilibria	254
9.3	Phase Boundaries and Boundary Layers	258
9.3.1	Driving Force for the Formation of Space-Charge Layers	258
9.3.2	Enrichment and Weak Depletion Layers	260
9.3.3	Strong Depletion Layers	261
9.3.4	Nanosize Effects on Space-Charge Regions	263
9.3.5	Nanosize Effects due to Surface Curvature	265
9.3.6	Formation of New Phases at Phase Boundaries	265
9.4	Nucleation and Growth	266
9.4.1	Macroscopic View	266
9.4.2	Atomistic Theory	267
9.5	Electromotive Force	269
9.5.1	Electrochemical Cells of Different Half Cells	269
9.5.2	Emf Caused by Surface Curvature Effects	270
9.5.3	Emf Caused by Concentration Differences	271
9.5.4	Diffusion Potentials	273
9.6	General Transport Processes and Chemical Reactions	274
9.7	Solid-State Reactions	275
9.8	Electrochemical (Electrode) Reactions	280
9.8.1	Charge-Transfer Process Limitations	280
9.8.2	Diffusion-Limited Electrochemical Processes	282
9.9	Stoichiometry Polarization	283
	Summary	285
	Acknowledgments	286
	References	286
10	Electroforming Processes in Metal Oxide Resistive-Switching Cells	289
	<i>Doo Seok Jeong, Byung Joon Choi, and Cheol Seong Hwang</i>	
10.1	Introduction	289
10.1.1	Forming Methods	290
10.1.2	Dependence of the Bipolar Switching Behavior on the Forming Conditions	291
10.1.3	Factors Influencing Forming Behavior	294
10.1.4	Forming in Bipolar and Unipolar Switching	295
10.1.5	Phenomenological Understanding of Forming	297
10.2	Forming Mechanisms	297
10.2.1	Early Suggested Forming Mechanisms	298
10.2.2	Conducting Filament Formation	298
10.2.3	Redox Reactions and Ion or Ionic Defect Migration during Forming	300
10.2.4	Point Defect Introduction	302

10.2.5	Point Defect Dynamics during the Forming Process	304
10.2.6	Microscopic Evidence for CF Formation during Forming	308
10.3	Technical Issues Related to Forming	310
10.3.1	Problems of Current Overshoot Forming	310
10.3.2	Nonuniform Forming Voltage Distribution	311
10.3.3	Forming-Free Resistive Switching	311
10.4	Summary and Outlook	312
	Acknowledgments	313
	References	313
11	Universal Switching Behavior	317
	<i>Daniele Ielmini and Stephan Menzel</i>	
11.1	General Properties of ReRAMs and Their Universal Behavior	317
11.2	Explaining the Universal Switching of ReRAM	320
11.3	Variable-Diameter Model	321
11.4	Variable-Gap Model	329
11.5	Coexistence of Variable-Gap/Variable-Diameter States	334
11.6	Summary	337
	Acknowledgment	337
	References	338
12	Quasistatic and Pulse Measuring Techniques	341
	<i>Antonio Torrezan, Gilberto Medeiros-Ribeiro, and Stephan Tiedke</i>	
12.1	Brief Introduction to Electronic Transport Testing of ReRAM	341
12.2	Quasistatic Measurement of Current–Voltage Characteristics	342
12.2.1	Dependence of Switching Parameters on Sweep Rate	345
12.3	Current Compliance and Overshoot Effects	346
12.4	Pulsed Measurements for the Study of Switching Dynamics	350
12.4.1	Experimental Setup and Results for Nanosecond Switching with Real-Time Monitoring of Device Dynamics	353
12.4.2	Experimental Setup and Results for Subnanosecond Switching with Real-Time Monitoring of Device Dynamics	354
12.5	Conclusions	358
	Acknowledgment	359
	References	359
13	Unipolar Resistive-Switching Mechanisms	363
	<i>Ludovic Goux and Sabina Spiga</i>	
13.1	Introduction to Unipolar Resistive Switching	363
13.2	Principle of Unipolar Switching	364
13.2.1	Basic Operation of Unipolar Memory Cells	364
13.2.2	Structure of Unipolar Memory Arrays	365
13.2.3	Experimental Evidences of Filamentary-Switching Mechanism	366
13.2.4	Typical Materials Used in Unipolar-Switching Cells	367
13.3	Unipolar-Switching Mechanisms in Model System Pt/NiO/Pt	368

13.3.1	Microscopic Origin of Switching in NiO Layers	368
13.3.1.1	Defect Chemistry	368
13.3.1.2	Microscopic Mechanism of the Switching	371
13.3.2	Physics-Based Electrical Models	372
13.3.2.1	Modeling of the Reset Switching	372
13.3.2.2	Modeling of the Set Switching	373
13.3.3	Model Implications on the Device Level	375
13.3.3.1	CF Size and R_{LRS} Scaling with I_C	375
13.3.3.2	I_{reset} Scaling with CF Size Scaling	376
13.3.3.3	Switching Speed	377
13.4	Influence of Oxide and Electrode Materials on Unipolar-Switching Mechanisms	379
13.4.1	Influence of the Oxide Material	380
13.4.1.1	The Specific Case of TiO_2	380
13.4.1.2	Influence of the Oxide Microstructure	380
13.4.1.3	Random Circuit Breaker Model	381
13.4.1.4	Coexistence of Bipolar and Unipolar Switching	382
13.4.1.5	Switching Variability and Endurance	383
13.4.2	Impacts and Roles of Electrodes	384
13.4.2.1	Anode-Mediated Reset Operation	384
13.4.2.2	Selection Criteria of Electrode Materials	385
13.5	Conclusion	386
	References	387
14	Modeling the VCM- and ECM-Type Switching Kinetics	395
	<i>Stephan Menzel and Ji-Hyun Hur</i>	
14.1	Introduction	395
14.2	Microscopic Switching Mechanism of VCM Cells	395
14.3	Microscopic Switching Mechanism of ECM Cells	397
14.4	Classification of Simulation Approaches	398
14.4.1	Ab initio and Molecular Dynamics Simulation Models	398
14.4.2	Kinetic Monte Carlo Simulation Models	398
14.4.3	Continuum Models	398
14.4.4	Compact Models	399
14.5	General Considerations of the Physical Origin of the Nonlinear Switching Kinetics	399
14.6	Modeling of VCM Cells	402
14.6.1	Ab initio Models and MD Models	402
14.6.1.1	HRS and LRS State Modeling	402
14.6.1.2	Electron Transfer	404
14.6.1.3	Phase Transformations and Nucleation	405
14.6.1.4	Calculation of Migration Barriers	406
14.6.2	Kinetic Monte Carlo Modeling	407
14.6.3	Continuum Modeling	410
14.6.4	Compact Modeling	417

14.7	Modeling of ECM Cells	422
14.7.1	Ab initio Models and MD Models	422
14.7.2	KMC Modeling	423
14.7.3	Continuum Modeling	426
14.7.4	Compact Modeling	428
14.8	Summary and Outlook	431
	Acknowledgment	433
	References	433
15	Valence Change Observed by Nanospectroscopy and Spectromicroscopy	437
	<i>Christian Lenser, Regina Dittmann, and John Paul Strachan</i>	
15.1	Introduction	437
15.2	Methods and Techniques	439
15.3	Interface Phenomena	442
15.3.1	Reactive Metal Layers on Insulating Oxides	442
15.3.2	Formation of a Blocking Layer on Conducting Oxides	443
15.3.3	Electrically Induced Redox Reactions at the Interface	444
15.4	Localized Redox Reactions in Transition Metal Oxides	446
15.4.1	Single Crystalline Model System – Doped SrTiO ₃	446
15.4.2	Localized Structural and Compositional Changes in TiO ₂	448
15.4.3	Compositional Changes in Ta ₂ O ₅ and HfO ₂	450
15.5	Conclusions	453
	Acknowledgment	453
	References	453
16	Interface-Type Switching	457
	<i>Akihito Sawa and Rene Meyer</i>	
16.1	Introduction	457
16.2	Metal/Conducting Oxide Interfaces: I – V Characteristics and Fundamentals	459
16.2.1	Schottky-Like Metal/Conducting Oxide Interfaces	459
16.2.2	Electronic Properties of Donor-Doped SrTiO ³	460
16.2.3	Electronic Properties of Mixed-Valent Manganites	461
16.3	Resistive Switching of Metal/Donor-Doped SrTiO ₃ Cells	463
16.4	Resistive Switching of p-Type PCMO Cells	465
16.5	Resistive Switching in the Presence of a Tunnel Barrier	469
16.5.1	Device Structure and Materials	469
16.5.2	Electrical Characteristics	470
16.5.3	Mechanism and Modeling	472
16.5.4	Passive Cross-Point Arrays	473
16.6	Ferroelectric Resistive Switching	475
16.6.1	Classification of Ferroelectric Resistive Switching	475
16.6.2	Ferroelectric Resistive-Switching Diode	475
16.7	Summary	479

	Acknowledgment	480
	References	480
17	Electrochemical Metallization Memories	483
	<i>Michael N. Kozicki, Maria Mitkova, and Ilia Valov</i>	
17.1	Introduction	483
17.2	Metal Ion Conductors	484
17.2.1	Materials	484
17.2.2	Ion Transport	490
17.3	Electrochemistry of CBRAM (ECM) Cells	492
17.3.1	Fundamental Processes	492
17.3.2	Filament Growth and Dissolution	495
17.3.3	Filament Morphology	500
17.4	Devices	503
17.4.1	Device Operation	503
17.4.2	Memory Arrays	506
17.5	Technological Challenges and Future Directions	508
	Acknowledgment	509
	References	510
18	Atomic Switches	515
	<i>Kazuya Terabe, Tohru Tsuruoka, Tsuyoshi Hasegawa, Alpana Nayak, Takeo Ohno, Tomonobu Nakayama, and Masakazu Aono</i>	
18.1	Introduction	515
18.1.1	Brief History of the Development of the Atomic Switch	516
18.1.2	Basic Working Principle of the Atomic Switch	517
18.2	Gap-Type Atomic Switches	519
18.2.1	Switching Time	519
18.2.2	Electrochemical Process	521
18.2.3	Cross-Bar Structure	523
18.2.4	Quantized Conductance	524
18.2.5	Logic-Gate Operation	526
18.2.6	Synaptic Behavior	527
18.2.7	Photo-Assisted Switch	528
18.3	Gapless-Type Atomic Switches	529
18.3.1	Sulfide-Based Switch	529
18.3.2	Oxide-Based Switch	530
18.3.3	Effect of Moisture	533
18.3.4	Switching Time	534
18.3.5	Quantized Conductance and Synaptic Behavior	535
18.3.6	Polymer-Based Switch	536
18.4	Three-Terminal Atomic Switches	537
18.4.1	Filament-Growth-Controlled Type	537
18.4.2	Nucleation-Controlled Type	539
18.5	Summary	541
	References	542

19	Scaling Limits of Nanoionic Devices	547
	<i>Victor Zhirnov and Gurtej Sandhu</i>	
19.1	Introduction	547
19.2	Basic Operations of ICT Devices	547
19.3	Minimal Nanoionic ICT	549
19.3.1	Switching Mechanisms and the Material Systems	549
19.3.2	Atomic Filament: Classical and Quantum Resistance	551
19.3.2.1	Classical Resistance	551
19.3.2.2	Quantum Resistance	552
19.3.2.3	Conductance in the Presence of Barriers	553
19.3.2.4	Barriers in Atomic Gaps: Nonrectangular Barrier	555
19.3.2.5	Transmission through Atomic Gaps	555
19.3.3	Interface Controlled Resistance (ICR)	556
19.3.3.1	Electrical Properties of Material Interfaces	557
19.3.3.2	Contact Resistance in a M–S (M–I) Structure	560
19.3.4	Stability of the Minimal Nanoionic State	563
19.4	Energetics of Nanoionic Devices	565
19.4.1	Switching Speed and Energy	565
19.4.2	Heat Dissipation and Transfer in a Minimal Nanoionic Device	567
19.5	Summary	569
	Acknowledgment	569
	Appendix A Physical Origin of the Barrier Potential	569
	References	571
20	Integration Technology and Cell Design	573
	<i>Fred Chen, Jun Y. Seok, and Cheol S. Hwang</i>	
20.1	Materials	573
20.1.1	Resistance Switching (RS) Materials	573
20.1.1.1	Insulating Oxides	573
20.1.1.2	Semiconducting Oxides	574
20.1.1.3	Electrolyte Chalcogenides	574
20.1.1.4	Phase-Change Materials	575
20.1.2	Electrode Materials, Including Reductants	575
20.2	Structures	576
20.2.1	Planar Stack	576
20.2.2	Sidewall-Conforming Stack	577
20.2.3	Lateral Structure	578
20.3	Integration Architectures	579
20.3.1	Transistor in Series with RRAM (1T1R)	579
20.3.2	Transistor in Parallel with RRAM (T R)	582
20.3.3	1S1R Stacked Crosspoint	583
20.3.3.1	The Selector Device	583
20.3.3.2	Sensing Margin	584
20.3.3.3	Write Margin	586
20.3.3.4	Cumulative Line Resistance	586

20.3.4	Through-Multilayer via Array	588
20.3.4.1	Through-Multilayer Vias	588
20.3.4.2	Staircase Connections	589
20.3.4.3	Horizontal Electrodes	589
20.3.4.4	Bathtub-Type Peripheral Connection	592
20.3.5	Array Area Efficiency	592
20.4	Conclusions	593
	Acknowledgment	594
	References	594
21	Reliability Aspects	597
	<i>Dirk J. Wouters, Yang-Yin Chen, Andrea Fantini, and Nagarajan Raghavan</i>	
21.1	Introduction	597
21.2	Endurance (Cyclability)	598
21.2.1	Endurance Summary of Bipolar Switching TMO RRAM	598
21.2.2	Balancing the Bipolar Switching for Better Endurance	599
21.2.3	Understanding of Endurance Degradation	600
21.3	Retention	601
21.3.1	Retention Summary of Bipolar TMO RRAM	601
21.3.2	Understanding of Retention Degradation in Bipolar TMO RRAM	603
21.3.3	Trade-Off between Retention/Endurance	604
21.4	Variability	605
21.4.1	Introduction	605
21.4.2	Experimental Aspects of Variability	605
21.4.2.1	Variability of Forming Operation	605
21.4.2.2	Intrinsic and Extrinsic Variability	606
21.4.3	Physical Aspects of Variability	607
21.4.3.1	Variability in Unipolar Devices	607
21.4.3.2	Variability in Bipolar Devices	607
21.5	Random Telegraph Noise (RTN)	609
21.5.1	Introduction	609
21.5.2	Charge Carrier Transport-Induced RTN	610
21.5.3	Oxygen Vacancy Transport-Induced RTN	611
21.5.3.1	Experimental Identification of Vacancy Perturbations	611
21.5.3.2	Vacancy-Induced RTN for Shallow to Moderate Reset	612
21.5.3.3	Vacancy-Induced RTN for Very Deep Reset	613
21.5.3.4	Bimodal Filament Configuration and Disturb Immunity	614
21.5.3.5	Role of Dielectric Microstructure on RTN Immunity	614
21.5.4	Summary of RTN Analysis Studies	615
21.6	Disturb	615
21.6.1	Phenomena	615
21.6.2	Understanding and Modeling	616
21.6.3	Anomalous Disturb Behavior	616
21.7	Conclusions and Outlook	617

	Acknowledgment	618
	References	618
22	Select Device Concepts for Crossbar Arrays	623
	<i>Geoffrey W. Burr, Rohit S. Shenoy, and Hyunsang Hwang</i>	
22.1	Introduction	623
22.2	Crossbar Array Considerations	624
22.2.1	Problems Associated with Large Subarrays	625
22.2.2	Considerations During NVM-Write	625
22.2.3	Considerations During NVM-Read	627
22.3	Target Specifications for Select Devices	627
22.4	Types of Select Devices	629
22.4.1	Si Based	629
22.4.2	Oxide Diodes	631
22.4.2.1	Oxide PN Junction	631
22.4.2.2	Metal-Oxide Schottky Barrier	632
22.4.3	Threshold Switch	633
22.4.3.1	Ovonic Threshold Switching	634
22.4.3.2	Metal–Insulator Transition (MIT)	636
22.4.3.3	Threshold Vacuum Switch	637
22.4.4	Oxide Tunnel Barrier	638
22.4.4.1	Single Layer Oxide-(Nitride-)Based Select Device (TiO_2 and SiN_x)	638
22.4.4.2	Multi-Layer Oxide-Based Select Device ($TaO_x/TiO_2/TaO_x$)	638
22.4.5	Mixed-Ionic-Electronic-Conduction (MIEC)	639
22.5	Self-Selected Resistive Memory	643
22.5.1	Complementary Resistive Switch	645
22.5.2	Hybrid ReRAM-Select Devices	647
22.5.3	Nonlinear ReRAM	649
22.6	Conclusion	651
	References	652
23	Bottom-Up Approaches for Resistive Switching Memories	661
	<i>Sabina Spiga, Takeshi Yanagida, and Tomoji Kawai</i>	
23.1	Introduction	661
23.2	Bottom-Up ReRAM Fabrication Methods	662
23.2.1	Vapor–Liquid–Solid Method	662
23.2.2	Template-Assisted Fabrication Methods of NWs	663
23.3	Resistive Switching in Single (All-Oxide) NW/Nanoisland ReRAM	664
23.3.1	Resistive Switching in Single NiO NWs and Nanoislands	665
23.3.2	Resistive Switching in Oxide NWs Alternative to NiO	669
23.3.3	Study of Switching Mechanisms in Oxide NW ReRAM	671
23.3.4	Resistive Switching in NW ReRAM with Active Electrodes: ECM Mechanisms	675
23.4	Resistive Switching in Axial Heterostructured NWs	678
23.5	Core–Shell NWs toward Crossbar Architectures	680

23.5.1	Crossbar Devices with Si(core)/a-Si(shell) NWs and Ag Electrodes	681
23.5.2	Crossbar Devices with Ni(core)/NiO(shell) NWs and Ni Electrodes	683
23.6	Emerging Bottom-Up Approaches and Applications	686
23.6.1	1D1R Nanopillar Array	686
23.6.2	Block-Copolymer Self-Assembly for Advanced ReRAM	687
23.7	Conclusions	688
	References	689
24	Switch Application in FPGA	695
	<i>Toshitsugu Sakamoto, S. Simon Wong, and Young Yang Liauw</i>	
24.1	Introduction	695
24.2	Monolithically 3D FPGA with BEOL Devices	696
24.3	Resistive Memory Replacing Configuration Memory	698
24.3.1	Architecture	698
24.4	Resistive Configuration Memory Cell	699
24.5	Resistive Configuration Memory Array	700
24.5.1	Prototype	702
24.5.2	Measurement Results	703
24.6	Complementary Atomic Switch Replacing Configuration Switch	706
24.6.1	Complementary Atomic Switch (CAS)	706
24.6.2	Cell Architecture with CAS	707
24.6.3	Demonstration of CAS-Based Programmable Logic	709
24.7	Energy Efficiency of Programmable Logic Accelerator	710
24.8	Conclusion and Outlook	712
	References	712
25	ReRAM-Based Neuromorphic Computing	715
	<i>Giacomo Indiveri, Eike Linn, and Stefano Ambrogio</i>	
25.1	Neuromorphic Systems: Past and Present Approaches	715
25.2	Neuromorphic Engineering	715
25.3	Neuromorphic Computing (The Present)	716
25.4	Neuromorphic ReRAM Approaches (The Future)	718
25.4.1	ReRAM-Based Neuromorphic Approaches	718
25.4.2	Nonvolatility and Volatility of Resistive States	721
25.4.3	Nonlinear Switching Kinetics	722
25.4.4	Multilevel Resistance Behavior	722
25.4.5	Capacitive Properties	725
25.4.6	Switching Statistics	725
25.5	Scaling in Neuromorphic ReRAM Architectures	728
25.6	Applications of Neuromorphic ReRAM Architectures	729
	References	731

Preface

In the area of nanoelectronics, resistive switching describes reversible phenomena of two-terminal circuit elements, which change their resistance upon electrical stimuli in a nonvolatile manner. The nonvolatility guarantees that the resistance change remains for a sufficiently long retention time after the electrical stimulus has been released. Empirically, the resistance change can be described by the variation of an internal state variable of the device in response to the stimulus. In other words, the resistance values are memorized by the elements that are, therefore, also called *memristive elements* or *devices*. The fundamental physical principles of the resistive switching and, hence, the nature of the internal state variable can be manifold, including ion migration and diffusion, chemical reactions, local temperature increase due to Joule heating, and many others. This book covers resistive switching based on nanoionics redox phenomena, that is, changes of the internal state of a device by ionic motion over nanoscale dimensions affecting the device resistance. Devices where this type of resistive switching can be observed go by the name of *redox-based resistive switching random access memory*, or simply *resistive switching random access memory* (ReRAM). Although ReRAM explicitly refers to the memory application, we often use the same term for resistive switching devices applied in functions beyond pure memory, that is, as a switch in integrated digital circuits, such as the field-programmable gate array (FPGA), or in other computing circuits, such as the neuromorphic systems aiming at mimicking the operation of the human brain. The purpose of this book is to address the fundamental physics and chemistry concepts as well as the technology, circuit, and architecture-related issues of the redox-based memristive phenomena and ReRAM devices.

The timeliness of the book can be estimated from the exponential increase in the number of papers published on ReRAM, rising from the few papers in the late 1990s to more than 1000 papers and more than 15 000 citations per year at the time this book is published.

The book is organized in 25 chapters, which are grouped in five parts. Each part comprises an area of information technology and starts with an Introduction. The first part (Chapters 1 and 2) provides a background in the concepts of resistive switching; the second part (Chapters 3–6) presents a survey of the relevant material properties; the third part (Chapters 7–12) covers microscopic physical/chemical processes that may occur in generic ReRAM devices; the fourth

part (Chapters 13–19) discusses the various types of ReRAM operation; and the last part (Chapters 20–25) provides an overview of the device technology, reliability aspects, and application areas.

The book is aimed at graduate students of physics, electrical engineering, and information technology, as well as those students studying material science or chemistry and focusing on electronic materials and technologies. We hope the book will also prove valuable to people working in the electronics industry interested in emerging technologies. The value of the book should lie on its linking the underlying principles with devices, technology, circuits, and applications. Prerequisites to understanding this book are undergraduate courses in sciences, particularly in physics and mathematics and an elementary introduction to electromagnetism, electronic materials, devices, and circuits. There are several different methods in which this book can be used. One method is to work through it as a classical textbook, chapter by chapter. Other readers might pick selected chapters. This is supported by the fact that the chapters are relatively self-contained. In some chapters on devices and applications, it might be advisable to read the corresponding basics chapter first.

Editing a book like this is only possible with the work of many dedicated people. We would like again to express our sincere appreciation to the authors who provided excellent contributions and have been willing to go through iterations in order to further enhance the consistency within the book. In addition, we are grateful to many colleagues in the world for numerous discussions and manifold suggestions. And we are particularly grateful to every member of our groups at the Research Center Jülich, RWTH Aachen University, and Politecnico di Milano, who supported the project by checking the symbols and formulas within the chapters and for their editorial assistance. All individual contributions are acknowledged at the end of every chapter. There has been a core team that did the technical and organizational work: First of all, we are grateful to Dagmar Leisten and Thomas Pössinger for editing and preparing hundreds of drawings and illustrations and for a careful project planning. Maria Garcia supported the project by assisting in getting the reprint permissions of figures and she contributed by organizing the editorial pages. Last but not least, we are indebted to Caroline Zurbelle-Waser for checking and improving the language of many of those chapters not written by native English speakers.

Milano, Italy
Aachen and Jülich, Germany
August 2015

Daniele Ielmini
Rainer Waser

List of Contributors

Hiro Akinaga

Nanodevice Innovation Research Center (NIRC)
National Institute of Advanced Industrial Science and Technology (AIST)
Central 2, 1-1-1 Umezono
Tsukuba
Ibaraki 305-8568
Japan

Stefano Ambrogio

Dipartimento di Elettronica
Informazione e Bioingegneria
Piazza L. da Vinci 32
20133, Milano
Politecnico di Milano
20133 Milano Italy

Masakazu Aono

National Institute for Materials Science (NIMS)
International Center for Materials
Nanoarchitectonics (MANA)
1-1 Namiki, Tsukuba
Ibaraki 305-0044
Japan

Geoffrey W. Burr

IBM Research - Almaden
650 Harry Road
San Jose
CA 95120
USA

Fred Chen

Industrial Technology Research Institute
Electronic and Optoelectronic Research Laboratories
195, Sec. 4, Chung Hsing Rd.
Chutung, Hsinchu, 31040
Taiwan Winbond Electronics Corporation
Taiwan

Yang-Yin Chen

Interuniversitair Micro-electronica Centrum (IMEC)
Kapeldreef 75
B-3001, Leuven
Belgium

Byung J. Choi
Hewlett Packard Labs
1501 Page Mill Road
Palo Alto
CA 94304-1100
USA

and

Seoul National University of
Science and Technology
Department of Materials Science
and Engineering
Seoul 139-743
Republic of Korea

Roger A. De Souza
RWTH Aachen University
Institute of Physical Chemistry
Landoltweg 2
52056 Aachen
Germany

Massimiliano Di Ventra
University of California
Department of Physics
San Diego
CA 92093-0319
USA

Regina Dittmann
Forschungszentrum Jülich
Peter Grünberg Institute (PGI-7)
Leo-Brandt-Straße
52425 Jülich
Germany

Andrea Fantini
Interuniversitair
Micro-electronica Centrum
(IMEC)
Kapeldreef 75
B-3001 Leuven
Belgium

Ludovic Goux
Interuniversitair
Micro-electronica Centrum
(IMEC)
Kapeldreef 75
B-3001, Leuven

Giuliano Gregori
Max Planck Institute for Solid
State Research
Heisenbergstr. 1
70569 Stuttgart
Germany

Tsuyoshi Hasegawa
National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki, Tsukuba
Ibaraki 305-0044
Japan

Ji-Hyun Hur
Samsung Advanced Institute of
Technology
Semiconductor Laboratory
Yongin
Gyeonggi-Do 446-712
Republic of Korea

Cheol S. Hwang
Inter-University Semiconductor
Research Center
Seoul National University
Department of Materials Science
and Engineering
Seoul 151-744
Republic of Korea

Hyunsang Hwang

Pohung University of Science and Technology (POSTECH)
Department of Materials Science and Engineering
77 Cheongam-ro
Nam-gu, Pohang
Gyungbuk 790-784
Republic of Korea

Daniele Ielmini

Politecnico di Milano
Dipartimento di Elettronica
Informazione e Bioingegneria
Piazza L. da Vinci 32
20133 Milano
Italy

Giacomo Indiveri

University of Zurich and ETH Zurich
Institute of Neuroinformatics
8057 Zurich Switzerland

Doo S. Jeong

Electronic Materials Research Centre
Korea Institute of Science and Technology
Hwarangno 14-gil 5
Seongbuk-gu
Seoul, 136-791
Republic of Korea

Tomoji Kawai

Osaka University
Institute of Scientific and Industrial Research
8-1 Mihogaoka
Ibaraki
Osaka, 567-0047
Japan

Michael N. Kozicki

Arizona State University
School of Electrical Computer and Energy Engineering
Center for Solid State Electronics Research (CSSER)
551 E Tyler Mall
Tempe
AZ 85287-5706
USA

Christian Lenser

Forschungszentrum Jülich
Peter Grünberg Institute (PGI-7)
Leo-Brandt-Straße
52425 Jülich
Germany

Eike Linn

RWTH Aachen University
Institute of Electronic Materials
52074 Aachen
Germany

Manfred Martin

RWTH Aachen University
Institute for Physical Chemistry
Landoltweg 2
52056 Aachen
Germany

Gilberto Medeiros-Ribeiro
Hewlett Packard Laboratories
1501 Page Mill Road
Palo Alto
CA 94304-1100
USA

and

Universidade Federal de Minas
Gerais
Brazil

Stephan Menzel
University
Forschungszentrum Jülich
Peter Grünberg Institute (PGI-7)
52425 Jülich Germany
Institut für Werkstoffe der
Elektrotechnik
Forschungszentrum Jülich
Sommerfeldstr. 24
52074 Aachen
Germany

Rene Meyer
Rambus
1050 Enterprise Way
Suite 700
Sunnyvale
CA 94089
USA

now

Memoria Technology
Cupertino
USA

Enrique Miranda
Universitat Autònoma de
Barcelona
Departament d'Enginyeria
Electrònica
Edifici Q, Bellaterra Bellaterra
celona
Spain

Maria Mitkova
Boise State University
Department of Electrical and
Computer Engineering
1910 University Drive
Boise
ID 83725-2074
USA

Monica Morales Masis
Ecole Polytechnique
Fédérale de Lausanne (EPFL),
Institute of Microengineering
(IMT),
Photovoltaics and Thin-Film
Electronics Laboratory,
Rue de la Maladière 71b,
2002 Neuchâtel,
Switzerland

Tomonobu Nakayama
National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki
Tsukuba
Ibaraki 305-0044
Japan

Alpana Nayak

National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki
Tsukuba
Ibaraki 305-0044
Japan

Takeo Ohno

National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki
Tsukuba
Ibaraki 305-0044
Japan

Yuriy V. Pershin

University of South Carolina
Department of Physics and
Astronomy
Columbia
SC, 29208
USA

Kin L. Pey

School of Electrical and
Electronic Engineering
Nanyang Technological
University
50 Nanyang Avenue
Singapore, 639798
Singapore

Nagarajan Raghavan

Interuniversitair
Micro-electronica Centrum
(IMEC)
Kapeldreef 75
B-3001 Leuven
Belgium

and

Singapore University of
Technology and Design (SUTD)
Singapore, 138 682
Singapore

Nagarajan Raghavan

School of Electrical and
Electronic Engineering
Nanyang Technological
University
50 Nanyang Avenue
Singapore, 639798
Singapore

Shriram Ramanathan

School of Materials Engineering
Purdue University
West Lafayette
IN 47907
USA

Takeshi Y. Sakamoto

Low-Power Electronics
Association and Project
Tsukuba 305-8569
Japan

Gurtej Sandhu

Micron Technology, Inc.
8000 S. Federal Way
Boise
ID 83707
USA

Akihito Sawa

National Institute of Advanced
Industrial Science and
Technology (AIST)
Electronics and Photonics
Research Institute
Tsukuba Central 5
1-1-1 Higashi 5
Tsukuba
Ibaraki 305-8565
Japan

Peter C. Schmidt

Technical University of
Darmstadt
Eduard-Zintl-Institut für
Anorganische und Physikalische
Chemie
Alarich-Weiss-Str. 8
64287 Darmstadt
Germany

Jun Y. Seok

Inter-university Semiconductor
Research Center
Seoul National University
Department of Materials Science
and Engineering
Seoul, 151-744
Republic of Korea

Rohit S. Shenoy

IBM Research - Almaden
650 Harry Road
San Jose
CA 95120
USA

Hisashi Shima

Nanodevice Innovation Research
Center (NIRC)
National Institute of Advanced
Industrial Science and
Technology (AIST)
Central 2, 1-1-1 Umezono
Tsukuba
Ibaraki 305-8568
Japan

Sabina Spiga

IMM-CNR
Consiglio Nazionale delle
Ricerche
Laboratorio MDM
Via Olivetti 2
20041
Agrate Brianza MB
Italy

John Paul Strachan

Hewlett-Packard Labs
1501 Page Mill Road
Palo Alto
CA 94304-1100
USA

Jordi Suñé

Universitat Autònoma de
Barcelona
Departament d'Enginyeria
Electrònica
08193 Edifici Q,
08193 Barcelona
Spain

Kazuya Terabe

National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki
Tsukuba
Ibaraki 305-0044
Japan

Stephan Tiedke

aixACCT Systems GmbH
Talbostraße 25
52068 Aachen
Germany

Antonio Torrezan

Hewlett Packard Laboratories
1501 Page Mill Road
Palo Alto
CA 94304-1100
USA

Tohru Tsuruoka

National Institute for Materials
Science (NIMS)
International Center for
Materials
Nanoarchitectonics (MANA)
1-1 Namiki
Tsukuba
Ibaraki 305-0044
Japan

Ilia Valov

University
Forschungszentrum Jülich
Peter Grünberg Institute (PGI-7)
52425 Jülich
Germany

Jan. M. Van Ruitenbeek

Leiden University
Kamerlingh Onnes Laboratorium
Leiden Institute of Physics
Niels Bohrweg 2
2333 CA Leiden
The Netherlands

Rainer Waser

RWTH Aachen University
Institut für Werkstoffe der
Elektrotechnik
Sommerfeldstraße 24
52074 Aachen
Germany

and

Forschungszentrum Jülich
Peter Grünberg Institute (PGI-7)
52425 Jülich
Germany
and
JARA-FIT

Franklin J. Wong

Harvard School of Engineering
and Applied Sciences
Cambridge
MA 02138
USA

H.-S. P Philip. Wong

Stanford University
Department of Electrical
Engineering
Center for Integrated Systems
420 Via Palou
Stanford
CA 94305-4075
USA

Dirk J. Wouters

Interuniversitair
Micro-electronica Centrum
(IMEC)
Kapeldreef 75
B-3001 Leuven
Belgium

now

RWTH Aachen University
Institute of Electronic Materials
Somerfeldstraße 24
074 Achen
Germany

Takeshi Yanagida

Osaka University
Institute of Scientific and
Industrial Research
8-1 Mihogaoka
Ibaraki
Osaka, 567-0047
Japan

Joshua J. Yang

University of Massachusetts
Department of Electrical and
Computer Engineering
100 Natural Resources Road
Amherst
MA 01003-929
USA

now

Department of Electrical and
Computer Engineering
University of Massachusetts,
Amherst
Amherst MA 01003-9292
USA

Young Yangliauw

Stanford University
Department of Electrical
Engineering
Center for Integrated Systems
420 Via Palou
Stanford
CA 94305-4075
USA

Simon Yu

Arizona State University
Department of Electrical
Engineering and Computer
Engineering
781 Terrace Rd, Rm 591

Tempe
AZ 85287
USA

Tobias Zacherle

RWTH Aachen University
Institute of Physical Chemistry
Landoltweg 2
52056 Aachen
Germany

Victor Zhirnov

Semiconductor Research
Corporation
1101 Slater Road
Suite 120
Durham
NC 27703
USA

1

Introduction to Nanoionic Elements for Information Technology

Rainer Waser, Daniele Ielmini, Hiro Akinaga, Hisashi Shima, H.-S. Philip Wong, Joshua J. Yang, and Simon Yu

This chapter provides the basic definitions of nanoionic redox-based resistive switching elements, their main switching modes, and their most important performance parameters. Furthermore, it sketches the scope of this book, which spans from nanoscale physics and chemistry of the switching phenomenon to devices, technology, and application areas. Three types of redox-based resistive switching elements are introduced, and a brief history of the phenomenon and the designated device application is given. This is followed by a discussion of the performance requirements and the specific themes of the application in memories and of the application in logic circuits, concluding with prospects and challenges.

1.1

Concept of Two-Terminal Memristive Elements

1.1.1

Classifications Based on Behavior, Mechanisms, and Operation Modes

Resistive switching denotes reversible phenomena of two-terminal elements, which change their resistance upon electrical stimuli in a nonvolatile fashion [1]. The *reversibility* is achieved by repeated applications of suitable stimuli, which control the resistance value between two or more levels. *Nonvolatility* means that the resistance change remains for a (long) retention time after the stimulus has been released. Phenomenologically, the stimulus affects an internal state variable of the element, which controls the resistance. For this reason, the resistance values are memorized by the elements, which are, therefore, also called *memristive elements* or *devices* [2]. The required switching speed and the retention times depend on the area of application and will be discussed later.

The fundamental physical principles of resistive switching and, hence, the nature of the internal state variable can be manifold. In a coarse-grained classification, one can distinguish between nanomechanical phenomena [3],

magnetoresistive effects such as spin-transfer torque (STT) [4], electrical effects such as leakage current through flash gate stacks in which trapping/detrapping phenomena occur [5], phase change between amorphous and crystalline phases [6], and *nanoionic redox phenomena* [7]. This book exclusively covers the last one, that is, ionic motion over nanoscale dimensions in two-terminal elements, which leads to local redox phenomena and, in turn, affects the resistance of the elements. Purely thermal resistance changes, which rely on the heat capacity of the elements after a temperature increase due to an electrical stimulus (as, e.g., in PTC thermistors), are not typically called resistive memories because the retention times related to the time constant of the temperature relaxation are too short.

Throughout this book, we will often use the term *redox-based resistive switching random access memory* (usually called *ReRAM*), in particular, if binary elements (i.e., elements that utilize *two* logic levels) are addressed, despite the fact that these elements may be used for functions beyond pure memory, that is, in the area of logic functions in various computational concepts. The term *memristive* will preferentially be used to describe the general characteristics of these elements and for applications of multilevel or analog features. In a more general sense, we will use the terms: resistive switching or memristive phenomena, elements, cells, devices.

Depending on the specific type of ReRAM, different operation modes have to be used. Figure 1.1 shows schematically characteristic current–voltage (I – V) diagrams recorded by periodic voltage sweeps (left) and pulse sequences with voltage pulse excitation and current responses (right). By far, the most device applications will use the *pulse mode*. However, the *I-V sweep mode* is helpful for obtaining an overview of the characteristics and for determining the (very) approximate threshold voltages V_{th} for the pulse operation. The resistance states of a ReRAM cell are called high resistance state (*HRS*) or *OFF state* and low resistance state (*LRS*) or *ON state*. For multilevel operation, intermediate resistance states are utilized as well. We assign the logic “0” state to the HRS and the logic “1” state to the LRS. A *write operation* changing a ReRAM cell from the HRS to the LRS is called a *SET operation*, while the opposite write operation is called a *RESET operation*. It should be noted that ReRAM cells often need an *electroforming* (short: forming) step prior to their first write/read operation. This electroforming step involves different voltage and current level than the write/read operations.

Most ReRAM systems reported in the literature are operated in the *bipolar resistive switching* (*BRS* or *BS*) mode (Figure 1.1a). Starting in the HRS, a SET process can be triggered by a voltage $V_{\text{SET}} > V_{\text{th}1}$, which leads to the LRS. Often, a current compliance (*cc*) is used for the SET operation in order to avoid damage to the cell and to optimize the operation. A read operation is performed at a much smaller voltage magnitude V_{rd} to detect the current while avoiding a detectable change of the state. A voltage signal V_{RES} of opposite polarity and an amplitude $V_{\text{RES}} < V_{\text{th}2}$ is used for the RESET process to switch the cell back into the HRS.

The *unipolar resistive switching* (*URS* or *US*) mode (Figure 1.1b) is characterized by the fact that all write and read operations can be performed with only one voltage polarity. Starting in the HRS, the SET process takes place at a voltage

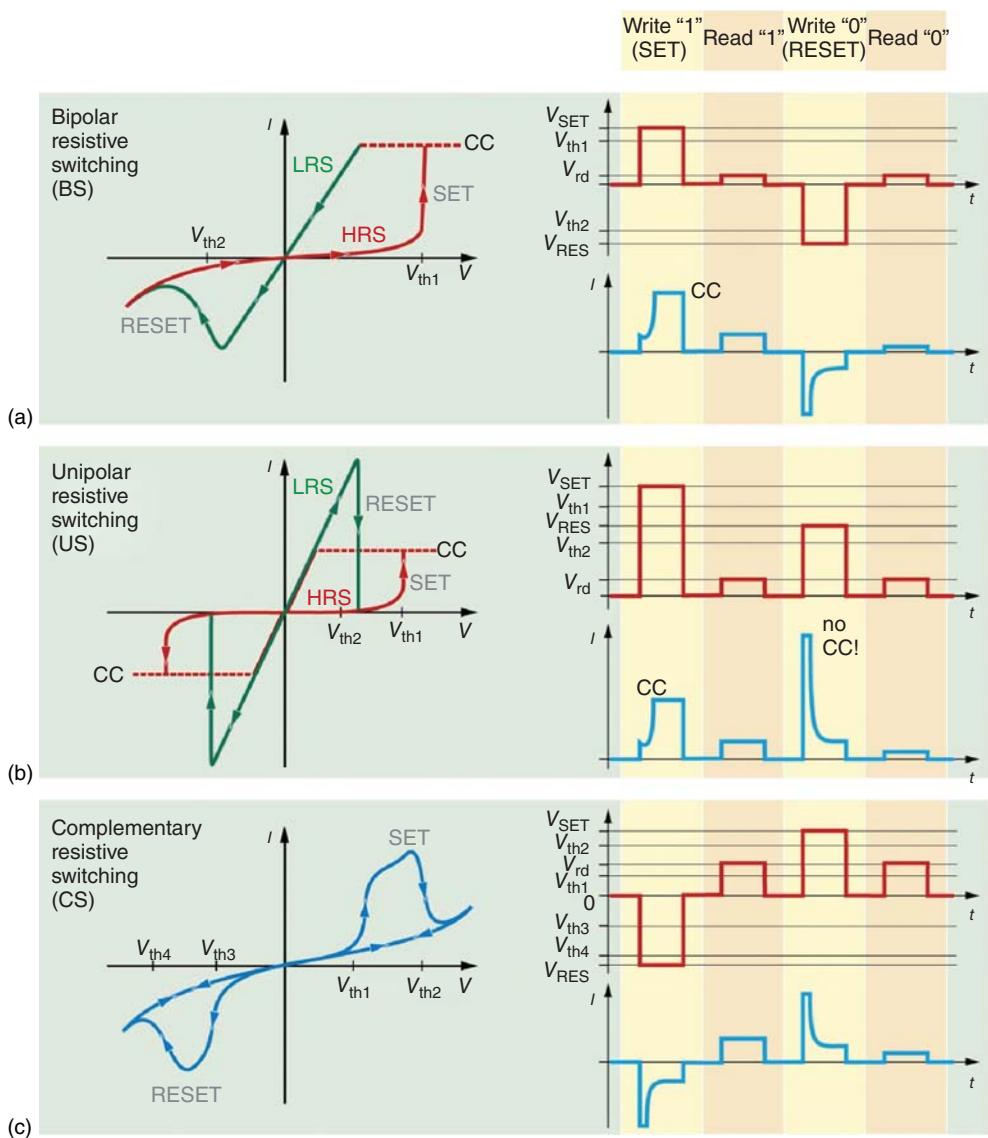


Figure 1.1 The three most common operation modes of different types of ReRAM elements shown for the I - V sweep operation (left) and the pulse operation (right). Details are described in the main text. Please note that the elements are nonvolatile. At first glance, the CS (CRS) mode resembles the so-called threshold switching, which shows

a hysteresis above a certain voltage bias but which disappears at voltages below this bias. The difference is the fact that the information is lost in the case of a threshold bias while it is maintained in a CRS cell and can be read out in the indicated manner. (From Ref. [8].)

$V_{\text{SET}} > V_{\text{th1}}$, with a LRS current limited by a cc. It is important that the cc is released in the RESET process with $V_{\text{RES}} > V_{\text{th2}}$, so that the current can exceed the cc value, which leads to changing back into the HRS. The read operation is performed at a small voltage V_{rd} as in the bipolar operation.

The *complementary resistive switching* (CRS or CS) mode (Figure 1.1c) can be obtained by connecting two BRS-type ReRAM cells in an antiparallel manner as suggested by Linn *et al.* [9]. In some cases, such a CRS behavior is also obtained by suitable processing and operation of single ReRAM cells [10]. Typically, the state of a CRS cell cannot be read at small voltages because the cell then always appears to be in a HRS. The state of the cell is only recognized at voltages $V > V_{\text{th1}}$. A read voltage $V_{\text{rd}} > V_{\text{th1}}$ will lead to a higher current (upper I - V trace in Figure 1.1c, left) in the case of a logic “1” state and to a lower current (lower I - V trace in Figure 1.1c, left) in the case of a logic “0” state. The write “0” is obtained by a positive voltage $V_{\text{wr}} > V_{\text{th2}}$, and a write “1” is obtained by a negative voltage $V_{\text{wr}} < V_{\text{th4}}$. Because of the relatively high read voltage amplitude $V_{\text{rd}} > V_{\text{th1}}$ and the corresponding currents, the internal state is affected by the read operation, that is, the read voltage may destroy the logic state (the so-called destructive readout, DRO). As a consequence, the last logic state needs to be rewritten into the cell after every read operation. This is the same situation as in the case of the standard (volatile) dynamic random access memory (DRAM) cells.

The major performance parameters of ReRAM devices are:

- *Resistance Values*: R_{LRS} and R_{HRS} (or: R_{ON} and R_{OFF}) and the resistance ratio $R_{\text{HRS}}/R_{\text{LRS}}$.
- *SET and RESET Voltages*: V_{SET} and V_{RES} , respectively.
- *Write Currents*: in particular, the current in the ON state, at a voltage amplitude just above V_{SET} .
- *Write Speed*: the shortest electrical pulse able to change the resistive state.
- *Retention Time*: the time for which a resistive state is maintained without a voltage applied to the cell.
- *Endurance*: the number of switching cycles before the resistance ratio fatigues to an unacceptable value.
- *Operation Energy per Bit*: the energy required to write a cell, that is, to change its resistive state.
- *Scalability*: the geometrical size to which a cell can be miniaturized before it encounters inherent (physical) limits.
- *Stackability*: the option to stack several layers of cell on top of one another by fabrication technology.
- *Multilevel Storage*: the option to store more than one bit of information in one cell.

A *microscopic analysis* of ReRAM cells reveals that the electrodes and, in particular, the material “I” between the metal electrodes M in the metal–insulator–metal (MIM) structure of the cell participate in a different manner in the resistive switching process. Typically, MIM structures are sandwich stacks. But, additionally, lateral MIM structures on surfaces and metal electrodes

are used as scanning probe tips, in particular for research purposes. Without addressing the physics and chemistry of the switching at this point, one can distinguish the categories of the locations of the switching event in a MIM stack. Along the path from one electrode to the other (which we will call the vertical direction), one can distinguish the switching locations:

- at/near one of the electrode interfaces
- at the (approximate) center between the electrode interfaces
- involving the entire path between the electrode interfaces.

In the perpendicular direction, that is, in the plane of cross section of the cell (which we will call the lateral direction), one can distinguish:

- *filamentary switching*, that is, switching by the formation and dissolution of a single *conducting filament* (CF),
- *area-proportional switching*, that is, a switching event that takes part over the entire cross section of the cell so that the currents become proportional to the cross-sectional area of the cell. Because this switching typically occurs close to one electrode, it is also called *interface-type switching*.

Figure 1.2 illustrates these scenarios schematically. Of course, this brief description represents a very simple, coarse-grained classification only. In reality, many intermediate cases have been reported, such as multiple filament switching or switching areas that are relatively extended spots without spreading over the entire cross section of the cell. It should further be pointed out that these categories and combinations thereof do not occur equally. The most frequent resistive switching described in the literature is filamentary switching (with evidence provided, for example, in Refs. [11–25]). If the cell shows a bipolar operation, there are usually clear indications that the switching is localized both laterally (filamentary) and vertically (near one of the electrode interfaces).

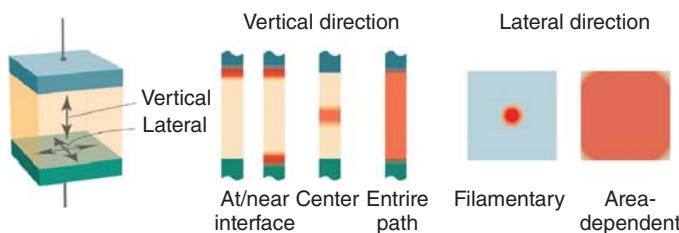


Figure 1.2 Geometrical location of the switching event in a ReRAM cell. In the vertical direction, the switching may happen close to one or both electrodes, in the center between the electrodes, or over the entire path between the electrodes. In the lateral direction, we distinguish between a localized filamentary switching and a switching that involves the entire cross

section. For example, a bipolar filamentary resistive switch constituting the majority of ReRAM devices reported to date combines a localized event in the lateral direction at/near an electrode interface in the vertical direction. Still, of course, the geometries shown here are limiting cases and intermediate situations may be encountered too.

The simple ReRAM concept might resemble that of a switch, which can be settled in two different configurations to allow a current flow or not. Although apparently simple, the explanation of the physical processes responsible for the switching has not been completely unveiled at the time of publication of this book, not to mention the high complexity to achieve a working technology based on the ReRAM concept, for example, a crossbar nonvolatile memory or a logic circuit outperforming the existing technologies. The purpose of this book is to address the fundamental physics and chemistry concepts as well as the technology, circuit, and architecture-related issues of the redox-based memristive phenomena and ReRAM devices.

1.1.2

Scope of the Book

Figure 1.3 is a sketch of the book topics and structures, covering the concept, the materials involved, processes in generic cells, the types of ReRAM cells, aspects of technology, and applications.

A phenomenological introduction of ReRAM cells in the context of two-terminal devices as dynamical systems will be provided in Chapter 2.

The material properties, described in Chapters 3–6, provide the basis for understanding the *physics and chemistry* of the processes that occur during redox-based resistive switching. These chapters cover the atomic and electronic structures with an emphasis on metal oxides used for ReRAM cells (Chapter 3), the lattice disorder and defect chemistry of these materials (Chapter 4), the ionic transport mechanisms (Chapter 5), and the electronic transport properties (Chapter 6). Transport properties in fact dominate resistive switching, since, microscopically, the switching is due to the electrical and ionic transport phenomena. For instance, electrical transport is essential in the read operation, where the conductance state of the memory is probed. On the other hand, ionic transport is responsible for the change of the chemical composition profile, leading to resistance switching. Electrical and ionic transport are essential in dictating the switching time, the switching energy, and the scale of the switching phenomena, which are key parameters in assessing the potential application space of ReRAM within memory and computing circuits [26, 27].

For a more detailed description of redox-based resistive switching, however, we need to peer inside the cell, where the mass transport and chemical composition change leads to the memory effect. In this book, the fundamental phenomena taking place within a generic cell are grouped according to quantum point contact (QPC) transport (Chapter 7), controlled dielectric breakdown (Chapter 8), and several microscopic processes responsible for resistive switching such as phase formation, space-charge formation, electron transfer at electrochemical interfaces, electromotive force (emf) formation, and kinetic demixing (Chapter 9). In most cases, device operation relies on an initial forming operation, which is introduced in Chapter 10. A universal switching behavior for filamentary-type

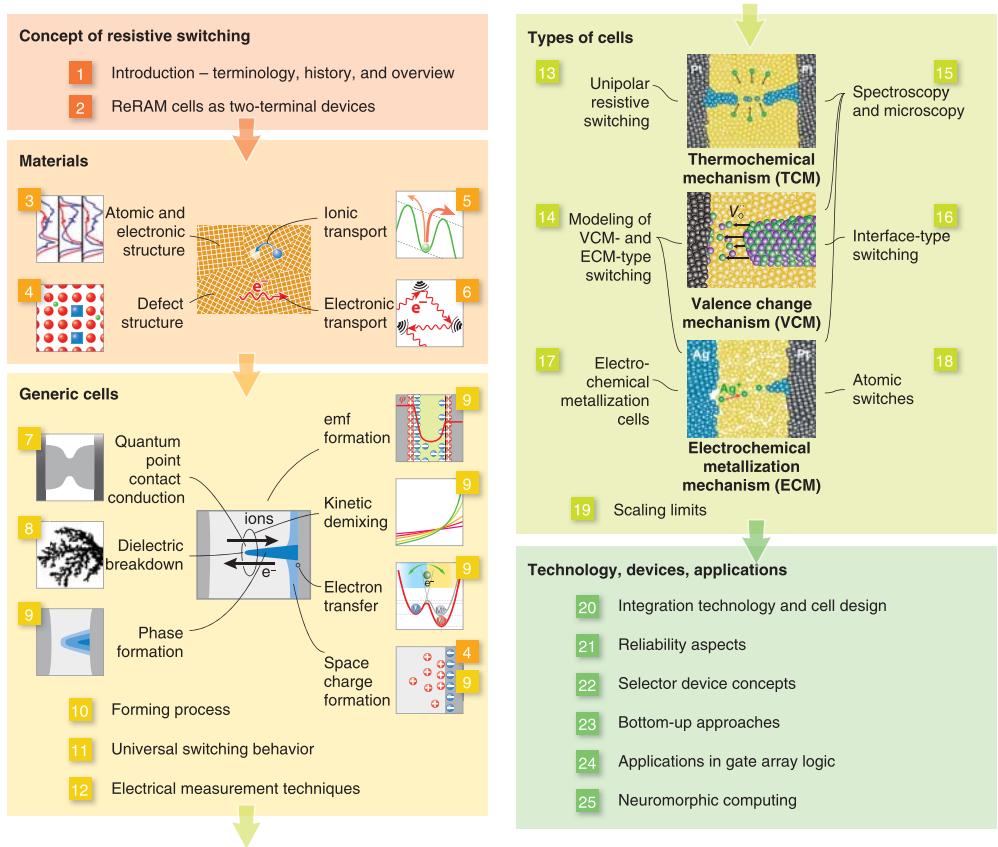


Figure 1.3 Synopses of the ReRAM concepts, materials, generic cells, ReRAM cell types, and ReRAM technology and applications. The numbers indicate the chapters in this book.

switching is introduced in Chapter 11, while an overview of the electrical techniques for device characterization is given in Chapter 12.

ReRAM cell can be classified by their basic nanoionic switching mechanism, covered in Chapters 13–16. The three main *ReRAM types* are electrochemical memory (ECM), valence change memory (VCM), and thermochemical memory (TCM). Please note that the same resistive switching material, and even the same cell, might show more than one switching behavior, for instance, the coexistence of VCM and TCM behaviors was demonstrated in TiO_2 [28] and HfO_2 [29]. This indicates that the switching behavior is dictated not only by the material type but also by the electrical operation and the type of electrodes. In general, all of the three ReRAM types display *filamentary* switching, except for VCM, which might, in addition, take place as area-dependent *interface-type* switching (Chapter 16) or even in a uniform *bulk-type* switching [30] and forming [31].

Unipolar switching in TCM devices (Chapter 13) typically takes place in metal oxides such as NiO [32] and TiO₂ [33]. Electrodes are generally inert, for example, Pt or TiN, and a forming process is needed to first initialize the ReRAM operation. SET and RESET processes are induced by applying voltage signals of the same polarity. For this reason, TCM is believed to result from purely chemical redox transitions, where a metal-rich CF is formed by local reduction of the metal oxide, while the CF is then dissolved by local oxidation, possibly in combination with a surface tension controlled Rayleigh instability of the CF. Both reduction and oxidation are accelerated by the high temperature developed at the CF by Joule heating [34, 35].

Microscopic models to describe the two BRS mechanisms, the VCM-type and the ECM-type switching, are presented in Chapter 14. VCM cells are based on metal oxides sandwiched between metal electrodes of different oxygen affinity and/or work functions. Similarly to TCM, the cells must typically first be initialized by the forming process, consisting of a controlled dielectric breakdown that generates active ionic defects and typically leads to the formation of a reduced oxygen-deficient CF phase with a valence change in the cations and a corresponding increase in the electronic conductivity. The ionic defects can then migrate over nanoscale distances in response to an applied voltage, allowing for the formation and disruption of the CF under bipolar voltage operation. The VCM process is investigated by spectroscopical and microscopical methods (Chapter 15). Cases of an interface-type VCM mechanism are covered in Chapter 16.

In ECM cells, the MIM stack includes at least one electrochemically active metal electrode, of which the most typical case is Ag or Cu. Under a positive applied voltage, the active electrode can oxidize and release cations in the “I” layer, also known as the solid electrolyte. Cations (Ag^+ , Cu^{2+}) then migrate to the negatively biased counter electrode, where they get reduced back to the neutral state and grow as a CF or bridge. In the literature, the ECM devices are also called Conductive Bridge RAM, short: CBRAM. The formation of the CF corresponds to the SET process. The bridge can then be electrochemically dissolved under a negative voltage applied to the active electrode as the RESET process. The general form of ECM cells is covered in Chapter 17, whereas those that emphasize the atomic contact of the metallic CF and conductance quantization are called *atomic switches*, described in Chapter 18.

All ReRAM types are subject to scaling rules and ultimate physical limits of scaling, which are discussed in Chapter 19.

To achieve functional and cost-effective electronic circuits, ReRAM must be integrated with complementary MOS (CMOS) digital circuits by process technology and designed to match the specifications of certain applications. Cell design and integration technology based on top-down approaches, such as optical lithography, as well as 3-D integration concept for further increasing the device density per area on the chip are covered in Chapter 20. Highly important topics in the ReRAM area are the reliability with respect to endurance, retention, random telegraph noise, and disturb error as well as variability (Chapter 21). A main advantage of ReRAM for memory applications, compared to CMOS

technology, is the two-terminal architecture, which allows a small device area of only $4F^2$ (F is the lithographically defined minimum feature size) in the crossbar array [36]. However, implementation of ReRAM in the crossbar array requires that each resistive switching device is accompanied by a selector (Chapter 22), to prevent inevitable interference during read and program between array cells [9]. Besides the top-down integration technology described in Chapter 20, there are techniques based on bottom-up approaches utilizing chemical and physical methods of self-assembling and self-aligning of structures (Chapter 23).

Beyond memory applications, ReRAM devices have been proposed for the use in logic circuits. One area is the logic based on field-programmable gate arrays (FPGAs) where ReRAM devices can be employed in the configuration array and for routing switches (Chapter 24). Finally, ReRAM devices hold the promise that they can be used for artificial synapses in neuromorphic computing circuits (Chapter 25).

1.1.3

History

From a historical point of view, resistive switching has been studied in various solid-state materials since the early 1960s. Most typically, very early reports on resistive switching emphasized the negative differential resistance occurring in the I-V characteristics of MIM structures. The “I” in the MIM stack can be one of the various binary and multinary oxides, chalcogenides, as well as group-IV, III–V, and II–VI semiconductors. In 1962, Hickmott reported a large negative resistance in MIM stacks with oxide insulator, such as SiO_x , Al_2O_3 , Ta_2O_5 , ZrO_2 , and TiO_2 [37]. The negative resistance effect in oxide-based MIM structures was explained by a space-charge-limited current in the oxide layer. In insulators, the trapping of electrons in localized states influences the space-charge-limited current [38]. A double injection model was proposed by considering that the lifetimes for the injected electrons and holes were different in insulators and vary with injection level [39]. This model explained the curve of the negative resistance and showed the possibility of spontaneous oscillations under the application of an appropriate DC voltage or an apparent breakdown at some critical voltage followed by a marked hysteresis in the current with decreasing voltage after the breakdown.

Bistable resistance switching was reported in 1964 in NiO thin films on Ni substrate, where the switching was believed to be due to the formation and rupture of a nickel metallic filament in the NiO layer sandwiched by two electrodes [40]. Later in 1965, bistable resistive switching between two stable resistance states was shown in Nb_2O_5 [41]. Figure 1.4 shows the reported I - V curves $\text{Bi}/\text{Nb}_2\text{O}_5(125 \text{ nm})/\text{Nb}$, measured after formation (dielectric breakdown) of the initially insulating stack. Upon a first positive voltage sweep, the device is in an LRS (a). Application of a negative voltage leads to a RESET transition to high resistance (b), while application of a positive voltage causes the SET transition to the initial LRS (c). The bipolar switching is bistable in that both states are stable.

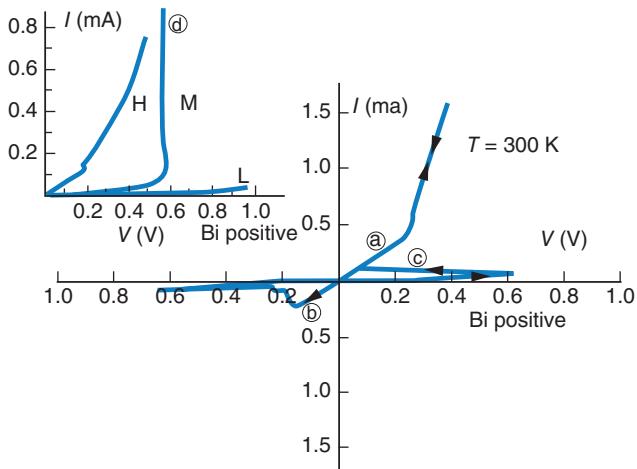


Figure 1.4 Measured I - V characteristics for an Nb–Nb₂O₅–Bi MIM stack showing bistable resistive switching. The device is initially in a low resistance state (a) due to the previous forming operation. Reset transition to the high resistance is shown for negative applied voltage (b), while set

transition to the low resistance appears at positive voltage (c). The inset shows the I - V curves of three stable states: a high resistance state H, a low resistance state L, and an intermediate state M (d). (Redrawn after Ref. [41].)

Studies on bistable resistive switching have also been reported for thin films of Ta₂O₅ [42], SiO [43], TiO₂ [44], and Al₂O₃ [45] and for ZnSe–Ge heterostructures [46]. The first report on what we today classify as ECM-type switching dates back to 1976 when Hirose and Hirose observed Ag dendrites being formed and dissolved between the Ag and Au electrodes in a bipolar operation mode of lateral Ag/As₂S₃/Au cells [47]. This early period of research faded in the late 1970s. Obviously, the interest in this area decreased because of the overwhelming progress of the Si-based integrated circuit technology, in particular, the Flash memories. In 1967, Kahng and Sze invented the floating gate MOSFET as a nonvolatile three-terminal device [48], which was developed into an EPROM (erasable programmable read-only memory) in the 1970s and into a flash memory in the 1980s. Flash memories have developed into dominant nonvolatile solid-state storage since the turn of the millennium. Another reason for the decrease in research on metal oxides and related compounds was presumably the lack of progress in understanding and controlling the resistive switching phenomena possibly due to insufficient analytical tools at that time. This period has been reviewed comprehensively by Dearnaley *et al.* [49], Oxley [50], and Pagnia and Sotnik [51].

A new era in research on resistive switching gradually started in the mid-1990s. The Tokura group found electrically triggered resistive switching in Pr_xCa_{1-x}MO₃ (PCMO) while investigating the magnetoresistive properties of this material [52], and Zhuang *et al.* used this material to fabricate the first integrated ReRAM memory, a 64-bit array, using a 500 nm CMOS technology [53]. In 2000, the

IBM Zurich lab reported the resistive switching of perovskite-type zirconates, including many properties that are essential to non-volatile memory (NVM) applications [54]. In the ECM-type area, Kozicki, Mitkova *et al.* started to study the Ag-GeSe systems in the late 1990s [55], while the Aono group published their first report on the so-called atomic switches in 2001 [56]. These devices make it possible to control the electrochemical formation and dissolution of, for example, an Ag atomic bridge in a nanogap between a mixed electronic–ionic conducting Ag^+ electrolyte and a metal electrode with the precision of Landauer conductance quantization [57]. In 2004, Samsung successfully demonstrated a high-density ReRAM chip using a 180 nm technology. It was based on unipolar switching Pt/NiO/Pt cells with an endurance of 10^6 SET/RESET cycles. These and related papers have been the beginning of an unprecedented rise of R&D activities, which led to the mega-trend that we encounter in the 2010s.

While there has been a basic understanding of the ECM mechanism from the beginning, a broad spectrum of mechanisms have been suggested as the underlying mechanisms of resistive switching in various metal oxide systems. In 2006, Waser's group was able to clarify the effect as a motion of oxygen ions and a coupled valence change in the cation sublattice on the nanometer scale at structural defects in the crystal lattice of the metal oxides near one electrode [58]. For this reason, the term VCM effect was suggested for bipolar metal oxide systems [59], extending the more detailed classification of the nanoionically driven, redox-process-based resistive switching memories [60]. The first spectroscopic evidence for the valence change was published for the dopants of Cr-doped SrTiO_3 by Janousch *et al.* [61] and for the host cation in TaO_x -based cells by Z. Wei *et al.* [62]. The filamentary nature of an oxygen-deficient phase generated during the forming process was shown by Kwon *et al.* [63] using high-resolution transmission electron microscopy. In 2008, Williams' group at the Hewlett-Packard Labs discovered that the electrical characteristics of resistive switching elements can be described in terms of the theory of memristive devices [64]. This theoretical concept was published in 1976 by Kang and Chua [2] as a generalized version of the theory of memristors introduced by Chua in 1971 [65], and this link led to a further increase in the international research activities.

Further historical notes will be given in the forthcoming part of this introduction and, in particular, in the individual chapters of this book. The technological realization has advanced to the point that a VCM-type 32 GB ReRAM chip was presented by Sandisk and Toshiba in 2013 [66], and an ECM-type 16 GB ReRAM chip was introduced by Micron and Sony in 2014 [67].

As a guidance for a further understanding of the nanoionic redox-based resistive switching elements covered in this book, it should be mentioned that there are other two-terminal resistive switching material systems based on different switching mechanisms: (i) Ferroelectric Schottky diodes were proposed as bipolar switching memory devices by Esaki *et al.* in 1971 [68]. A first realization based on epitaxial PbTiO_3 films grown on electronically conducting oxides was reported by Blom *et al.* 1994 [69]. These activities led to resistively switching ferroelectric and multiferroic tunnel junctions, briefly covered in Chapter 15. (ii) Ovshinsky

described a unipolar, thermally driven switching between the amorphous and the crystalline phase of Ge-Te-based compositions in 1966 [70, 71], which became important for rewritable optical storage media in the 1990s and as phase-change memories (PCMs) in the mid-2000s. (iii) For several decades, there have been reports on *organic* materials that show resistive switching (see e.g., Ref. [72]). In some prominent cases, it was discovered later that the switching, in fact, takes place in an oxide layer formed on an electrode metal used to contact the organic material. For example, Cu : TCNQ films sandwiched between Cu and Al electrodes [73] were found to switch resistively because of an ECM effect in the Al_2O_3 layer built during the processing of the system [74]. A similar situation was encountered for rose Bengal films between Al and Zn or indium tin oxide (ITO) electrodes [75] for which the resistive switching presumably is caused by a VCM-type effect in the metal oxide layer of one of the electrodes [76]. (iv) Nanosized dispersions of electronically conductive phases, for example, metals such as Pt, in insulators such as SiO_2 have also shown pronounced BRS effects if sandwiched between asymmetric electrode materials (e.g., $\text{Pt}/\text{SiO}_2:\text{Pt}/\text{Mo}$) and if the film thickness is close to the percolation threshold [77]. These systems have been described in terms of a purely electronic trap charging/discharging model for which, however, the origin of the highly nonlinear switching kinetics (see Section 1.2.1) needs to be further clarified. A clear electronic trap mechanism is found in the bipolar MemFlash cells, which are conventional flash memory cells with the source and gate contacts connected in order to obtain a two-terminal device [78]. Although this concept inheres the properties of flash cells with respect to write speed and write voltages, it may be further developed in order to match the requirements of neuromorphic devices [79]. (v) For narrowband-gap Mott insulator compounds AM_4X_8 ($\text{A} = \text{Ga}, \text{Ge}$; $\text{M} = \text{V}, \text{Nb}, \text{Ta}$; $\text{X} = \text{S}, \text{Se}$), a URS effect has been reported, which is apparently related to a Mott insulator–metal transition triggered by an electronic avalanche effect above a critical field strength. This effect may be described as an electronic phase-change effect in the crystalline phase in combination with local strain [80]. There are further reports about other purely electronic resistive switching mechanisms based on electron correlation, for example, the unipolar switching of NiO [81]. Further clarification of these effects is subject to ongoing research at the time of the publication of this book.

1.2

Memory Applications

1.2.1

Performance Requirements and Parameter Windows

This section reports on the major parameters of ReRAM devices for various types of memory applications and, in some aspects, also for logic concepts. The demonstrated and the expected performance of ReRAM devices are compared to the performance of alternative nonvolatile device technologies.

The development of oxide-based VCM-type ReRAM has progressed rapidly since the mid-2000s. In particular, binary oxides using materials that are familiar to the semiconductor industry have seen intense research and development in both industry and academia. The objective of an oxide ReRAM is a non-volatile memory that is fast, high-density, and compatible by integration with conventional silicon CMOS technology. The early ReRAM had large device areas ($\gg \mu\text{m}^2$), large programming currents ($> \text{mA}$), long programming times ($> \mu\text{s}$), and low endurance ($< 10^3$ cycles). Over the years, many of these drawbacks have been overcome.

Great *scalability* is a major differentiator of ReRAM from other competing technologies, including commercialized technologies such as Flash, SRAM, and DRAM, as well as some prototype and niche technologies such as magnetoresistive memory (e.g., STT-MRAM) and ferroelectric memory (FeRAM). Device sizes down to $10 \text{ nm} \times 10 \text{ nm}$ have been demonstrated [82]. The scalability into the single-digit nanometer regime arises from the fact that the resistance change in ReRAM is usually associated with ionic motions within a much more localized structure, which could be as small as a nanometer or less [58, 83]. This is one of the reasons that makes ReRAM a highly promising technology.

Reading of ReRAM devices is *nondestructive* (nondestructive readout, short: NDRO). This differs from other memories, such as FeRAM, where every read operation changes the information stored in the device and requires a restore cycle. The *resistance ratio* (also called: resistance window) is usually fairly large for ReRAM, usually between 10 and > 1000 . Presumably, this is because a small change in ion positions may induce exponentially large changes in the electronic conductance of the cell. Compared with the resistance window of magnetic RAMs (MRAMs), which is typically less than 5, the large $R_{\text{HRS}}/R_{\text{LRS}}$ ratio of ReRAM offers a larger sensing margin.

The absolute values of the R_{HRS} and, in particular, of the R_{LRS} are of importance too when it comes to the design of ReRAM circuits. They have to be compared to the bit line resistance R_{BL} and the resistance of the select device R_{SD} , such as a transistor in the ON state in the case of an active matrix. The bit line capacitance C_{BL} has to be taken into account as well. For example, if one considers a 512×512 matrix and an $F = 10 \text{ nm}$ technology, then C_{BL} and $(R_{\text{BL}} + R_{\text{SD}})$ are in the range of 30 fF and $1 - 10 \text{ k}\Omega$, respectively. For a read voltage of 0.3 V and a read speed of $t_{\text{rd}} < 100 \text{ ns}$, one can estimate a read current of $> 100 \text{ nA}$ and a maximum R_{LRS} of approximately $300 \text{ k}\Omega$. Please note that, as a consequence, the current density across the $10 \text{ nm} \times 10 \text{ nm}$ ReRAM cell will be as high as 10^5 A/cm^2 . For a reasonable sense voltage margin, the minimum R_{HRS} needs to be approximately $500 \text{ k}\Omega$. The minimum value of R_{LRS} is obtained by the write operation in which a sufficiently large voltage must drop over the cell, that is, $R_{\text{LRS}} > (R_{\text{BL}} + R_{\text{SD}})$ should be fulfilled. Details are provided, for example, in Refs. [26, 84, 85].

These resistance values and the current density should be kept in mind, when possible switching mechanisms are compared and also, for example, when promises made in the literature concerning the scalability of area-proportional, interface-type ReRAM concepts are to be judged.

The write speed of ReRAM is typically in the order of nanoseconds or tens of nanoseconds, which is orders of magnitude faster than the write speed of Flash cells. In dedicated studies, ReRAM devices have been observed to switch as fast as 100 ps [86] and potentially even faster. One possible reason for the ultrahigh speed is that the distance for the ionic species to move during switching is likely to be fairly short, approximately a nanometer or less. The other key reason is that the nanoscale active region inside the switching oxide material is under very high electric field for enhanced ion mobility [87] and is heated up during switching by Joule heat [88], which greatly increases the mobility of the ionic species and enables fast switching. Upon removing the switching electrical pulse, the heated region cools down very quickly, probably within a nanosecond, and the mobility of the ionic species drops to a very low level again. This leads to a great *retention* at relatively low electrical bias and temperature. Long retention times are kept under read voltage bias, because a low voltage bias avoids any temperature and field acceleration of the ions and, hence, their mobility stays very low.

The required *ultrahigh nonlinearity of the switching kinetics*, that is, nanosecond switching types for a certain write voltage and many years of retention even if constantly read out at, for example, 1/10th of the write voltage, represents a challenge for many proposed mechanisms. In many papers, purely electronic switching mechanisms have been suggested based on the idea of trapping/detrapping charges in the “I” layer of the MIM stack of ReRAM cell, obviously inspired by the flash memory. However, in contrast to the gate dielectrics, the “I” layer in a ReRAM cell is highly conductive. Hence, excess charges stored by trapping or detrapping can only be kept for an extremely short retention time. In a very coarse approach, this can be compared to the idea of storing excess charge on a very leaky capacitor in which the RC time of the device determines the retention time. Details are provided in Ref. [89].

The reports on the maximum number of switching cycles, the *endurance* of ReRAM cells vary a lot. An endurance of 10^{12} cycles has been reported for a TaO_x -based VCM cell [90]. Typically, 10^6 – 10^8 cycles are reported, which by far excels highly dense NAND flash, which shows 10^3 – 10^5 cycles.

Another key parameter for any nanoelectronic device is the *operation energy per bit*. The energy consumption in reading memory cells is normally not a concern since it is usually much lower than that of write operation. There is not much ability to reduce energy by reducing switching voltages because the switching voltages are close to each other even for different ReRAM types. In contrast, the switching current can be significantly reduced by reducing the device size or optimizing the switching materials. Furthermore, it varies tremendously by the ReRAM types. While TCM devices often need switching currents in the millampere range, ECM cells can be tuned to switch at currents below 1 nA (with low retention). As an example of VCM, TaO_x -based nanodevices ($50 \text{ nm} \times 50 \text{ nm}$) can switch at sub 2 ns times under sub 2 V with less than $10 \mu\text{A}$, resulting in a sub pJ/bit operation energy [91].

In many ReRAM devices, the device resistance can be switched to any value in between HRS and LRS, displaying an analog behavior. This analog switching

behavior can be utilized for *multilevel storage*, that is, to store more bits per cell [92, 93]. Landauer conductance steps may be regarded as natural levels for ultimately scaled devices, as demonstrated by ECM-type atomic switches [57]. Multilevel storage serves as the basis for analog computing and neuromorphic computing [94, 95].

There are no exotic materials or processes involved in many of the ReRAM device concepts. The fabrication processes and the materials are *CMOS compatible*, which allows for the integration of these devices with CMOS chips in a normal CMOS fabrication. Because of the materials involved, *stackability* is feasible for most of the ReRAM types. The device stack may be as simple as a conductor/insulator/conductor trilayer, and all the layers can be polycrystalline or even amorphous. The practical limitation of the number of layers that can be stacked is the alignment and the electrical addressability of these stacked layers. Some details are sketched in Section 1.2.3, and more details in Chapter 20.

In comparison with ReRAM, three prototypical memories, parameters for FeRAM, STTRAM, and Phase change RAM (PCRAM), are summarized in Table 1.1. While FeRAM was the first to be commercialized, its difficulties in scalability make it a poor candidate for high-density memory applications, despite its excellent latency, power, and endurance characteristics. STTRAM is a very promising candidate to reduce the power consumption of the memory system, particularly for embedded applications, coming from the highest performance among emerging memories in terms of speed and endurance. On the other hand, the complexity of the materials involved and their scalability including the thermal stability, the low write power, and the resistance distributions (e.g., the low ON/OFF ratio) are current challenges. PCRAM has a certain advantage regarding the 3D integration, because PCRAM requires only unipolar selection device in passive arrays, instead of bipolar selectors for ECM- and VCM-type ReRAM (see Section 1.2.2 and Chapter 22). Overall, ReRAM shows many advantages, also in comparison with the flash technology. In Table 1.1, the major performance parameters are outlined for the three types of ReRAM for 2009 and 2015.

While ReRAM has the potential to be a *stand-alone*, high-capacity, nonvolatile memory technology, it may be even more suitable for *embedded applications*. This is because it offers the low programming voltage that flash does not provide in combination with a speed that is comparable to that of DRAM. For example, an 8T2R nonvolatile SRAM (static random access memory) cell with two ReRAMs at two storage nodes of SRAM has been demonstrated for dynamical power management [97]. The ability to position a ReRAM at the contact vias of the MOSFET without extensive process steps [98] is also an attractive device feature, especially for embedded applications where only a low memory capacity, multiple-time programmable, nonvolatile memory is required. Therefore, there is an enormous opportunity to completely rethink the design of the system to gain new functionalities and even orders of magnitude improvements in speed and/or power consumption. In 2013, Panasonic released the first microcontroller with embedded ReRAM [99].

Table 1.1 Potential of the current and emerging nonvolatile memory candidates.

Parameter	FeRAM	STTRAM	FCRAM	ReRAM			
				Year	ECM (CBRAM)	VCM (Bipolar redox)	TCM (Unipolar redox)
Maturity	●	●	●	2015	Product	Product	Research
	●	●	●	2009	Prototype	Research	Research
Scalability/CMOS process compatibility	▲	●	●	2015	20 nm	5 nm	35 nm
	●	●	●	2009	90 nm	180 nm	180 nm
3D integration	▲	●	●	2015	–	Demonstrated	–
	●	●	●	2009	–	–	–
Endurance	●	●	●	2015	$>10^{10}$	$>10^{12}$	$>10^6$
	●	●	●	2009	$>10^9$	$>10^6$	$>10^6$
Retention	●	●	●	2015	>10 years	>10 years	>10 years
	●	●	●	2009	>10 years	>8 months	–
Latency	●	●	●	2015	<1 ns	<1 ns	5 ns
	●	●	●	2009	5 ns	10 ns	–
Power	●	●	●	2015	~1 pJ	~1 pJ	–
	●	●	●	2009	~1 pJ	~1 pJ	–
Variability	●	●	●	2015	–	–	–
	●	●	●	2009	–	–	–



Maturity	Research	Prototype	Product
Scalability	$45 \text{ nm} < F$	$10 < F < 45$	$F < 10 \text{ nm}$
3D integration	Difficult	Possible	Feasible
Endurance	$\leq 10^5$	$\leq 10^{10}$	$>10^{10}$
Retention	< 1 year	< 10 years	>10 years
Latency	>10 μs	>100 ns	$\leq 100 \text{ ns}$
Power	High	Medium	Low
Variability	Problematic	Reasonable	Low

Parameters for ReRAM show the situation for 2009 and 2015 to illustrate the progress in this area. Data are taken from Ref. [96] with modifications.

1.2.2

Device Isolation in Crossbar Arrays

At memory array level, one can distinguish between active arrays and passive arrays, by which the device density can be determined. In active arrays, every memory element is connected to a select transistor as in the DRAM or NOR flash architecture. Such cells are called 1 transistor – 1 resistor cells (1T1R). The 1T1R cell area ranges from $6F^2$ using aggressive DRAM-like design to approximately $18F^2$ for a very relaxed design.

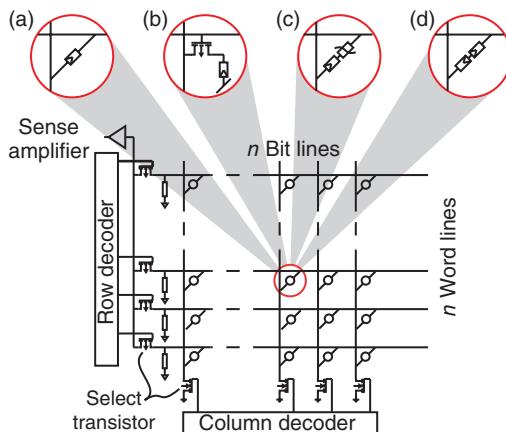


Figure 1.5 The cross-point memory architecture (a) without and (b–d) with cell selection elements. (b) Represents 1T1R cells, (c) cells with a diode or varistor type selector, and (d) a CRS cell. Details are provided in Chapter 22.

For high-density integration, a passive array with a cross-point architecture with $4F^2$ area is preferred. The simplest way is to connect the word and bit lines at each node by a ReRAM cell. A *sneak path problem* with this simple cross-point architecture may arise (Figure 1.5): if the cell to be read out happens to be in HRS with surrounding cells in LRS, the reading current can easily flow through the surrounding cells in the LRS, and, thus, an LRS data will be mistakenly read out. For further discussions about the sneak path problem and the interconnect effect on the cross-point architecture design, one can refer to Chapter 22 and Refs. [100–103]. The conclusions from these works indicate that in order to magnify the write/read margin of the memory array, larger ratio of the ReRAM resistance over the interconnect resistance is desired, thus increasing the LRS resistance is helpful. In order to further suppress the sneak paths, a cell *selection element* (also: *select device, selector*) with strong *I-V* nonlinearity should be added at each node. The cell selection element can be a diode for unipolar switching ReRAM as the 1 diode–1 resistor (1D1R) architecture, or a bidirectional selector (with a strong nonlinear *I-V* characteristics in both polarities, that is, a varistor or Z diode) for bipolar switching ReRAM as the 1 selector-1 resistor (1S1R) architecture. The diode or the selector effectively cuts off the leakage current paths at reversed bias or low bias, thus the interference between neighboring cells is prevented.

Here we briefly discuss the cell selection element materials. For unipolar switching, a p–n diode is the most common device for the cell selection element. Although a high-performance p–n diode is easily fabricated with the current epitaxial silicon technology for the planar device structures, it is not feasible to implement epitaxial silicon-based p–n diode with the ReRAM array at the back end of line (BEOL) in the fabrication process because it is difficult to grow epitaxial silicon on a metal layer and high processing temperatures are required.

On the other hand, amorphous silicon allows for lower processing temperatures. But it does not meet the requirement for current density for memory cell programming. Therefore, new materials need to be explored for the cell selection element, which should both allow for low processing temperatures and provide high current drivability. Compared with silicon p–n diode, oxide p–n diode is more attractive. It offers better flexibility in processing technology because it can be fabricated during BEOL processing even at room temperature. If the oxide material is oxygen-deficient with sufficient amount of oxygen vacancies (or cation interstitials), it is n-type; if the oxide material is metal-deficient with sufficient amount of metal vacancies, it is p-type. Thus, a combination of p-type oxide and n-type oxide essentially forms a p–n diode. Recently, several kinds of oxide p–n diodes [104], such as p-NiO/n-TiO₂, p-NiO/n-ZnO, p-NiO/n-InZnO, p-CuO/n-InZnO, have been demonstrated and stacked with Pt/NiO/Pt ReRAM in series, among which p-CuO/n-InZnO is regarded as the best candidate in terms of current drivability. Besides the p–n oxide diode, through oxide/electrode or oxide/oxide interface band engineering, unidirectional rectification of *I-V* for unipolar switching or bidirectional rectification of nonlinear I-V for bipolar switching can also be achieved. For examples, Schottky diode with Pt/TiO₂/Ti/Pt stack has been integrated with Pt/TiO₂/Pt unipolar ReRAM [105], Ni/TiO₂/Ni bidirectional selector has been integrated with Ni/HfO_x/Pt bipolar ReRAM [106], and Pt/TaO_x/TiO₂/TaO_x/Pt bidirectional selector has been integrated with Cu/HfO_x/Pt bipolar ReRAM [107]. A MIM tunnel diode based on atomic layer deposition (ALD)-processed Ta₂O₅ has been used for a bidirectional selector [108].

In addition, some materials with metal–insulator transition (MIT) such as VO₂, NbO₂, and Ti₂O₃ can be used for selectors as well. Unlike ReRAM devices, the resistive switching behavior in MIT materials is not bistable, and it is referred to as threshold switching. The threshold switching is not polarity dependent; thus, it is suitable for bipolar switching ReRAM. The sharp increase in current by orders of magnitude when reaching the threshold voltage shows an ideal behavior of a selector device. Pt/VO₂/Pt selector has been integrated with NiO unipolar ReRAM [109] and ZrO_x/HfO_x bipolar ReRAM [110]. However, VO₂ has a transition temperature of around 67 °C, beyond which the threshold switching behavior disappears [111], which is a major drawback for practical applications. Alternatively, NbO₂ has a transition temperature around 800 °C; thus, it is more attractive due to its thermal stability. W/NbO₂/Pt selector has been integrated with NbO_x [112] and a Pt/TaO_x/TiO_{2-x}/Pt with TiO_{2-x} [113] in bipolar ReRAM, respectively. The primary concern with this type of selector is operation energy since phase transitions in MIT materials are generally believed to be driven by Joule heating [114]. If the ReRAM device is in its LRS state, even a reading operation needs to turn on the negative differential resistance (NDR) device once. In addition, the endurance of NDR devices might also be a potential issue [115].

Alternatively, Cu ion motion in the Cu-containing mixed ionic electronic conduction (MIEC) materials also shows a good selector behavior for bipolar switching ReRAM [116].

Table 1.2 Representatives of selection devices for ReRAM in the literature.

Type	Stack	Voltage range (V)	Current drivability (A cm ⁻²)	Ratio
p-n diode	p-CuO/n-InZnO [104]	-2 to +2	3×10^4	3×10^4
Schottky diode	Pt/TiO ₂ /Ti/Pt [105]	-2 to +2	3×10^5	2.4×10^6
Bidirectional selector	Ni/TiO ₂ /Ni [119]	-4 to +4	10^5	10^3
Bidirectional selector	Pt/TaO _x /TiO ₂ / TaO _x /Pt [107]	-2.5 to +2.5	3.2×10^7	10^4
MIT	Pt/VO ₂ /Pt [110]	-0.5 to +0.5	6×10^6	50
MIT	W/NbO ₂ /Pt [112]	-1.5 to +1.5	2×10^6	50
Bidirectional selector	MIEC [116]	-0.5 to +0.5	$5 \times 10^{4a})$	3×10^3

The voltage range states the maximum voltage where the current density is measured. For diodes, the ratio is defined as the forward/reverse current at the maximum voltage. For bidirectional and MIT selectors, the ratio is defined as the current at maximum voltage over half of the maximum voltage.

a) MIEC has a higher current drivability $\sim 50 \times 10^6$ A/cm² at pulse mode (1.6 V).

Last but not least, two ReRAM devices can also be stacked antiferentially (head-to-head) to form a CRS [9]. Under a switching voltage pulse, these two devices always have opposite voltage polarities and switch in opposite directions, thus limiting the current through the device pair since there is always one device in the OFF state in the pair. This CRS device concept can also be realized in a single ReRAM device with two instead of one switchable interface [117, 118]. The disadvantage of this scheme is that it becomes difficult to operate these switches as multilevel cells or analog devices. A destructive read is also much more in demand for device endurance in this scheme.

Table 1.2 compares several aforementioned selection element candidates in aspects of voltage range, current drivability, rectifying ratio. Although the reported diodes or selectors show a promising current drivability and rectifying ratio, many of them use Pt in the stacks, which is not CMOS fabrication process-friendly.

A detailed description of all relevant selector device concepts is given in Chapter 22.

1.2.3

3-D Technology

The primary target of ReRAM is to replace flash technology, as ReRAM has significant advantages in terms of speed and energy consumption and as flash is facing scaling limitations beyond 10 nm technology node [120]. For mass-storage application, a key challenge for ReRAM is to improve the integration density in terms of cost per bit (bit cost), so it can compete with the multibit storage NAND flash.

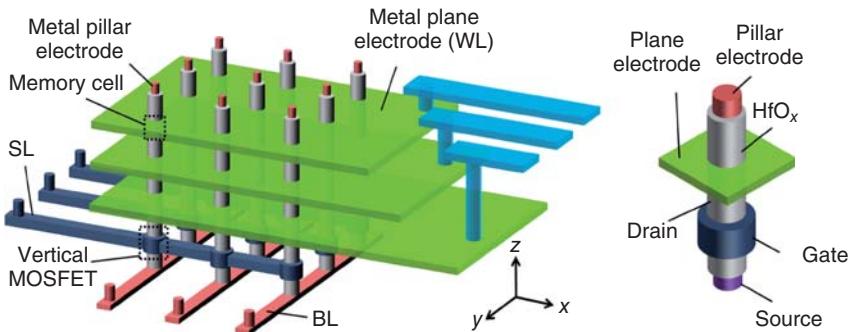


Figure 1.6 Schematic of the proposed 3D cross-point architecture using the vertical ReRAM cell. The vertical ReRAM cells are formed at the intersections of each pillar electrode and each plane electrode: the resistive switching oxide layer surrounds the pillar electrode and is also in contact with the plane electrode. To enable

the random access of each memory cell, three-dimensional decoding is needed through WL (decoding in z-direction), BL (decoding in y-direction), and SL of the gate of the vertical MOSFET (decoding in x-direction). WL, BL, and SL denote the Word Line, Bit Line, and Select Line, respectively. (Reprinted with permission from Ref. [130].)

State-of-the-art NAND flash has been scaled down to sub-20 nm regime, and the 3D stackable NAND flash is emerging [121–123]. To achieve similar device density to that of the 3D NAND flash, a technology path toward the 3D stackable ReRAM is required. Two 3D integration approaches are available for ReRAM: one is the planar ReRAM stacked layer by layer [124]; however, it does not save lithography steps or masks because in each layer, a lithography step is needed to pattern the features, and, therefore, the bit cost remains high. The other one is the vertical ReRAM sandwiched between the pillar electrodes and multilayer plane electrodes [125–127], which requires only one critical lithography step or mask; thus, it is a more promising approach for reducing the bit cost. In 2012, Yu *et al.*, proposed a cost-effective technology path toward 3D integration (Figure 1.6) [128–130]. As a proof of concept work, a two-layer vertical HfO_x ReRAM structure was fabricated and characterized [128]. Programming schemes for random access in the 3D array were proposed and the 3D array operation was experimentally demonstrated [129]. Furthermore, the scaling trend of the 3D vertical ReRAM array was analyzed by experiments and 3D circuit simulation. More details and alternative approaches will be described in Chapter 20.

1.2.4

Memory Hierarchy

The pyramidal hierarchical structure of memories and the storage devices is shown in Figure 1.7 [96, 131]. The vertical arrows on the left and right side of the figure indicate the capacity (bit), access time (ns), and also the retention time of typical memories, respectively. Within this hierarchy, storage class memory

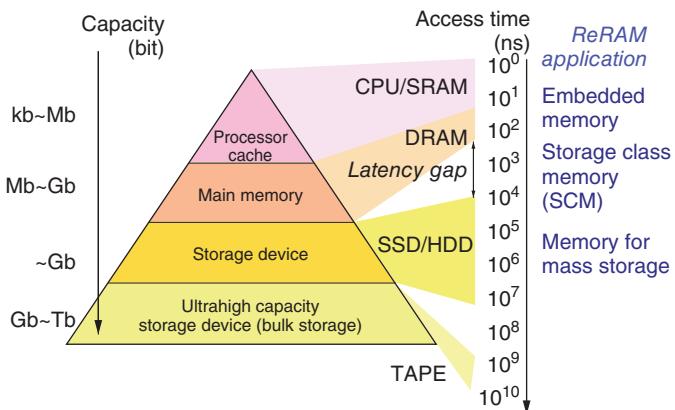


Figure 1.7 Pyramidal hierarchical structure of memories and the storage devices.

(SCM) bridges the latency gap between cache, main memories, and storage devices, such as hard disk drive (HDD) and solid-state drive (SSD). As mentioned earlier, ReRAM possesses two primary target application areas. One is the embedded application, such as microcontroller units with embedded ReRAMs [132, 133]. Another is the mass-storage application, in particular, SCM. Owing to the high possibility of the 3D integration, ReRAM is a candidate to replace the NAND flash technology [134]. As shown in Figure 1.7, the access time of DRAMs, which are used as the main memory, is 2 orders of magnitude faster than that of SSD, in which the NAND flash memory is utilized as a storage medium. Of course, architectural changes are required to utilize the advantages of SSD effectively, in terms of the low-power performance and durability. If circuit design adaption is realized, the use of SCM in mobile terminals improves the speed of data transfer within a network, enabling a marked reduction in the power consumption of the system as a whole [135, 136]. In summary, since the ReRAM technology has measured up the requirements with regard to the capacity and latency, ReRAM is the most promising candidate for SCM and beyond.

1.3 Logic Circuits

There are three main routes along which ReRAM devices are considered for application in logic operations in electronic circuits: (i) as nonvolatile switches in FPGA circuits, (ii) as logic gates, and (iii) as elements in neuromorphic circuits. These applications exploit the extreme scalability the nonvolatility of the ReRAM and also the fact that many ReRAM systems can be integrated in the so-called BEOL in the integration process, after any CMOS part is fabricated.

An *FPGA* consists of a logic-computing circuit where logic units are connected together by switch units, which can be reprogrammed to reconfigure the

connections and the functionality of the FPGA. In a conventional FPGA, logic cells are connected by volatile switches consisting of SRAM circuits. An SRAM circuit occupies a relatively large area of more than $100F^2$. To reduce the switch area in the FPGA, SRAM can be replaced by ReRAM devices with a minimum area of $4F^2$ [137]. In addition, ReRAM devices can be fabricated in the BEOL and, thus, can be stacked on top of the logic CMOS plane to further optimize the circuit area.

An early example of ReRAM-based FPGA is the so-called CMOL proposal, where CMOS is combined with switches that consist of molecular devices (hence the name) or any other type of ReRAM devices [138]. The crossbar array of reconfigurable nanoswitches is fabricated on top of the CMOS logic, displaying a tilted orientation with respect to the CMOS array. This peculiar architecture allows for the exact matching of the switch array size with the logic array size, to take advantage of the smaller ReRAM pitch F in order to achieve extremely high density in the switch array combined with a more relaxed pitch F_{CMOS} of the CMOS circuit. ReRAM–CMOS hybrid integrated circuits have been fabricated based on this concept. For example, ReRAM-based FPGA circuits have adopted a 1T2R architecture of the switch [139] or a complementary atom switch (CAS) structure, where two ReRAM switches are connected in parallel [140]. For application in FPGA circuits, the requirements of the ReRAM device are different from those expected for memory applications. First, the writing speed of ReRAM must be close to that of CMOS-based SRAM, which is capable of sub-nanosecond switching. Second, ReRAM resistance must be low to allow for low-voltage operation and for low RC parasitics. Finally, reliability with respect to tolerance against low-voltage disturbs must be ensured to avoid unwanted reprogramming during operation of the FPGA logic. More details are provided in Chapter 24.

Logic gates as fundamental elements for sequential binary computing have been realized by resistive switches. For instance, the Boolean function *material implication* (IMP) was implemented as a gate based on two memristive elements as shown in Figure 1.8. Here, switches P and Q are connected in parallel and initially contain the input logic states p and q . In the computational stage, pulse voltages of amplitude V_{set} and V_{cond} are applied to Q and P, respectively, where V_{set} is just above the threshold for SET transition, whereas V_{cond} is just below the same threshold. If p is 0, namely P has a high resistance, then the whole voltage V_{set} drops across Q, which thus unconditionally switches to the set state. If p is 1, namely P has a low resistance, then the voltage drop across q is just $|V_{\text{set}} - V_{\text{cond}}| \approx 0$, which cannot induce any change in q . The resulting truth table in Figure 1.8 demonstrates IMP functionality. All other logic operations can be achieved through iteration of IMP over successive cycles [141, 142], thus making it possible to achieve logic completeness. This concept has been extended to CRS cells [143, 144]. Switch-based logic gates can allow for a huge reduction of circuit area, thanks to the small size of the switch in a crossbar array, and a huge reduction in energy consumption, thanks to the nonvolatile behavior of resistance switches, which enables the normally off operation of digital circuits [145].

Another emerging application is using ReRAM cells as *artificial synaptic device* for the hardware implementation of *neuromorphic circuits* mimicking the

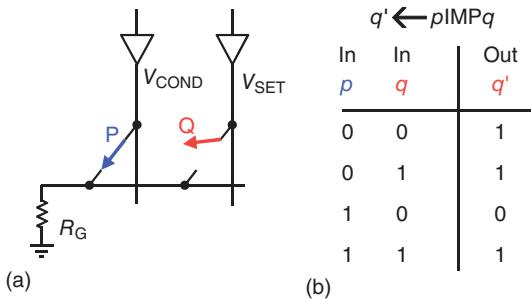


Figure 1.8 Schematic illustration of the IMP gate implemented through two ReRAM switches (a) and the corresponding truth table (b). Memristive elements P and Q

contain the input logic states p and q , respectively, then Q can conditionally switch to the output state q' . (Reprinted with permission from Ref. [141].)

biological neural operations and aiming at cognitive functionality. Neuromorphic circuits elaborate spiking neural signals through neural networks where CMOS neurons are connected by artificial synapses.

Biological neurons are formed by a cell body (soma), connected to other neurons through dendrites, which provide the input stimuli for neuron elaboration, and an axon, which transmits the output signal or “fire.” An output axon and an input dendrite are separated by an interfacial gap called synapse. Each synapse has a characteristic conductivity, or weight, describing the strength of the coupling between presynaptic and postsynaptic neurons. Therefore, neuron computation relies on both the input signals coming by the synapses and their corresponding weights. Traditionally, electronic neural networks have been developed by integration of CMOS circuits, playing the role of the neurons, with artificial synapses consisting of Si-based floating gate cells [146] or SRAM circuits [147]. Instead of the pure CMOS realization, ReRAM may take over the key functionality for artificial synapses (Figure 1.9), which is the weighted resistive connection of two neurons. ReRAM devices can be considered as binary, multilevel, or even analog switching devices, that is, the synapse conductance is changed continuously between low and high to configure the neural network. ReRAM is highly suitable for such application, since the resistance can be changed from the set state to the reset state and vice versa through a proper choice of the voltage and current. In fact, the resistance can be controlled by the maximum current during the set operation and by the maximum voltage during the reset transition [148]. Instead of voltage and current, however, the time delay between spikes is the parameter dictating the change in the synapse conductance in the biological system through the so-called spike-timing-dependent plasticity (STDP). Besides their functionality, another advantage of ReRAMs is the fact that they are two-terminal devices, that they can be strongly scaled in their size, and that they can be integrated on top of CMOS circuits in the BEOL process steps. It should further be noted that the requirements on the performance of ReRAM cell for neuromorphic

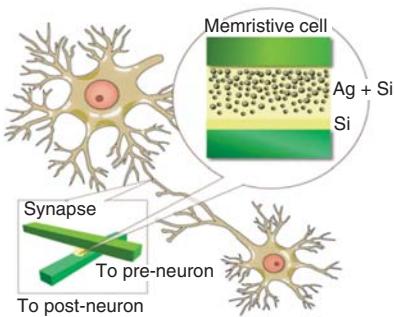


Figure 1.9 Schematic illustration of the natural (biological) and ReRAM-based artificial synapses. (Reprinted with permission from Ref. [95].)

computing such as the operation speed and the retention time may be significantly relaxed as compared to memory and FPGA applications. In addition to spiking concepts, there are also level-based concepts utilizing threshold gates, realizing, for example, content addressable memories [149] or single-layer perceptrons [150]. A common fingerprint of ReRAM-based neuromorphic systems is that they comprise hybrid analog–digital electronic circuits. Details about the different concepts of how to use ReRAM in neuromorphic computing are provided in Chapter 25.

The results published suggest that ReRAM might be able to emulate the functionalities of synapses and even entire neurons and become an essential ingredient in neuromorphic circuits, thus enlarging the spectrum of possible applications in the computing domain.

1.4 Prospects and Challenges

As described, due to their attractive properties, ReRAM devices are promising for numerous applications, including memory and storage as well as digital, analog, and neuromorphic computing. Accordingly, these applications set a variety of requirements for the devices. The device performance requirements for memory and storage applications are more demanding than, for instance, for neuromorphic applications with respect to speed, retention time, and device density. In the following, the promises of the ReRAM devices are briefly summarized. Then the focus is on the challenges and possible solutions to these challenges.

ReRAMs offer a great scalability, ultrafast switching speed, nonvolatility, large HRS/LRS window, analog resistance change, nondestructive reading, simple structures with common materials, 3D stackability, and great CMOS compatibility and manufacturability. Besides these great prospects, ReRAM devices face a number of challenges. These challenges are application dependent. The most critical issues include device isolation in crossbar arrays, device variability, and reliability.

ReRAMs face a challenge with respect to an intrinsic cycle-to-cycle variability, which is typically relatively large compared to other nonvolatile memory technologies, owing to the filamentary nature of the active region in the ReRAM and the stochasticity of the filament buildup on the atomistic scale. As a consequence, the effective memory window due to wide distributions of the LRS and HRS resistances may shrink below critical values. Of course, this strongly depends on the type of ReRAM cell, the material system, and the cell design. Another important reliability requirement is nonvolatility (retention). In general, higher resistive filaments (desired for low current operation) tend to show a reduced retention. Presumably, this is due to atomic-scale sized filaments and processes on the atomic scale affecting the filament properties, such as interdiffusion, phase formation, and interface reactions. More insights into the different reliability mechanisms are provided in Chapter 21.

In summary, based on the progress in the microscopic understanding of the electroforming and switching process, solutions to the challenges will come from a combination of materials engineering, device structure optimization, as well as innovations in addressing/readout circuitry and programming algorithm.

Acknowledgments

The authors are indebted to Stephan Menzel (FZ Jülich, Germany) for carefully reading the manuscript and for helpful discussions, to Eike Linn (RWTH Aachen, Germany) for in-depth discussions on neuromorphic computing, and to Yun-Leei Chiou (Temple Terrace, USA) for firsthand information on the history of ReRAM in the 1960s. RW would like to gratefully acknowledge the support from the Deutsche Forschungsgemeinschaft through the SFB 917 and the Leibniz Award 2014.

References

- Waser, R. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 603–620.
- Chua, L.O. and Kang, S.M. (1976) *Proc. IEEE*, **64** (2), 209–223.
- Akarvardar, K. and Philip Wong, H.-S. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 375–387.
- Dieny, B., Sousa, R., Nozieres, J.-P., Redon, O., and Prejbeanu, I.L. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 655–668.
- Zhirnov, V. and Mikolajick, T. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 621–634.
- Raoux, S. and Wuttig, M. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 669–682.
- Waser, R., Bruchhaus, R., and Menzel, S. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser),

- Wiley-VCH Verlag GmbH,
pp. 683–710.
8. Menzel, S., Dittmann, R., and Waser, R. to be published in *Adv. Physics* (2015).
 9. Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010) *Nat. Mater.*, **9** (5), 403–406.
 10. Nardi, F., Balatti, S., Larentis, S., Gilmer, D.C., and Ielmini, D. (2013) *IEEE Trans. Electron Devices*, **60** (1), 70–77.
 11. Lee, M.-J. *et al.* (2009) *Nano Lett.*, **9**, 1476–1481.
 12. Yang, J.J. *et al.* (2009) *Nanotechnology*, **20** (21), 215201.
 13. Muenstermann, R., Menke, T., Dittmann, R., and Waser, R. (2010) *Adv. Mater.*, **22** (43), 4819–4822.
 14. Kim, K., Jeong, D.S., and Hwang, C.S. (2011) *Nanotechnology*, **22** (25), 254002.
 15. Liu, Q. *et al.* (2012) *Adv. Mater.*, **24** (14), 1844–1849.
 16. Maehne, H. *et al.* (2012) *Solid-State Electron.*, **72**, 73–77.
 17. Yao, J., Zhong, L., Natelson, D., and Tour, J. (2012) *Sci. Rep.*, **2** (00242), 1–5.
 18. Chen, J. *et al.* (2013) *Nano Lett.*, **13**, 3671–3677.
 19. Celano, U., Chen, Y.Y., Wouters, D.J., Groeseneken, G., Jurczak, M., and Vandervorst, W. (2013) *Appl. Phys. Lett.*, **102** (12), 121602/1–121602/3.
 20. Park, G.-S. *et al.* (2013) *Nat. Commun.*, **4**, 2382/1–2382/9.
 21. Calka, P. *et al.* (2013) *Nanotechnology*, **24** (8), 85706/1–85706/9.
 22. Hsu, C.-W. *et al.* (2013) Proceedings of the 2013 VLSI Symposium on Technology.
 23. Chen, Y.Y. *et al.* (2013) Electron Devices Meeting (IEDM), 2013 IEEE International.
 24. Ninomiya, T., Wei, Z., Muraoka, S., Yasuhara, R., Katayama, K., and Takagi, T. (2013) *IEEE Trans. Electron Devices*, **60**, 1384–1389.
 25. Kamaladasa, R.J. *et al.* (2015) *Microsc. Microanal.*, **21**, 140–153.
 26. Zhirnov, V.V. *et al.* (2010) *Proc. IEEE*, **98** (12), 2185–2200.
 27. Yang, J.J., Strukov, D.B., and Stewart, D.R. (2013) *Nat. Nanotechnol.*, **8** (1), 13–24.
 28. Jeong, D.S., Schroeder, H., and Waser, R. (2007) *Electrochim. Solid-State Lett.*, **10** (8), G51–G53.
 29. Nardi, F., Balatti, S., Larentis, S., and Ielmini, D. (2011) 2011 IEEE International Electron Devices Meeting (IEDM 2011), pp. 31.1.1–31.1.4
 30. Aoki, Y. *et al.* (2014) *Nat. Mater.*, **5**, 3473/1–3473/9.
 31. Kamaladasa, R.J. *et al.* (2013) *J. Appl. Phys.*, **113** (23), 234510/1–234510/7.
 32. Baek, I.G. *et al.* (2004) Internat. Electron Device Meeting IEDM Technical Digest, pp. 587–590.
 33. Choi, B.J. *et al.* (2005) *J. Appl. Phys.*, **98** (3), –33715-1-10.
 34. Russo, U., Ielmini, D., Cagli, C., and Lacaita, A.L. (2009) *IEEE Trans. Electron Devices*, **56** (2), 186–192.
 35. Ielmini, D., Bruchhaus, R., and Waser, R. (2011) *Phase Transitions*, **84** (7), 570–602.
 36. Baek, I.G. *et al.* (2005) IEEE International Electron Devices Meeting (IEDM) 2005, Technical Digest, pp. 769–772.
 37. Hickmott, T.W. (1962) *J. Appl. Phys.*, **33**, 2669–2682.
 38. Lampert, M.A. (1956) *Phys. Rev.*, **103** (6), 1648–1656.
 39. Lampert, M.A. (1962) *Phys. Rev.*, **125** (1), 126–141.
 40. Gibbons, J.F. and Beadle, W.E. (1964) *Solid-State Electron.*, **7** (11), 785–790.
 41. Hiatt, W.R. and Hickmott, T.W. (1965) *Appl. Phys. Lett.*, **6**, 106–108.
 42. Chopra, K.L. (1965) *J. Appl. Phys.*, **36**, 184–187.
 43. Simmons, J.G. and Verderber, R.R. (1967) *Radio Electron. Eng.*, **34**, 81–89.
 44. Argall, F. (1968) *Solid-State Electron.*, **11** (5), 535–541.
 45. Hickmott, T.W. (1969) *J. Vac. Sci. Technol.*, **6** (5), 828–833.
 46. Hovel, H.J. and Urgell, J.J. (1971) *J. Appl. Phys.*, **42**, 5076–5083.
 47. Hirose, Y. and Hirose, H. (1976) *J. Appl. Phys.*, **47** (6), 2767–2772.
 48. Kahng, D. and Sze, S.M. (1967) *Bell Syst. Tech. J.*, **46**, 1288–1295.
 49. Dearnaley, G., Stoneham, A.M., and Morgan, D.V. (1970) *Rep. Prog. Phys.*, **33** (3), 1129–1191.

50. Oxley, D.P. (1977) *Electrocomponent Sci. Technol.*, **3** (4), 217–224.
51. Pagnia, H. and Sotnik, N. (1988) *Phys. Status Solidi A*, **108** (1), 11–65.
52. Asamitsu, A., Tomioka, Y., Kuwahara, H., and Tokura, Y. (1997) *Nature*, **388** (6637), 50–52.
53. Zhuang, W.W. et al. (2002) IEEE Internat. Electron Device Meeting, Technical Digest IEDM, p. 193.
54. Beck, A., Bednorz, J.G., Gerber, C., Rossel, C., and Widmer, D. (2000) *Appl. Phys. Lett.*, **77** (1), 139–141.
55. Kozicki, M.N., Yun, M., Hilt, L., and Singh, A. (1999) *Electrochem. Soc. Proceedings of Solid-State Ionic Devices* **99**, Seattle, WA, May 02–07, 1999.
56. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2001) *Riken Rev.*, **37**, 7–8.
57. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005) *Nature*, **433** (6), 47–50.
58. Szot, K., Speier, W., Bihlmayer, G., and Waser, R. (2006) *Nat. Mater.*, **5** (4), 312–320.
59. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21** (25–26), 2632–2663.
60. Waser, R. and Aono, M. (2007) *Nat. Mater.*, **6** (11), 833–840.
61. Janousch, M., Meijer, G.I., Staub, U., Delley, B., Karg, S.F., and Andreasson, B.P. (2007) *Adv. Mater.*, **19**, 2232–2235.
62. Wei, Z. et al. (2008) IEEE Internat. Electron Device Meeting (IEDM), Technical Digest.
63. Kwon, D.-H. et al. (2010) *Nat. Nanotechnol.*, **5** (2), 148–153.
64. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) *Nature*, **453** (7191), 80–83.
65. Chua, L.O. (1971) *IEEE Trans. Circuit Theory*, **CT-18** (5), 507–519.
66. Liu, T.-Y. et al. (2013) Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC 2013).
67. Fackenthal, R. et al. (2014) 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC).
68. Esaki, L., Laibowitz, R.B., and Stiles, P.J. (1971) *IBM Tech. Discl. Bull.*, **13** (8), 2161.
69. Blom, P.W.M., Wolf, R.M., Cillessen, J.F.M., and Krijn, M.P.C.M. (1994) *Phys. Rev. Lett.*, **73** (15), 2107–2110.
70. Ovshinsky, S.R. (1966) Symmetrical current controlling device, US Patent 3271591.
71. Ovshinsky, S.R. (1968) *Phys. Rev. Lett.*, **21** (20), 1450–1453.
72. Lee, T. and Chen, Y. (2012) *MRS Bull.*, **37**, 144–149.
73. Potember, R.S., Poehler, T.O., and Cowan, D.O. (1979) *Appl. Phys. Lett.*, **34** (6), 405–407.
74. Kever, T., Boettger, U., Schindler, C., and Waser, R. (2007) *Appl. Phys. Lett.*, **91** (8), 083506-1–083506-3.
75. Bandyopadhyay, A. and Pal, A.J. (2004) *Appl. Phys. Lett.*, **84** (6), 999–1001.
76. Karthaeuser, S. et al. (2006) *J. Appl. Phys.*, **100** (9), 94504-1–94504-6.
77. Choi, B.J., Chen, A.B.K., Yang, X., and Chen, I. (2011) *Adv. Mater.*, **23** (33), 3847–3852.
78. Ziegler, M., Oberlaender, M., Schroeder, D., Krautschneider, W.H., and Kohlstedt, H. (2012) *Appl. Phys. Lett.*, **101**, 263504.
79. Riggert, C., Ziegler, M., Schroeder, D., Krautschneider, W.H., and Kohlstedt, H. (2014) *Semicond. Sci. Technol.*, **29**, 1040011.
80. Stoliar, P., Rozenberg, M., Janod, E., Corraze, B., Tranchant, J., and Cario, L. (2014) *Phys. Rev. B: Condens. Matter*, **90** (4), 45146/1(R).
81. Xue, K.H., de Araujo, C.A.P., Celinska, J., and McWilliams, C. (2011) *J. Appl. Phys.*, **109**, 091602. 21st International Symposium on Integrated Ferroelectrics and Functionalities, Colorado Springs, CO.
82. Govoreanu, B. et al. (2011) 2011 IEEE International Electron Devices Meeting – IEDM '11, IEDM Technical Digest.
83. Zhirnov, V.V., Meade, R., Cavin, R.K., and Sandhu, G. (2011) *Nanotechnology*, **22** (25), 254027/1–254027/21.
84. Liang, J., Yeh, S., Wong, S.S., and Wong, H.-S.P. (2012) 2012 4th IEEE International Memory Workshop (IMW), p. 4.
85. Chen, A. (2013) *IEEE Trans. Electron Devices*, **60** (4), 1318–1326.

86. Torrezan, A.C., Strachan, J.P., Medeiros-Ribeiro, G., and Williams, R.S. (2011) *Nanotechnology*, **22**, 485203/1–485203/7.
87. Strukov, D.B. and Williams, R.S. (2009) *Appl. Phys. A: Mater. Sci. Process.*, **94** (3), 515–519.
88. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011) *Adv. Funct. Mater.*, **21** (23), 4487–4492.
89. Schroeder, H., Zhirnov, V.V., Cavin, R.K., and Waser, R. (2010) *J. Appl. Phys.*, **107** (5), 054517/1–054517/8.
90. Lee, M.-J. *et al.* (2011) *Nat. Mater.*, **10** (8), 625–630.
91. Miao, F. *et al.* (2011) *Adv. Mater.*, **23**, 5633–5640.
92. Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A.L., and Kozicki, M.N. (2009) *IEEE Trans. Electron Devices*, **56** (5), 1040–1047.
93. Chien, W.-C. *et al.* (2011) 2011 IEEE International Electron Devices Meeting – IEDM '11.
94. Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J.K., and Aono, M. (2011) *Nat. Mater.*, **10**, 591–595.
95. Jo, S.H., Chang, T., Ebong, I., Bhadviya, B.B., Mazumder, P., and Lu, W. (2010) *Nano Lett.*, **10** (4), 1297–1301.
96. The International Technology Roadmap for Semiconductors (ITRS) (2013).
97. Chiu, P.-F. *et al.* (2012) *IEEE J. Solid-State Circuits*, **47** (6), 1483–1496.
98. Lee, H.Y. *et al.* (2008) IEEE International Electron Devices Meeting 2008, Technical Digest, San Francisco, CA.
99. Panasonic (2013) press release.
100. Liang, J. and Wong, H.-S.P. (2010) *IEEE Trans. Electron Devices*, **57** (10), 2531–2538.
101. Flocke, A., Noll, T.G., Kugeler, C., Nauenheim, C., and Waser, R. (2008) Proceedings of the 8th IEEE Conference on Nanotechnology, pp. 319–322.
102. Liang, J., Yeh, S., Wong, S.S., and Wong, H.-S.P. (2013) *ACM J. Emerging Technol. Comput. Syst.*, **9** (1), 9:1–9:14.
103. Deng, Y. *et al.* (2013) *IEEE Trans. Electron Devices*, **60** (2), 719–726.
104. Lee, M.-J. *et al.* (2007) 2007 IEEE International Electron Devices Meeting – IEDM '07, IEDM Technical Digest.
105. Kim, G. *et al.* (2013) *Adv. Funct. Mater.*, **23** (11), 1440–1449.
106. Huang, J.-J., Tseng, Y.-M., Hsu, C.-W., and Hou, T.-H. (2011) *IEEE Electron Device Lett.*, **99**, 1–3.
107. Lee, W. *et al.* (2012) 2012 IEEE Symposium on VLSI Technology, pp. 37–38.
108. Govoreanu, B., Adelmann, C., Redolfi, A., Zhang, L., Clima, S., and Jurczak, M. (2014) *IEEE Electron Device Lett.*, **35** (1), 63–65.
109. Lee, M.-J. *et al.* (2007) *Adv. Mater.*, **19**, 3919–3923.
110. Son, M. *et al.* (2011) *IEEE Electron Device Lett.*, **32** (11), 1579–1581.
111. Ko, C. and Ramanathan, S. (2008) *Appl. Phys. Lett.*, **93** (25), 252101.
112. Kim, S. *et al.* (2012) 2012 IEEE Symposium on VLSI Technology, pp. 155–156.
113. Yang, J.J. *et al.* (2012) *Appl. Phys. Lett.*, **100** (11), 113501/1.
114. Pickett, M.D., Borghetti, J., Yang, J.J., Medeiros-Ribeiro, G., and Williams, R.S. (2011) *Adv. Mater.*, **23** (15), 1730.
115. Cha, E., Woo, J., Lee, D., Lee, S., and Hwang, H. (2014) 2014 IEEE International Symposium on Circuits and Systems, ISCAS 2014, Melbourne, VIC, Australia.
116. Gopalakrishnan, K. *et al.* (2010) Symposium on VLSI Technology (VLSIT), Honolulu, HI.
117. Yang, J.J., Borghetti, J., Murphy, D., Stewart, D.R., and Williams, R.S. (2009) *Adv. Mater.*, **21** (37), 3754–3758.
118. Shima, H., Zhong, N., and Akinaga, H. (2009) *Appl. Phys. Lett.*, **94** (8), 082905-1–082905-3.
119. Huang, J.-J., Tseng, Y.-M., Luo, W.-C., Hsu, C.-W., and Hou, T.-H. (2011) 2011 IEEE International Electron Devices Meeting – IEDM '11.
120. Goda, A., Parat, K., and January, M.T. (2012) IEEE International Electron Devices Meeting, pp. 13–16.
121. Tanaka, H. *et al.* (2007) IEEE Symposium on VLSI Technology, pp. 14–15.
122. Kim, W. *et al.* (2009) IEEE Symposium on VLSI Technology, pp. 188–189.
123. Jang, J. *et al.* (2009) IEEE Symposium on VLSI Technology, pp. 192–193.

124. Lee, M.J. *et al.* (2009) *Adv. Funct. Mater.*, **19** (10), 1587–1593.
125. Yoon, H.S. *et al.* (2009) 2009 Symposium on VLSI Technology, Digest of Technical Papers.
126. Baek, I. *et al.* (2011) Technical Digest - International Electron Devices Meeting, IEDM.
127. Chien, W. *et al.* (2012) 2012 IEEE Symposium on VLSI Technology, pp. 153–154.
128. Chen, H.-Y., Yu, S., Gao, B., Huang, P., Kang, J., and Wong, H.-S.P. (2012) 2012 IEEE International Electron Devices Meeting (IEDM 2012), pp. 20.7. 1–20.7. 4.
129. Yu, S. *et al.* (2013) 2013 Symposium on VLSI Technology, pp. T158–T159.
130. Yu, S., Chen, H.-Y., Gao, B., Kang, J., and Wong, H.-S.P. (2013) *Acs Nano*, **7** (3), 2320–2325.
131. Akinaga, H. (2013) *Jpn. J. Appl. Phys.*, **52**, 100001–1–100001–12.
132. Yamamoto, S., Shuto, Y., and Sugahara, S. (2010) *Jpn. J. Appl. Phys.*, **49**, 040209–1–040209–3.
133. Kawahara, A. *et al.* (2012) International Solid-State Circuit Conference (ISSCC), pp. 220–221.
134. Harari, E. (2012) International Solid-State Circuit Conference (ISSCC), pp. 10–15.
135. Fukuda, M., Higuchi, K., and Takeuchi, K. (2011) *Jpn. J. Appl. Phys.*, **50**, 04DE09–1–04DE09–7.
136. Takeuchi, K., Hatanaka, T., and Tanakamaru, S. (2012) *IEICE Electron. Express*, **9** (8), 779–794.
137. Kaeriyama, S. *et al.* (2005) *IEEE J. Solid-State Circuits*, **40** (1), 168–176.
138. Strukov, D.B. and Likharev, K.K. (2005) *Nanotechnology*, **16** (6), 888–900.
139. Liauw, Y.Y., Zhang, Z., Kim, W., Gamal, A.E., and Wong, S.S. (2012) IEEE International Solid-State Circuits Conference, pp. 406–408.
140. Miyamura, M. *et al.* (2012) IEDM Technical Digest, p. 247.
141. Borghetti, J., Snider, G.S., Kuekes, P.J., Yang, J.J., Stewart, D.R., and Williams, R.S. (2010) *Nature*, **464** (7290), 873–876.
142. Linn, E., Rosezin, R., Tappertzhofen, S., Böttger, U., and Waser, R. (2012) *Nanotechnology*, **23**, 305205.
143. You, T. *et al.* (2014) *Adv. Funct. Mater.*, **24**, 3357–3365.
144. Siemon, A. *et al.* (2015) *Adv. Funct. Mater.* (DOI: 10.1002/adfm.201500865)
145. Nikonov, D.E. and Young, I.A. (2012) *Nanotechnology*, **23**, 305205.
146. Diorio, C.J., Hasler, P.E., Mead, C.A., and Minch, B.A. (1996) *IEEE Trans. Device Mater. Reliab.*, **43**, 1972.
147. Rajendran, B. *et al.* (2013) *IEEE Trans. Electron Devices*, **60**, 246.
148. Yu, S., Wu, Y., Jeyasingh, R., Kuzum, D., and Wong, H.P. (2011) *IEEE Trans. Electron Devices*, **58**, 2729–2737.
149. Nielen, L., Siemon, A., Tappertzhofen, S., Waser, R., Menzel, S., and Linn, E. (2015) IEEE Journal on Emerging and Selected Topics in Circuits and Systems, pp. 1–9.
150. Prezioso, M., Merrikh-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K., and Strukov, D.B. (2015) *Nature*, **521** (7550), 61–64.

2

ReRAM Cells in the Framework of Two-Terminal Devices

E. Linn, M. Di Ventra, and Y. V. Pershin

2.1

Introduction

In this chapter, we emphasize that the framework of two-terminal devices is well suited to describe ReRAM cells in a compact way. In principle, one can either use the ideal two-terminal circuit elements (circuit model) to implement a device model or implement the device model by means of ordinary differential equations, that is, a dynamical system (mathematical model) [1]. In the case of ReRAM modeling, the latter approach is most widely used and termed “memristive modeling” [2]. The mathematical modeling offers a unified approach to describe virtually all possible two-terminal memory devices [3]. Furthermore, depending on the degree of knowledge one has on the device physics, either a black-box modeling approach (if one has no idea how the device works) or a white-box modeling (if the device physics is completely understood) can be applied. To develop highly accurate ReRAM device models, a combination of black- and white-box modeling is in general required since device operations are only partly understood.

The research on ReRAM devices is highly interdisciplinary, and scientists in the fields of physics, electrical engineering, chemistry, and materials science work together. Thus, to avoid any confusion, we begin by providing some basic definitions that are relevant to the discussion of ReRAM cells in the framework of two-terminal devices. Furthermore, we also provide a rigorous view on the limitations one should be aware of when dealing with ideal circuit elements from a physics point of view: *Any* system subject to external perturbations features a response function (e.g., its resistance) that always has some degree of memory (time nonlocality) [4].

The framework of two-terminal memristive elements can be considered as a convenient formalism to formulate the physical models of ReRAM devices. However, a postulation of memristive models, without referring to the actual device

operation physics, can lead to questionable results and predictions. One of such instances is the claim that the resistance switching memories are ideal memristors [5–7].

This chapter offers a critical analysis of this claim and highlights the limitations of ideal memristor models. Moreover, we emphasize that the correct classification of ReRAM cells is definitely in the class of memristive devices but not of ideal memristors. Thus, ReRAM elements cannot be considered the fourth nonlinear circuit elements [8].

This chapter is then organized as follows. First, we give a short introduction to the definitions relevant to two-terminal device modeling and cover the details of mathematical modeling of ReRAM cells (Section 2.2). In Section 2.3, we discuss a theoretical approach for describing elements with memory from a physics point of view. Next, we discuss ReRAM modeling from a practical viewpoint of a device engineer Section 2.4. Finally, we provide the conclusion in Section 2.5.

2.2

Two-Terminal Device Models

Electronic devices can be characterized by the number of terminals. For example, a MOSFET comprises four terminals (source, drain, gate, bulk), while a pn-diode comprises two terminals similar to ReRAM cells. Moreover, two-terminal device models often neglect the spatial dimension of the device and, thus, are considered lumped elements.

2.2.1

Lumped Elements

In electrical engineering, devices are in general considered as lumped elements. This assumption is valid if the device size is $\ll \lambda/4$ (wavelength of the excitation signal). A basic feature of lumped elements is the time invariance, that is, there is no explicit dependence on t in the device model.

2.2.2

Ideal Circuit Element Approach

There are three basic linear circuit elements: resistor R , capacitance C , and inductor L . According to Chua, ideal nonlinear circuit elements can also be defined by the variables V , I , ϕ , and q , resulting in four nonlinear passive devices: nonlinear resistor, nonlinear capacitor, nonlinear inductor, and (ideal) memristor. Even higher order circuit elements are feasible following this line [9]. One can use these ideal circuit elements for device modeling, especially black-box modeling, resulting in an abstract circuit representing the device. See, for example, [1].

2.2.3

Dynamical Systems Approach

Instead of using ideal circuit elements for modeling, one can directly apply the differential equations of the system to build the model. This approach is also very common in control theory where arbitrary systems and control thereof are considered. A dynamical system can be represented by two equations:

$$y = h(\mathbf{x}, \mathbf{u}, t) \quad (2.1)$$

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, t) \quad (2.2)$$

Equation (2.2) is the state equation, and (2.1) is the readout equation [10]. The functions $f(\cdot)$ and $h(\cdot)$ are nonlinear and dependent on time t as well as on variables \mathbf{x} and \mathbf{u} , which are multidimensional in general. The variable \mathbf{x} denotes the state of the system, the input variable \mathbf{u} reflects the external excitations, and y is the output or observation variable. This system-theoretical approach can be applied to arbitrary systems and is widely applied in electrical engineering.

2.2.3.1 Memristive Systems

A memristive system [11] is a special case of dynamical system (2.1)–(2.2) and displays a generic term for a complete class of two-terminal devices. These devices are characterized by a so-called pinched hysteresis loop [5, 12] and are nonlinear in general. A memristive system is defined by the state-dependent Ohm's law and the state equation, which is multidimensional and time-dependent in general. Note that a memristor [8] is only a special case of a first-order memristive system where the state variable equals the flown charge ($x = q$).

The relevance of the memristive system approach results from the possibility to model dynamic device behavior of resistive switching elements. This was first recognized by the HP group in 2008 [13] and has triggered many groups to work on memristive circuit models [12]. For a memristive system, y and \mathbf{u} are scalar values and y is a product of h and \mathbf{u} :

$$y = h(\mathbf{x}, u, t) \cdot u \quad (2.3)$$

$$\dot{\mathbf{x}} = f(\mathbf{x}, u, t) \quad (2.4)$$

Equation (2.3) shows that y is always zero when u is zero, which corresponds to a Lissajous figure with a pinched hysteresis loop. The definition of memristive systems was given in Ref. [11] and can be directly applied to two-terminal electronic devices. Memristive systems can be either current-controlled ($u = I$ and $h = R$) or voltage-controlled ($u = V$ and $h = G$). In this case, u is the input variable, while y is the output variable. A current-controlled memristive system reads

$$V = R(\mathbf{x}, I, t) \cdot I \quad (2.5)$$

$$\dot{\mathbf{x}} = f(\mathbf{x}, I, t), \quad (2.6)$$

while a voltage-controlled memristive system reads

$$I = G(\mathbf{x}, V, t) \cdot V \quad (2.7)$$

$$\dot{\mathbf{x}} = f(\mathbf{x}, V, t). \quad (2.8)$$

2.2.3.2 Memristor

A memristor is sometimes considered the fourth passive circuit element [8] and is a special case of a memristive system. A memristor has only one state variable, which is the flown charge $x = q$ in the current-controlled case

$$V = R(q) \cdot I \quad (2.9)$$

$$\dot{q} = I. \quad (2.10)$$

In the voltage-controlled case, the magnetic flux (flux linkage) is the state variable $x = \phi$

$$I = G(\phi) \cdot V \quad (2.11)$$

$$\dot{\phi} = V. \quad (2.12)$$

2.2.4

Significance of the Initial Memristor and Memristive System Definitions in the Light of Physics

At the heart of a memristive system theory is the definition of a general class of n th-order current-controlled memristive devices (Eqs. (2.5) and (2.6)) or voltage-controlled memristive devices (Eqs. (2.7) and (2.8)). Clearly, the memory feature of memristive devices is provided by the x -dependence of R in these equations.

Ideal memristor or simply memristor [14] is an electronic device whose memristance depends *only* on the charge $q(t)$ or flux linkage $\phi(t)$ (Eqs. (2.9)–(2.12)). These are two special cases of Eqs. (2.5) and (2.7). Currently, there is an understanding that the ideal memristor models [7] provide only a very superficial description of ReRAM response (see Section 2.2.4.1). In fact, more physically oriented memristive models – for example, the model of memristive systems with threshold proposed in Ref. [15] or the models discussed in Section 2.4 – need to be used for a realistic representation of the dynamic response of ReRAM cells.

Several general properties of memristive systems need to be mentioned. First of all, arguably the most known property is a “pinched hysteretic loop” – called the fingerprint of memristive devices – which is observed in I - V plots when the devices are subjected to a periodic input. Such loops have the form of “8” located in the first and third quadrants of I - V plane with a crossing at (0,0). Pinched hysteresis loops are regularly observed in experiments with ReRAM cells. However, the general form of Eqs. (2.5)–(2.7) and (2.6)–(2.8) and the fact that memristance can in principle take 0 or infinite values (considering passive resistive structures made of superconducting or insulating materials) allow, in principle, other shapes such as dumbbell-like loops or loops with several self-crossings. This means that pinched hysteretic curves are a typical feature of memristive elements, but not a necessary feature as argued in some literature [5, 6].

The hysteresis is more pronounced at frequencies of the external input that are comparable or lower¹⁾ than the frequencies of the internal processes that lead to memory. The hysteresis disappears at very high frequencies, when the internal state does not have enough time to follow the external input. A substantial hysteresis is normally seen at low frequencies, although it is reduced as compared to the most optimal case occurring when the frequency is comparable to the typical inverse relaxation time of the system.

Additionally, we note that the memristance can be continuously switched between two limiting values (usually, it is performed using an appropriate pulse sequence [16]). In many cases, however, the switching occurs only when the amplitude of the applied signal exceeds a certain threshold, which is not necessarily fixed or can be exactly defined due to nonlinear switching kinetics. The threshold-type switching is important for several potential future applications of memristive devices, in particular, in the area of Boolean logic [17, 18]. Moreover, real devices with purely memristive response are always passive. In this regard, they are different from memcapacitive and meminductive systems, which can in principle exhibit an active response [3].

Finally, the state variables – whether defining a continuum or a discrete set of states – may follow a *stochastic differential equation* rather than a deterministic one [19]. This case has received much less attention even though interesting phenomena, such as a noise-induced hysteresis [20] and a skewed charge probability density [21], have been predicted.

2.2.4.1 Limitations of Ideal Memristor Models

The ideal memristor models (see, e.g., Ref. [7]) are used by some researchers in the context of ReRAM cells although they lead to some inherent insufficiencies [4, 22].

The main prediction of these models, namely, that the memristance is a function of the cumulative charge flown through device ($R = R(q(t))$) or integral of the applied voltage ($R = R(\phi(t))$) is incompatible with experimental observations and certain physical principles such as symmetries of electrodynamics [4]. In particular, the ideal models predict a high sensitivity of the device state to input fluctuations, leading to what we call a *stochastic catastrophe* [4]. Additionally, these models violate the Landauer principle of the minimal heat generation for irreversible computation [22, 23] and result in an unphysical overdelayed switching effect and switching without a threshold [4].

2.2.5

Memristive, Memcapacitive, and Meminductive Systems

The concept of memristive systems was extended to memcapacitive and meminductive systems in Ref. [3]. Note, in the description of electronic devices with

1) But not very low as there is always such a timescale that the system has enough time to relax to its preferable configuration.

memory, the most important task is to identify the mechanism of memory response and associated internal state variables. In all possible cases, the memory is related to a change in the internal state of the device, which can be its phase, spin polarization, structure, shape, and so on (see our recent review paper [19] for many specific examples). As suggested in Ref. [2] for memristive systems, and generalized for instances of memcapacitive and meminductive systems in Ref. [3], in order to mathematically describe the internal state of a device, one can often use a vector x of n -internal state variables (or a continuous function giving rise to continuum memory elements [19]²⁾).

Here, we focus on the most common cases of discrete memory elements [19]. Following the definition in Ref. [3], let us then consider all possible relations between any two complementary constitutive circuit variables such as charge, current, voltage and flux linkage. Quite generally, with $u(t)$ and $y(t)$ denoting the input and output of the system and g a generalized response, we can define a general class of n th-order u -controlled memory systems (devices) according to the following relations:

$$y(t) = g(x, u, t)u(t), \quad (2.13)$$

$$\dot{x} = f(x, u, t), \quad (2.14)$$

where f is a continuous n -dimensional vector function.

Generally [19], the relation between current and voltage (when, in Eq. (2.13), $y(t)$ is the current and $u(t)$ is the voltage or vice versa) defines a memristive system [2], while the relation between charge and voltage specifies a memcapacitive system [3], and the flux–current relation gives rise to a meminductive system [3]. Two other pairs (charge–current and voltage–flux) are linked through equations of electrodynamics and therefore do not define any element. Devices defined by the relation between charge and flux (which is the time integral of the voltage) are not considered as a separate group since such devices can be redefined on the current–voltage relation basis [2]. Memristive, memcapacitive, and meminductive devices form a class of memory circuit elements or *memelements*.

Memristors (short for memory resistors) [14], memcapacitors (for memory capacitors) [3], and meminductors (memory inductors) [3] are special *ideal* cases of memristive, memcapacitive, and meminductive devices [3]. Models of ideal memristors exemplifying the definition of these ideal elements can be found in many publications (see, e.g., [5, 6, 14]). Figure 2.1 shows the circuit symbols of the three memelements.

2.2.6

ReRAM: Combination of Elements, Combination of Memory Features, and Consideration of Inherent Battery Effects

Memristive, memcapacitive, and meminductive systems are mathematical descriptions of a certain device behavior. With a device that cannot be described

2) In the literature, the former case of discrete memory elements is the most studied.

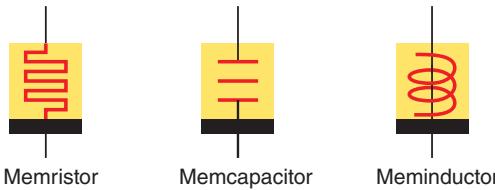


Figure 2.1 Symbols of memory circuit elements. Reprinted with permission from Ref. [3].

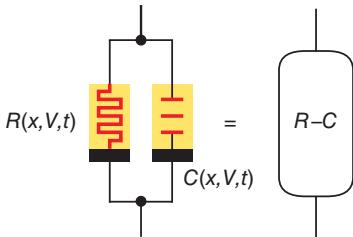


Figure 2.2 The combination of two or more memory-circuit elements in parallel or in series may result in the compact description of a single circuit element with nonalgebraic dependence on the control parameters [19]. Reprinted with permission from Ref. [19].

as a single memristive system, one has two options: first is the setup of a dynamical system that includes all the device behaviors. Alternatively, one can combine a memristive system with other circuit elements, for example, basic circuit elements or a memcapacitive system.

Typically, ReRAM memory cells are fabricated in the parallel-plate capacitor geometry. Therefore, the main memristive response of these cells should always be accompanied by a certain degree of capacitive response. Additionally, one can also expect a meminductive component as well as a battery-like component (the latter is associated with the internal dynamical/chemical processes in the ionic subsystem). In the cases when deviations from the purely resistive response needs to be taken into account, ReRAM cells can be described as effective memristive systems with a nonalgebraic dependence on the input [19]. Such a description provides a compact model of devices with complex functionality (see Figure 2.2).

Following Ref. [19], let us derive effective memristive system equations combining memristive and memcapacitive responses. For the sake of simplicity, let us assume that the same vector of state variables x determines the memory of both memristive and memcapacitive components. Using Kirchhoff's law, one can find that the device can be described as a voltage-controlled memristive system with current

$$I(t) = G_{\text{eff}}(x, V, t)V(t), \quad (2.15)$$

where the effective memductance (memory conductance) $G_{\text{eff}}(x, V, t)$ has a non-algebraic dependence on the bias V , namely

$$G_{\text{eff}}(x, V, t) = G(x, V, t) + \frac{1}{V(t)} \frac{d}{dt} [C(x, V, t) V(t)]. \quad (2.16)$$

Here, $G(x, V, t)$ is the memductance and $C(x, V, t)V(t)$ is the memcapacitance of the resistive and capacitive components of response, respectively. Experimentally,

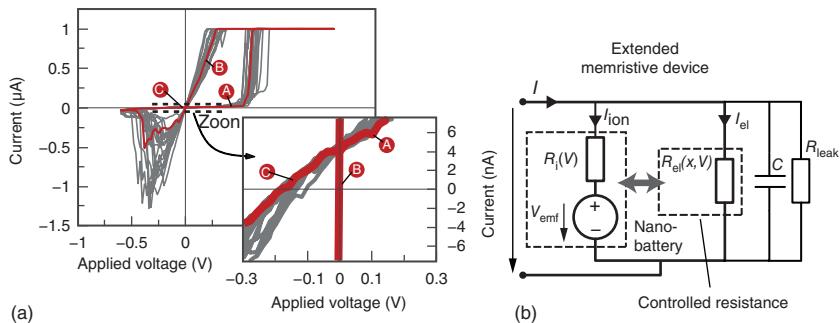


Figure 2.3 Equivalent circuit model of an ionic-type ReRAM device. Reprinted with permission from Ref. [27].

coexistence of memristive and memcapacitive responses was reported in, for example, perovskite oxide thin films [24] and vanadium dioxide structures [25, 26].

Having a closer look into ionic-type ReRAM devices (ECM, VCM, and TCM) reveals the fact that these devices require some modifications in terms of device modeling: these devices offer battery voltage and, thus, must be actually considered as active devices. Figure 2.3a depicts an I - V curve of an Ag/SiO₂/Pt ECM cell offering a nonpinched behavior due to the battery effect. The microscopic origin of the emf is either a Nernst potential or a diffusion potential or a Gibbs–Thomson effect [27]. Furthermore, ionic-type ReRAM devices are mixed electronic–ionic conductors and, thus, comprise two current branches, one electronic I_{el} and one ionic I_{ion} . The battery (V_{emf}) and the ionic resistance R_{ion} are located in the ionic current branch, which is the relevant current for the change of state variable (gap size x). The electronic resistance R_{el} , which is controlled by V and x , dominates the overall current in the low resistive state (see Figure 2.3b).

Strictly speaking, such ionic-type ReRAM devices cannot be termed memristive systems in the definition of [11]. For the sake of convenience, one of us (EL) suggested the use of the term “memristive device” also for those extended memristive systems [27]. Clearly, this is a reasonable classification as, in this case, the main component of the device response is resistive.

2.3

Fundamental Description of Electronic Devices with Memory

An ideal circuit element is assumed to respond instantaneously upon external perturbations. This is a simplification that is valid for a wide range of device models, but is not exactly true in physical systems. Here, we closely follow what we (YP, MD) outlined in a previous publication [4]. In particular, we show explicitly that the very fact that a physical system cannot respond instantaneously to an external perturbation leads to response functions that are nonlocal in time (memory-dependent).

To show this explicitly, it is enough to use the (classical or quantum) perturbation theory (also known as Kubo formalism [28, 29]) formulated in the 1950s, which dates back to the early developments of quantum mechanics. We focus on resistances, although similar derivations can be used for other response functions (see Ref. [4]). Let us then assume that we know the many-body Hamiltonian $\mathcal{H}(\{x\})$ of the system, which is dependent on the internal state variables $\{x\}$ (e.g., the classical positions and momenta of all ions). The state variables follow their own equation of motion (e.g., the classical Newton equation of motion):

$$\{\dot{x}\} = f(\{x\}, t). \quad (2.17)$$

We now switch on the electric field $\vec{E}(\mathbf{r}, t)$ at time t_0 and compute, in the linear regime, the corresponding induced current density. Note that in doing so, the right-hand side of Eq. (2.17) becomes electric-field-dependent. Simple algebra (also known as the Kubo response theory) leads us to ($\mu, \nu = x, y, z$)

$$j_\mu(\mathbf{r}, t) = \sum_\nu \int d\mathbf{r}' \int_{t_0}^t dt' \sigma_{\mu\nu}(\mathbf{r}, \mathbf{r}'; t, t'; \{x\}) E_\nu(r', t'), \quad (2.18)$$

where $\sigma_{\mu\nu}(\mathbf{r}, \mathbf{r}'; t, t'; \{x\})$ is a rank-2 tensor representing the electrical conductivity of the system in the direction μ under an electric field component in the direction ν .

By assuming a negligible magnetic field induced by the varying electric field, we can write $\vec{E} = -\vec{\nabla}V$, with V , the electric potential. Consider, now, a two-terminal device along the x axis. By integrating the current density over far-left and far-right device surfaces C_1 and C_2 (shown schematically in Figure 2.4), we obtain the total current

$$I(t) = G(\{x\}, t)V(t), \quad (2.19)$$

where the conductance G is

$$G(\{x\}, t) = - \int_{C_1} d\mathbf{r} \int_{C_2} d\mathbf{r}' \int_{t_0}^t dt' \sigma_{xx}(\mathbf{r}, \mathbf{r}'; t, t'; \{x\}). \quad (2.20)$$

If we go beyond linear response, we would need to include higher orders in the electric field perturbation, which in turn would introduce a nonlinear dependence on the bias V . The final result would then be [4]

$$I(t) = G(\{x\}, V, t)V(t), \quad (2.21)$$

which together with Eq. (2.17) (with the appropriate dependence on bias) represents what we now call a memristive system [2, 3]. We have thus shown what

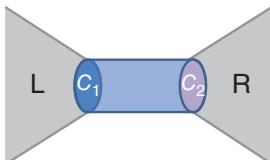


Figure 2.4 Schematic of the system coupled to left (L) and right (R) electrodes. C_1 and C_2 are integration areas in Eq. (2.20).

we had set out to do from the beginning: *any* resistive physical system subject to an external field would show some degree of memory (time nonlocality) in its response, and the equations describing the system response can be cast in the form of memristive system equations [2].

2.4

Device Engineer's View on ReRAM Devices as Two-Terminal Elements

One of the goals of device engineers is to formulate compact models of electronic devices that can be used later for the purposes of device optimization, prediction, and circuit simulations. While such models should be relatively compact – to be programmed, for example, in popular SPICE environments [30] – they should also include the device operation physics since, otherwise, the model predictive power becomes limited. Clearly, the first-principle approaches, such as that described in Section 2.3 being based on a large number of internal state variables (positions and velocities of all atoms), are of limited value to device engineers interested in compact models.

As mentioned earlier, we can distinguish three main approaches used for device modeling: black-box modeling and white-box modeling, a combination of both, and a gray-box method. On the one hand, the interpretation of the internal state variables in the black-box models is not trivial. On the other hand, the white-box modeling requires complete knowledge of the internal physical behavior of the device. This is rarely the case. Most frequent is the gray-box method, when some features of device dynamics are known, but all aspects of behavior or some model parameters cannot be extracted by experiment. Thus, a gray-box approach using some prior knowledge and combining this information with the measurement data is the most suitable for ReRAM modeling (cf. Figure 2.5).

Although the operation of all ReRAM devices based on metal–insulator–metal (MIM) structures is based on redox mechanisms, a unified view of ReRAM devices is only possible to a certain extent since the internal physical behavior differs quite strongly. Commonly, the ReRAM cells can be classified according to their mechanisms of switching as follows:

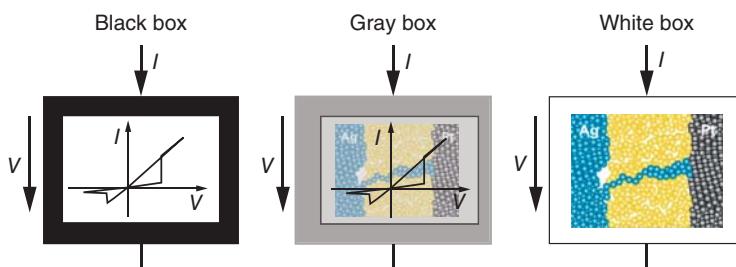


Figure 2.5 The black-box approach (left) only considers I - V data, while the white-box approach considers the actual internal device physics (right). The gray-box approach is based on both the I - V data and some knowledge of the device physics.

- Electrochemical metallization (ECM) cells;
- Valence change mechanism (VCM) cells;
- Thermochemical mechanism (TCM) cells.

Thus, it is appropriate to consider the basic ReRAM classes separately for two-terminal device modeling.

Typically, in ECM cells, one electrode consists of Cu or Ag and the other electrode is an inert material, for example, Pt or TiN. The I-layer can be either a bulk isolator (e.g., SiO_2) or a solid electrolyte (e.g., GeS) or an ion conductor (e.g., AgI). The basic principle in ECM is the formation/rupture of an Ag or a Cu filament in the I-layer of the device under bipolar operation voltage. For details on ECM, see Chapter 17.

For VCM cells, the metallic electrodes can be either symmetric (e.g., both TiN) or asymmetric (e.g., Ti and TiN). The I-layer is in general a layer of metal oxide (e.g., TiO_x , SrTiO_x , HfO_x , TaO_x). By an initial forming step, a device asymmetry is either inserted or reinforced, leading to a filamentary conducting channel. Movement of oxygen vacancy is considered the main switching mechanism in VCM devices, modulating the device resistance under bipolar voltage excitation (see Chapters 14 and 16 for details on the mechanism).

A typical I-layer material for TCM devices is NiO , but TiO_x is often used. TCM can be distinguished from VCM by means of operation, which is unipolar for TCM (see Chapter 13).

Note that in terms of VCM devices, a finer granularity of modeling attempts is required because the actual device equations and state variables might be different for different VCM devices. In the following, two exemplary ECM and VCM device models are presented. These models are formulated using memristive system equations.

2.4.1

Modeling of Electrochemical Metallization (ECM) Devices

In this section, a SPICE modeling approach for an ECM cell is illustrated. The physical 1D model was introduced by Menzel *et al.* [31], and the resulting SPICE implementation can be found in Ref. [32]. This model can be directly linked to a voltage-controlled memristive modeling approach described in Section 2.2:

$$I = G(x, V) \cdot V, \quad (2.22)$$

$$\dot{x} = f(x, V). \quad (2.23)$$

The corresponding ECM model equations are:

$$I = I_{\text{Tu}}(x, V) + I_{\text{ion}}(x, V), \quad (2.24)$$

$$\frac{\partial x}{\partial t} = -\frac{M_{\text{Me}}}{z e \rho_{m,\text{Me}}} J_{\text{ion}}, \quad (2.25)$$

where M_{Me} is the molecular mass, $\rho_{m,\text{Me}}$ the mass density of the deposited metal, z the ionic charge of the cations, and J_{ion} the corresponding ionic current density.

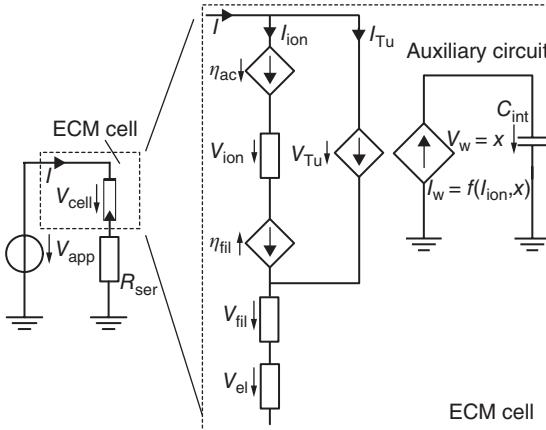


Figure 2.6 ECM memristive model.

The state variable x corresponds to the gap between the filament and the active electrode, I_{ion} is the current through the ionic path, while I_{Tu} represents the electronic tunneling current path. Figure 2.6 depicts the equivalent circuit model of the ECM cell. Details on the ECM model can also be found in Chapter 14 or in Ref. [31].

Our equivalent circuit implementation is based on the following equations:

$$I_{\text{fil}} = \begin{cases} j_0 A_{\text{fil}} \left(\exp \left(-\frac{\alpha e z}{k_B T} \Delta V_{\text{fil}} \right) - 1 \right) & \text{for } V \geq 0 \text{ V} \\ j_0 A_{\text{fil}} \left(\exp \left(\frac{(1-\alpha) e z}{k_B T} \Delta V_{\text{fil}} \right) - 1 \right) & \text{for } V < 0 \text{ V} \end{cases} \quad (2.26)$$

$$I_{\text{ac}} = \begin{cases} j_0 A_{\text{ac}} \left(\exp \left(\frac{(1-\alpha) e z}{k_B T} \Delta V_{\text{ac}} \right) - 1 \right) & \text{for } V \geq 0 \text{ V} \\ j_0 A_{\text{ac}} \left(\exp \left(-\frac{\alpha e z}{k_B T} \Delta V_{\text{ac}} \right) - 1 \right) & \text{for } V < 0 \text{ V} \end{cases} \quad (2.27)$$

I_{fil} is the ionic current at the filament/insulator interface and I_{ac} is the ionic current at the active electrode/insulator interface. j_0 is the exchange current density, α is the charge transfer coefficient, A_{fil} is the area of the filament/insulator interface (=filament area), and A_{ac} is the effective area of the active electrode/insulator interface. ΔV_{fil} and ΔV_{ac} are the overpotentials at the filament/insulator and active electrode/insulator interface, respectively. The ionic resistance of the gap leads to an additional voltage drop in this path

$$V_{\text{ion}} = \frac{\rho_{\text{ion}} \cdot x}{A_{\text{ion}}} \cdot I_{\text{ion}}, \quad (2.28)$$

where A_{ion} denotes the effective area of ionic drift in the switching layer and ρ_{ion} is the specific ionic resistivity of the switching layer. Finally, the electronic current

Table 1. Simulation parameters.

Parameter	Symbol	Value
Molecular mass	M_{Me}	$1.06 \times 10^{-22} \text{ g}$
Charge number	z	1
Mass density	$\rho_{\text{m,Me}}$	8.95 g cm^{-3}
Relative electron mass	m_r	0.86
Barrier height	ΔW_0	3.6 eV
Charge transfer coefficient	α	0.5
Exchange current density	j_0	$1 \times 10^{-2} \text{ A m}^{-2}$
Temperature	T	300 K
Filament area	A_{fil}	25.13 nm^2
Area of the active electrode	A_{ac}	25.13 nm^2
Effective area of ionic drift	A_{ion}	25.13 nm^2
Initial gap cell A	$x_{\text{init,A}}$	0.19 nm
Initial gap cell B	$x_{\text{init,B}}$	20 nm
Switching layer thickness	L	20 nm
Filament conductance	ρ_{fil}	$2 \times 10^{-8} \Omega \text{ m}$
Electrode resistance	R_{el}	76.4 mΩ

Figure 2.7 Table of typical parameters. From Ref. [32].

path is described by a tunneling equation:

$$I_{\text{Tu}} = \frac{3\sqrt{2m_{\text{eff}}\Delta W_0}}{2x} \left(\frac{e}{h} \right)^2 \exp \left(-\frac{4\pi x}{h} \sqrt{2m_{\text{eff}}\Delta W_0} \right) A_{\text{fil}} V_{\text{Tu}}, \quad (2.29)$$

where $m_{\text{eff}} = m_r \cdot m_0$ is the effective electron tunneling mass, ΔW_0 is the tunnel barrier height. In summary, Eqs. (2.26)–(2.28) describe the ionic current path, while Eq. (2.29) describes the electronic current path. This modeling approach is very close to the white-box approach since all equations have a clear physical meaning, as does the state variable x , which is the gap size. However, most parameters are unknown and, thus, must be either estimated or taken from the literature if available. A plausible parameter set can be found in Figure 2.7, reflecting basic Cu/SiO₂/Pt cell properties. In Figure 2.8, the device simulation results are depicted, Figure 2.8a for a single cell and Figure 2.8b for a three-element circuit, a complementary resistive switch plus a series resistor [33].

2.4.2

Modeling of Valence Change Mechanism (VCM) Devices

In this section, an exemplary memristive VCM modeling approach for Pt/TiO_x/Pt devices is presented [34–36]. This approach is physics-oriented, and the selection of equations is based on trial-and-error modifications. In so far, this “gray”-box modeling approach is closer to black-box approach compared to the presented ECM modeling approach. However, by this approach, realistic switching characteristics can be obtained and proper multielement simulations can be conducted.

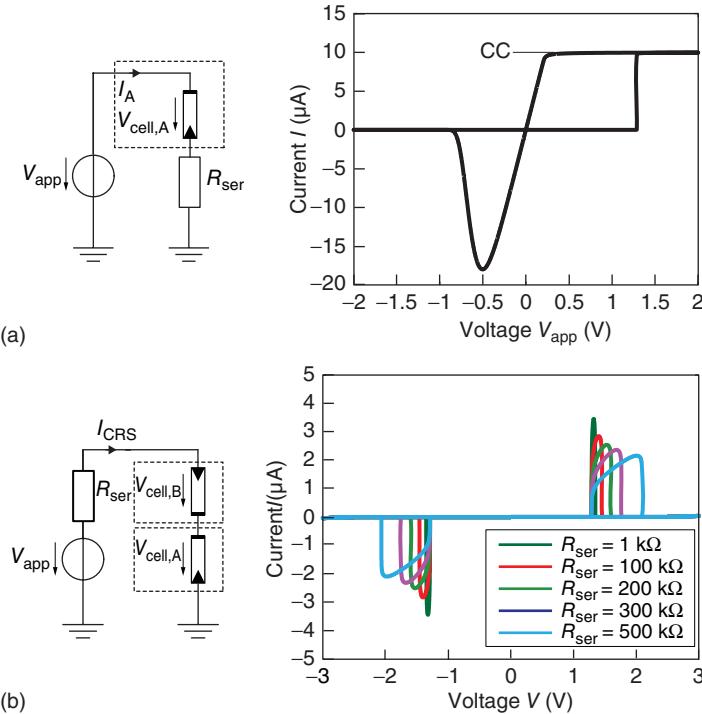


Figure 2.8 (a) ECM quasistatic sweep for a current compliance (CC) of $10 \mu\text{A}$. (b) Complementary resistive switch simulation for several series resistors. Adapted from Ref. [32].

In this model, the current is modeled as a tunneling current:

$$I = \frac{J_0 A}{\Delta x^2} \cdot \left\{ \phi_I \cdot \exp \left(-B \sqrt{\phi_I} \right) - \left(\phi_I + e |V_g| \right) \cdot \exp \left(-B \sqrt{\phi_I + e |V_g|} \right) \right\}. \quad (2.30)$$

The state variable x corresponds to the tunneling gap, \dot{x} , which is calculated as follows:

$$\dot{x} = \begin{cases} f_{off} \sinh \left(\frac{|I|}{I_{off}} \right) \exp \left(-\exp \left(\frac{x - a_{off}}{x_c} - \frac{|I|}{b} \right) - \frac{x}{x_c} \right) I > 0 \\ -f_{on} \sinh \left(\frac{|I|}{I_{on}} \right) \exp \left(-\exp \left(\frac{a_{on} - x}{x_c} - \frac{|I|}{b} \right) - \frac{x}{x_c} \right) I > 0 \end{cases}, \quad (2.31)$$

where ϕ_I is the barrier height:

$$\phi_I = \phi_0 - e |V_g| \left(\frac{x_1 + x_2}{x} \right) - \left(\frac{1.15 \lambda x}{\Delta x} \right) \ln \left(\frac{x_2 (x - x_1)}{x_1 (x - x_2)} \right). \quad (2.32)$$

Further parameters are defined as follows:

$$B = \frac{4\pi \Delta x \sqrt{2m}}{h}, \quad (2.33)$$

$$x_2 = x_1 + x \left(1 - \frac{9.2\lambda}{3\phi_0 + 4\lambda - 2e|V_g|} \right), \quad (2.34)$$

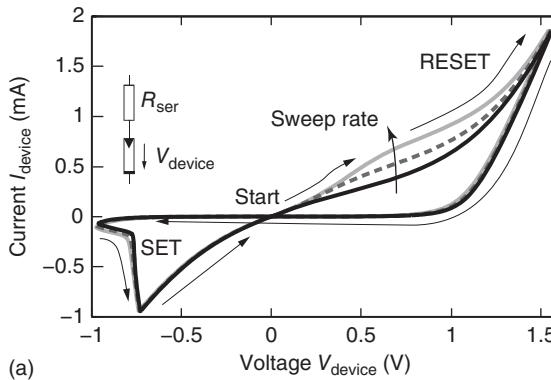
$$\lambda = \frac{e^2 \ln(2)}{8\pi\kappa\epsilon_0 x}, \quad (2.35)$$

$$J_0 = \frac{e}{2\pi h}, \quad (2.36)$$

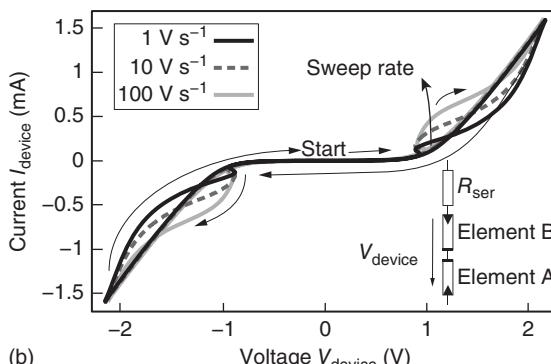
$$\Delta x = x_2 - x_1, \quad (2.37)$$

$$V_g = V_{\text{device}} - I \cdot R_s. \quad (2.38)$$

The parameter values can be found in Ref. [36]. In Figure 2.9a, the basic I - V characteristics for sweep rates of 1, 10, and 100 V s^{-1} are depicted. Note that an additional series resistance R_{ext} is considered in the simulation, but only



(a)



(b)

Figure 2.9 (a) Single-cell I - V characteristic for the following sweep rates: 1, 10, and 100 V s^{-1} . (b) Complementary resistive switching cell for sweep rates of 1, 10, and 100 V s^{-1} .

the actual device voltage is shown. This is the reason for the snapback during SET operation:

$$V_{\text{device}} = V_{\text{app}} - I \cdot R_{\text{ext}}. \quad (2.39)$$

In Fig. 2.9b, the feasibility of multielement simulation is shown by means of antiserially connected cells. The characteristics reflect the basic properties of a CRS cell. Although offering realistic dynamic I - V curves, this model has some weak points. First, as already mentioned, there is a lack of a clear physical understanding of VCM device physics; hence, each modeling equation is only a guess. The behavior of a different device, for example, of scaled size, might require a completely different set of fitting equations. No temperature effects are considered until now but are required for simulation [37]. Furthermore, the robustness of simulation of this model is a problem, as mentioned in Ref. [36]. For example, a symmetrical triangular voltage sweep of $+2/-2$ V as an input causes simulation stability problems. In conclusion, what we need in the future to achieve improved VCM models are:

- Physical equations for each device or better device class;
- Incorporation of temperature as a state variable;
- Optimized implementation strategies/approaches for highly nonlinear devices for robust circuit simulations.

2.5

Conclusions

ReRAM devices can be very well modeled in the framework of memristive systems or, more generally speaking, as a dynamical system. We have shown that memory properties are inherent to any physical system. While it may appear that the modern physics and electrical engineering fields offer two seemingly different approaches to model the resistance switching phenomenon, these two approaches are fundamentally the same. Obviously, the fundamental description of memristive devices in terms of the Kubo theory can be formulated exactly in the same form as the basic equations of the memristive theory. Clearly, the fundamental microscopic description of memristive switching is not always required, and, in many instances, not even desirable for compact device models. Therefore, the set of variables describing memristive devices can be significantly reduced. This results in effective models, whose formulation should be, however, based on physical theories or at least on results of experimental observations and not on postulates or presuppositions, which may or may not be physically correct.

As an example of compact ReRAM modeling, we have provided effective models of ECM and VCM memristive switching cells whose performance is similar to the experimental observation. Such a great match of these models and the actual devices is mainly due to the fact that the description relies on physical understanding of the device operation.

Acknowledgment

This work has been partially supported by NSF grants No. DMR-0802830 and ECCS-1202383 and the Center for Magnetic Recording Research at UCSD.

References

1. Chua, L. (1980) Device modeling via basic non-linear circuit elements. *IEEE Trans. Circuits Syst.*, **27** (11), 1014–1044.
2. Chua, L.O. and Kang, S.M. (1976) Memristive devices and systems. *Proc. IEEE*, **64**, 209–223.
3. Di Ventra, M., Pershin, Y.V., and Chua, L.O. (2009) Circuit elements with memory: memristors, memcapacitors, and meminductors. *Proc. IEEE*, **97** (10), 1717–1724, doi: 10.1109/JPROC.2009.2021077.
4. Di Ventra, M. and Pershin, Y.V. (2013) On the physical properties of memristive, memcapacitive, and meminductive systems. *Nanotechnology*, **24**, 255 201.
5. Chua, L. (2011) Resistance switching memories are memristors. *Appl. Phys. A*, **102** (4), 765–783.
6. Chua, L.O. (2011) The fourth elements. *Proc. IEEE*, **100**, 1920–1927.
7. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) The missing memristor found. *Nature*, **453**, 80–83.
8. Chua, L. (1971) Memristor – the missing circuit element. *IEEE Trans. Circuit Theory*, **CT-18** (5), 507–519.
9. Chua, L.O. (2003) Nonlinear circuit foundations for nanodevices, Part I: the four-element torus. *Proc. IEEE*, **91** (11), 1830–1859.
10. Marquez, H. (2003) *Nonlinear Control Systems: Analysis and Design*, Wiley-Interscience.
11. Chua, L. and Kang, S. (1976) Memristive devices and systems. *Proc. IEEE*, **64** (2), 209–223.
12. Pershin, Y.V. and Ventra, M.D. (2011) Memory effects in complex materials and nanoscale systems. *Adv. Phys.*, **60** (2), 145–227.
13. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) The missing memristor found. *Nature*, **453** (7191), 80–83.
14. Chua, L.O. (1971) Memristor - the missing circuit element. *IEEE Trans. Circuit Theory*, **18**, 507–519.
15. Pershin, Y.V., La Fontaine, S., and Di Ventra, M. (2009) Memristive model of amoeba learning. *Phys. Rev. E*, **80**, 021926, doi: 10.1103/PhysRevE.80.021926.
16. Pershin, Y.V. and Di Ventra, M. (2010) Practical approach to programmable analog circuits with memristors. *IEEE Trans. Circuits Syst. I*, **57**, 1857–1864.
17. Borghetti, J., Snider, G.S., Kuekes, P.J., Yang, J.J., Stewart, D.R., and Williams, R.S. (2010) ‘Memristive’ switches enable ‘stateful’ logic operations via material implication. *Nature*, **464**, 873–876.
18. Pershin, Y.V. and Di Ventra, M. (2012) Neuromorphic, digital and quantum computation with memory circuit elements. *Proc. IEEE*, **100**, 2071–2080.
19. Pershin, Y.V. and Di Ventra, M. (2011) Memory effects in complex materials and nanoscale systems. *Adv. Phys.*, **60**, 145–227.
20. Stotland, A. and Di Ventra, M. (2012) Stochastic memory: memory enhancement due to noise. *Phys. Rev. E*, **85**, 011116, doi: 10.1103/PhysRevE.85.011116.
21. Slipko, V.A., Pershin, Y.V., and Di Ventra, M. (2013) Changing the state of a memristive system with white noise. *Phys. Rev. E*, **87**, 042103, doi: 10.1103/PhysRevE.87.042103.
22. Meuffels, P. and Soni, R. (2012) Fundamental issues and problems in the realization of memristors. arXiv preprint arXiv:1207.7319.

23. Landauer, R. (1961) Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, **5**, 261–269.
24. Liu, S., Wu, N., Ignatiev, A., and Li, J. (2006) Electric-pulse-induced capacitance change effect in perovskite oxide thin films. *J. Appl. Phys.*, **100**, 056101.
25. Driscoll, T., Kim, H.T., Chae, B.G., Kim, B.J., Lee, Y.W., Jokerst, N.M., Palit, S., Smith, D.R., Di Ventra, M., and Basov, D.N. (2009) Memory metamaterials. *Science*, **325**, 1518–1521.
26. Driscoll, T., Quinn, J., Klein, S., Kim, H.T., Kim, B.J., Pershin, Y.V., Di Ventra, M., and Basov, D.N. (2010) Memristive adaptive filters. *Appl. Phys. Lett.*, **97**, 093502.
27. Valov, I., Linn, E., Tappertzhofen, S., Schmelzer, S., van den Hurk, J., Lentz, F., and Waser, R. (2013) Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nat. Commun.*, **4**, 1771.
28. Kubo, R. (1957) Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.*, **12** (6), 570–586.
29. Di Ventra, M. (2008) *Electrical Transport in Nanoscale Systems*, Cambridge University Press.
30. Bielek, D., Di Ventra, M., and Pershin, Y.V. (2013) Reliable SPICE simulations of memristors, memcapacitors and meminductors. *Radioengineering*, **22**, 945–968.
31. Menzel, S., Böttger, U., and Waser, R. (2012) Simulation of multilevel switching in electrochemical metallization memory cells. *J. Appl. Phys.*, **111** (1), 014501/1–5.
32. Linn, E., Menzel, S., Ferch, S., and Waser, R. (2013) Compact modeling of CRS devices based on ECM cells for memory, logic and neuromorphic applications. *Nanotechnology*, **24** (38), 384008.
33. Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010) Complementary resistive switches for passive nanocrossbar memories. *Nat. Mater.*, **9** (5), 403–406.
34. Yang, J.J., Pickett, M.D., Li, X., Ohlberg, D.A.A., Stewart, D.R., and Williams, R.S. (2008) Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat. Nanotechnol.*, **3** (7), 429–433.
35. Pickett, M.D., Strukov, D.B., Borghetti, J.L., Yang, J.J., Snider, G.S., Stewart, D.R., and Williams, R.S. (2009) Switching dynamics in titanium dioxide memristive devices. *J. Appl. Phys.*, **106** (7), 074508.
36. Abdalla, H. and Pickett, M.D. (2011) Spice modeling of memristors. 2011 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1832–1835.
37. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011) Origin of the ultra-nonlinear switching kinetics in oxide-based resistive switches. *Adv. Funct. Mater.*, **21** (23), 4487–4492.

3**Atomic and Electronic Structure of Oxides**

Tobias Zacherle, Peter C. Schmidt, and Manfred Martin

3.1**Introduction**

Macroscopic properties of a crystal such as transport properties and thermodynamic stability are determined by the electronic and atomic structures, which are closely related via the general notion of local or global crystal symmetry. By inspecting the simplest and most symmetric case, one can most readily see the connection between the geometrical structure and the electronic structure: an isolated atom. In free atoms, the electron density distribution $\rho(\mathbf{r})$ is spherically symmetric and the corresponding rotational symmetry dictates the appropriate quantum numbers (orbital momentum l , magnetic quantum number m_l). So, the atomic orbitals (AOs) are denoted as l and m_l and the principal quantum number as n . The situation changes for a solid, where the continuous full rotational symmetry is lowered to a discrete rotational symmetry. It turns out that the orbital momentum is not a good quantum number for a solid. Due to the translational symmetry of a crystal, it is replaced by the allowed values of the wave vector \mathbf{k} . It is not possible to provide a realistic description of a solid without taking into account both the electronic and the atomic structure. This is the subject of this chapter.

In the first section, we discuss the most important crystal structures of metal oxides (see also Section 6.2.1) and focus in particular on the structural building blocks and their connectivity, for example, edge-sharing octahedra. In the following section, we discuss the basics of the electronic structure of solids by first considering metal clusters and oxide clusters and how they form quasi-continuous states with increasing cluster size. Then, we focus on the band structure model, density of states (DOS), crystal field splitting, and exchange and correlation. We apply these concepts to understand and describe the electronic structure of different material classes. In the last section, we give selected examples of the electronic

structure of metal oxides. For further information on crystal structures and electronic structure, we refer the reader to the standard textbooks, “Introduction to solid state physics” by C. Kittel [1] and “Solid state physics” by N. W. Ashcroft, N. D. Mermin [2], which were used as main sources for the theoretical chapters on electronic structure. For further information on transition metal oxides, we recommend the book “Transition metal oxides: An introduction to their electronic and structural properties” by P. A. Cox [3].

We mainly focus on the concepts and theoretical methods, experimental methods such as XRD (X-ray diffraction), XES (X-ray emission spectroscopy), XAS (X-ray absorption spectroscopy), or EELS (electron energy-loss spectroscopy) in their various types, which are used to explore the atomic and electronic structures of oxides; however, these methods are not discussed here and we refer the reader to the more specialized literature [4].

3.2 Crystal Structures

The key symmetry property of crystalline materials is their translational invariance. The translation symmetry can be described by three *primitive lattice vectors* \mathbf{a} , \mathbf{b} , and \mathbf{c} forming the *primitive unit cell*. The vectors

$$\mathbf{R} = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}, \quad (3.1)$$

with integers n_i define the translational lattice (*Bravais lattice*). In Figure 3.1a, we show a simple example of a Bravais lattice, namely the face-centered cubic (fcc) lattice with its primitive lattice vectors $\mathbf{a} = a \left(0, \frac{1}{2}, \frac{1}{2} \right)$, $\mathbf{b} = a \left(\frac{1}{2}, 0, \frac{1}{2} \right)$, $\mathbf{c} = a \left(\frac{1}{2}, \frac{1}{2}, 0 \right)$ (a is the length of the cube, which forms the *conventional cell*) and the corresponding primitive unit cell. This crystal structure is common for various metals, for example, Cu or Al. If there are two atoms in the unit cell, that is, if a *diatomic basis* is chosen, for example, Mg at $(0, 0, 0)$ and O at $a \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right)$, we can construct the *rock-salt structure* (*NaCl structure*) by translating this basis with the primitive lattice vectors of the fcc structure, see Figure 3.1b. This structure is often found for oxides, as we will see later. By choosing another basis, for example, Zn at $(0, 0, 0)$ and S at $a \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)$, one obtains with the same fcc-lattice vectors the *zincblende structure*, see Figure 3.1c. Another example of a simple Bravais lattice is the body-centered cubic (bcc) lattice, see Figure 3.1d, which is realized, for example, for Na or the high-pressure phase of Mg.

It has to be noted that the choice of the unit cell is not unique. The same structure can be built up by choosing a different basis with different numbers of atoms and correspondingly different lattice vectors. For example in Figure 3.1b, the primitive cell contains one formula unit MgO, whereas the conventional cell (the cube) has four formula units.

The rock-salt structure is part of a larger class of structures, which are called *eutactic* [5, 6] and are common for binary oxides. These structures are formed by

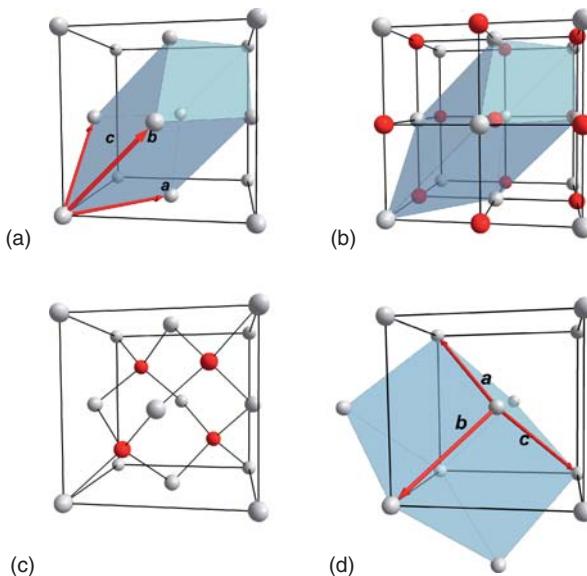


Figure 3.1 (a) Face-centered cubic (fcc) lattice with primitive lattice vectors and primitive cell (blue polyhedron), the conventional cubic cell being the surrounding cube. (b) Rock-salt structure: a superposition of two fcc lattices, which are shifted by $a\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. (c) Zincblende structure: a superposition

of two fcc lattices, which are shifted by $a\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. If all atoms are of the same type, (c) corresponds to the diamond structure. (d) Body-centered cubic (bcc) lattice with primitive lattice vectors and primitive cell (blue polyhedron), the conventional cell being the surrounding cube.

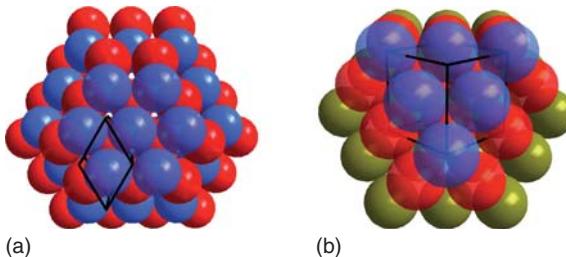


Figure 3.2 Close-packed structures: (a) ABAB stacking (hexagonal, close-packed) and (b) ABCABC stacking (cubic, close-packed), which also corresponds to the face-centered cubic (fcc) structure. The corresponding unit cells are also shown by the black lines.

choosing a close-packed anion lattice, see Figure 3.2, as the oxygen ions O^{2-} are mostly larger than the cations. The octahedral/tetrahedral holes in the structure are filled with cations. One can choose from two different close-packed structures: ABAB stacking for hexagonal close-packed (hcp) lattice, see Figure 3.2a, or ABCABC stacking for cubic close-packed (ccp) lattice, see Figure 3.2b. The latter is identical to the fcc lattice.

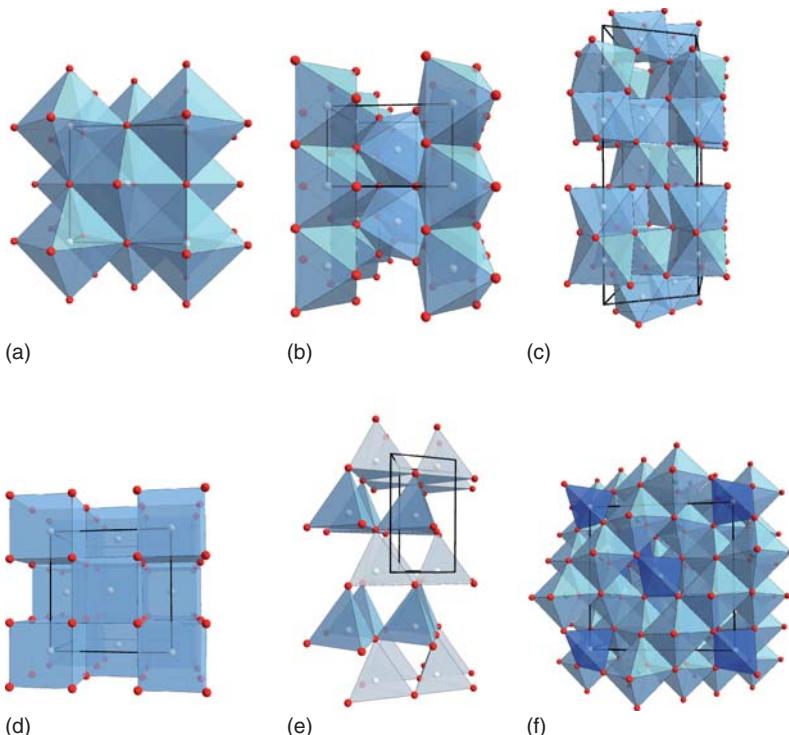


Figure 3.3 Eutectic crystal structures: (a) rock-salt structure with edge-sharing octahedra, (b) rutile structure with edge-sharing octahedra along the c -axis, whereas the columns are connected at the corners, (c) corundum structure with pairs of face-sharing octahedra along the c -axis, (d) fluorite structure with the simple cubic sublat-

tice of the anions, (e) wurtzite structure with corner-sharing tetrahedra, and (f) spinel structure: the cations are both octahedrally and tetrahedrally coordinated by anions. The corresponding conventional unit cells are shown by the black lines. Anions are shown in red and cations in white.

The percentage of filled tetrahedral/octahedral holes determines the final crystal structure. First, we consider the octahedral holes. By filling all the octahedral holes of a CCP structure, one obtains the *rock-salt* structure, see Figures 3.1b and 3.3a, for example, for MgO , NiO , or MnO . Here, we have an FCC lattice with edge-sharing octahedra.

Alternatively, half-filling of the octahedral holes of a HCP structure results in the *rutile* structure, see Figure 3.3b, for example, TiO_2 . The distorted octahedra are edge sharing along the tetragonal c axis, and the ones centered at the corners and at the center of the unit cell are rotated by 90° about the c -axis. The interactions along this axis have much debated consequences due to the possibility of dimerization, which will be discussed in detail later for the compounds NbO_2 and VO_2 .

For 2/3 filling of the octahedral holes of a hcp lattice, one obtains the *corundum* structure, see Figure 3.3c, which is named after the compound Al_2O_3 . Here, we have a network of edge- and face-sharing distorted octahedra, but of particular importance are the pairs of face-sharing octahedra along the trigonal *c*-axis. The corresponding cation–cation distance can be so small that metal–metal bonding is observed in some cases, for example, for Ti_2O_3 .

Now we consider tetrahedral holes. If all the tetrahedral holes are filled in a ccp structure, one obtains the *antifluorite* structure, for example, for Na_2O . For the *fluorite* structure, the cation lattice is close-packed and the anions are placed in all of the corresponding tetrahedral holes (CeO_2 , high-temperature phases of HfO_2 and ZrO_2). Here, the anions form a simple cubic sublattice and the cations are located in the centers of oxygen cubes, see Figure 3.3d.

If one fills half of the tetrahedral holes in a hcp anion lattice one obtains the *wurtzite* structure (e.g., ZnO). This lattice can be observed as composed of corner-sharing tetrahedra, see Figure 3.3e. Also, more complicated structures can be generated in this way. For example, the *spinel* structure (Figure 3.3f), named after the compound MgAl_2O_4 , is built up with a ccp anion sublattice, where 1/2 of the octahedral holes and 1/8 of the tetrahedral holes are filled. Therefore, both octahedrally and tetrahedrally coordinated cation positions exist. A well-known transition metal compound displaying this structure is Fe_3O_4 .

A simple structure wherein neither the anion sublattice nor the cation sublattice is close-packed is the *rhenium trioxide* structure (e.g., ReO_3 and WO_3 in distorted variants). The octahedra are corner-sharing and build a simple cubic lattice. If the center of the cube is occupied by a large cation (e.g., K, Sr, La, Ba), one obtains the *perovskite* structure, which is very common for ternary ABO_3 compounds, for example, BaTiO_3 and SrTiO_3 , see Figure 3.4.

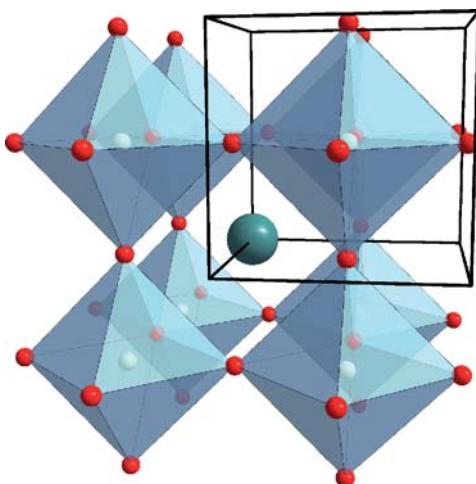


Figure 3.4 Perovskite structure with large cations in between the octahedra. Without the central cation, the structure corresponds to the ReO_3 structure.

Table 3.1 Different structure types with their corresponding space group.

Structure type	Space group	Close-packing	Holes filled (type, fraction)	Metal coordination	Connectivity	Examples
Rock salt	$Fm\bar{3}m$	O ccp	oct, 1	6	e.s. oct	NiO, MnO, MgO, SrO, TiO, FeO, CoO, and TaO
Rutile	$P42/mnm$	O hcp	oct, 1/2	6	c.s./e.s. oct	TiO_2 , VO_2 , NbO_2 , and TaO_2
Corundum	$R\bar{3}c$	O hcp	oct, 2/3	6	e.s./f.s. oct	Al_2O_3 , Ti_2O_3 , Cr_2O_3 , and V_2O_3
Fluorite	$Fm\bar{3}m$	M ccp	tet, 1	8	e.s. cub	ZrO_2 , HfO_2 , and CeO_2
Wurtzite	$P63mc$	O hcp	tet, 1/2	4	c.s. tet	ZnO
Spinel	$Fd\bar{3}m$	O ccp	oct, 1/2 tet, 1/8	4 and 6	e.s. oct c.s. tet/oct	Fe_3O_4
ReO_3	$Pm\bar{3}m$	—	—	6	c.s. oct	WO_3 and ReO_3
Perovskite	$Pm\bar{3}m$	—	—	A: 12 B: 6	c.s. oct	$SrTiO_3$
ABO_3						

The close-packed structure (e.g., O (M) ccp (hcp) for oxygen (metal) sublattice is cubic (hexagonal) close-packed), the amount of filled octahedral (oct) or tetrahedral (tet) holes in the close-packed structure, the metal coordination, and the connectivity of the building blocks (e.s. for edge sharing, c.s. for corner sharing, f.s. for face sharing, oct for octahedra, tet for tetrahedra, and cub for cubes). We also give examples for the mentioned structure types.

Finally, we discuss a class of structures known as *Magnéli phases*, which are more complicated than the structures previously discussed and have been reported first for the molybdenum oxides [7]. They form a homologous series with the formula M_nO_{2n-1} . For $n=2$, one obtains the corundum structure M_2O_3 , whereas the limit structure for $n \rightarrow \infty$ is rutile MO_2 . The structure is composed of rutile-like slabs of infinite extension separated by crystallographic shear planes, which are structurally similar to corundum. The formula $M_nO_{2n-1} = M_2O_3 + (n-2) MO_2$ also validates this qualitative picture [8]. Well-known examples for Magnéli phases are found in the titanium–oxygen system, where they have been observed for $3 \leq n \leq 10$ [9], as well as in the vanadium–oxygen system ($3 \leq n \leq 8$ [10]).

In Table 3.1, we summarize the properties of the structures discussed thus far. In some materials (further), distortions of octahedra/tetrahedra lead to different symmetries, although the structures are derived from the aforementioned ones, see Section 3.5.4.

3.3 Electronic Structure

Solid materials can roughly be divided into three classes: metals, semiconductors, and insulators. This classification can only be understood in terms of

electronic bands. Whereas in isolated atoms we have discrete electronic energy levels, in a solid we have – due to interactions of the atoms – ranges of energy with quasi-continuous energy levels called electronic bands. These energetically allowed bands can be separated by energetically forbidden energy ranges, the *energy gaps*. If the occupied and unoccupied states are separated by a band gap, the highest occupied band is termed *valence band*, and the lowest unoccupied band is called *conduction band*. The energy gap between the *valence band maximum* (VBM) and the *conduction band minimum* (CBM) is called *optical band gap* E_{opt} .

A semiconductor has a small optical band gap of <2 eV (typical values: 0.67 eV for Ge, 1.1 eV for Si, and 1.43 eV for GaAs [2]). For electronic transport, electrons have to be thermally excited to the conduction band and the resulting conductivities lie in the broad range from 10^{-8} to 10^3 S cm $^{-1}$.

An insulator has a larger gap of roughly >4 eV, so the conductivities at room temperature are very small (10^{-18} to 10^{-10} S cm $^{-1}$). However, the distinction between semiconductors and insulators is somewhat arbitrary, for example, Ga₂O₃ has an optical band gap of 4.9 eV and is nevertheless called a wide-band-gap semiconductor.

In a metal, there is no optical band gap, and the excitation of electrons to unoccupied states due to an applied electric field leads to high electrical conductivities of typically 10^4 – 10^6 S cm $^{-1}$ at room temperature.

In the following section, we provide a more detailed description of the aforementioned material classes, in particular in the context of oxides. To achieve this objective, we first briefly explain and review the theory of electronic structure starting from small clusters and then moving to crystals.

3.3.1

From Free Atoms to the Solid State

We consider the valence states of metal clusters and metal oxide clusters and start with a metal dimer (Mg₂). In a simple molecular orbital (MO) approach [11], the MOs are built from localized AOs

$$\psi_{\text{MO}}(\mathbf{r}) = c_a \psi_a(\mathbf{r}) + c_b \psi_b(\mathbf{r}), \quad (3.2)$$

where the AOs or basis functions are centered at the positions of the nuclei a and b .

The MOs for Mg₂ are shown in the left column of Figure 3.5a. For the lowest energy state E_1 , the atomic electron charge distributions overlap and we obtain covalent bonds with $c_a = c_b$ (*bonding state*) in the simple MO picture. This can be seen from the isosurface of the corresponding charge density distribution $\rho_1(\mathbf{r})$. For E_2 , one obtains $c_a = -c_b$ (*antibonding state*) and the MO charge density distribution ρ_2 has a node plane between the nuclei. E_1 and E_2 are the occupied states (blue color). By using a larger basis function set, one also obtains the unoccupied states (red color), for example, E_3 and E_4 .

Next, we consider linear clusters Mg _{m} as a function of the cluster size m . In Figure 3.5, we have plotted the energy levels E_i and, for some states, an isosurface

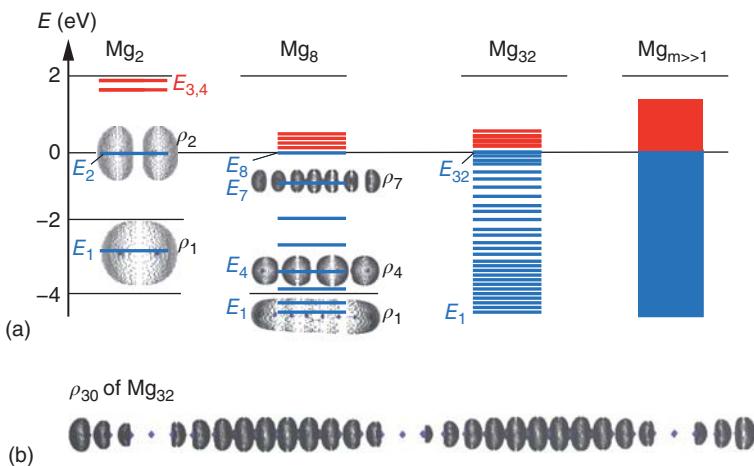


Figure 3.5 Dimer Mg_2 and clusters Mg_m . (a) Energy level diagram for the dimer Mg_2 and linear clusters Mg_m . The highest occupied state is the zero of the energy scale. For some states, isosurfaces of the charge den-

sity distribution $\rho_i(\mathbf{r})$ are shown. Blue levels are occupied, red levels unoccupied. (b) Iso-surfaces of the charge density distribution $\rho_{30}(\mathbf{r})$ of the Mg_{32} cluster.

of the valence charge densities $\rho_i(\mathbf{r})$. We follow the convention that the *zero of the energy scale is the highest occupied state*. With increasing cluster size m , the difference between neighbored energy values decreases and the energy levels of the eigenstates become quasi-continuous and for the extrapolation to the crystal, $m \rightarrow \infty$. One finds no energy gap E_{gap} between the occupied and unoccupied states, so the linear Mg chain is a metallic system.

Figure 3.5 also shows the bonding character in solids. Except for the lowest (fully bonding) and highest (fully antibonding) occupied states, the bonding is always a mixture of bonding and antibonding regions. This can be clearly seen from the pure covalent bonds of Mg_m . In this context, we recall that the antibonding states have a node plane in the middle of the axis between two neighboring nuclei. As can be seen from the densities of Mg_8 , there are $i-1$ node planes for each shown $\rho_i(\mathbf{r})$.

Coming now to the metal oxide clusters, we first obtain the MO levels for the molecule (MgO) as shown in the left column of Figure 3.6a. The AO energies for Mg 3s and 2p O are different, and therefore, we achieve a charge transfer $-\delta e$ from a to b and the molecule has a dipole moment. The covalent bond contribution is weaker than that in Mg_2 . However, additionally, there is an ionic bond contribution, which can be roughly estimated by Coulomb's law

$$U_{\text{ionic}} = \frac{1}{4\pi\epsilon_0} \frac{(+\delta e)(-\delta e)}{d}, \quad (3.3)$$

where d is the bond distance. The valence charge density is shown in the left column of Figure 3.6b.

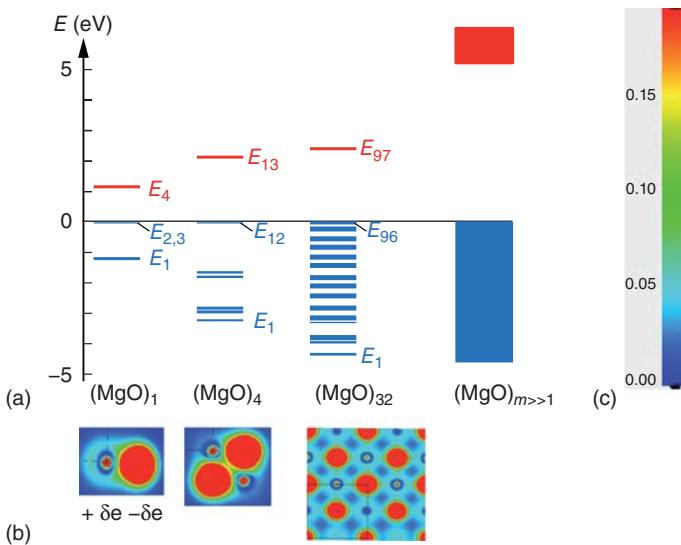


Figure 3.6 MgO molecule and $(\text{MgO})_m$ clusters. (a) Energy level diagram for the MgO molecule and $(\text{MgO})_m$ cubic clusters. Blue levels are occupied, red levels unoccupied.

The highest occupied state is the zero of the energy scale, (b) total valence charge density distribution (in $\text{e} \text{\AA}^{-3}$) for one plane. The scale is shown in (c).

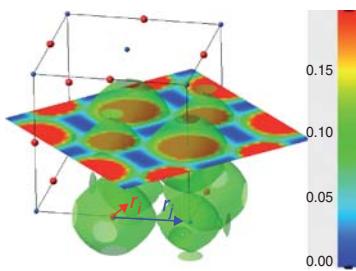


Figure 3.7 Total charge density distribution $\rho(\mathbf{r})$ in fcc MgO. $\rho(\mathbf{r})$ (in $\text{e} \text{\AA}^{-3}$) is shown for a (001) plane, and additionally, one isosurface of $\rho(\mathbf{r})$ is plotted.

The charge separation δ increases with increasing cluster size because the external potential of the partial charges $\pm\delta e$ lowers more and more the energy of the localized oxygen 2p states, that is, stabilizes more and more the O 2p states until δ approaches the value of 2 for the MgO crystal. In a first approximation, the electrostatic potential (*Madelung potential*) $\phi_M(\mathbf{r}_i)$ “seen” by the O 2p states can be described simply by

$$\phi_M(\mathbf{r}_i) = \frac{1}{4\pi\epsilon_0} \sum_j \frac{q_j}{|\mathbf{r}_i - \mathbf{r}_j|}; \quad q_j = (+\delta e) \text{ or } (-\delta e), \quad (3.4)$$

where the O nucleus is the zero of the coordinate system and \mathbf{r}_j are the nuclei sites of the surrounding ions, see Figure 3.7. The Figure shows the total charge density of the MgO crystal in contrast to the valence charge density shown in Figure 3.6.

The charge separation results in localized fully occupied O 2p states stabilized by the Madelung potential and empty Mg 3s states, and we obtain an optical band gap of several electron volts; therefore, crystalline MgO is an insulator.

In the pure ionic model, the crystalline charge distribution is spherically symmetric around the nuclei and the electrostatic energy (*Madelung energy*) of the compound is given by

$$V_M = \frac{1}{2} \frac{1}{4\pi\epsilon_0} \sum_{ij} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}; \quad q_{ij} = (+2) \text{ or } (-2e). \quad (3.5)$$

Although the ionic model might be quite satisfactory for MgO, for many other oxides, the pure ionic model is only a poor approximation to describe the real charge distribution, as there is a distinct covalent contribution to the chemical bond.

3.3.2

Electrons in Crystals

3.3.2.1 Free Electron Model (Sommerfeld Model)

In the previous section, we discussed the chemical bond on the basis of localized atomic orbitals centered at the atomic nuclei. In this section, we consider the charge density distribution starting from another set of basis functions, namely totally delocalized states, the electronic *plane waves*. One can get a first hint of this from $\rho_{30}(\text{Mg}_{32})$ shown in Figure 3.5b. The envelope function resembles a standing wave with four maxima.

We start with the one-electron Schrödinger equation [2]

$$\left(-\frac{\hbar^2}{2m} \nabla_i^2 + V_i(\mathbf{r}) \right) \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}), \quad (3.6)$$

where $\psi_i(\mathbf{r})$ is the orbital and $V_i(\mathbf{r})$ is the potential energy “seen” by the electron i . To describe the valence electron states of simple metals, in a first approximation, $V_i(\mathbf{r})$ can be replaced by a constant $(-V_0)$ as sketched in Figure 3.8 (particle in a box).

We consider the electron inside a cubic box of length L and assume a constant potential V_0 . We replace E_i in Eq. (3.6) by $E_i + V_0$ to obtain

$$-\frac{\hbar^2}{2m} \nabla_i^2 \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \quad (3.7)$$

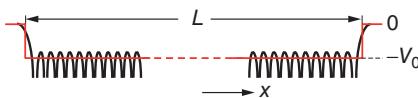


Figure 3.8 A rough sketch of the shape of the potential “seen” by an electron in a metal. To get the valence electron states of simple metals, in a first approximation, the real potential can be replaced by macroscopic potential of depth $-V_0$ and length L .

We recall that the solutions of this equation for free electrons are sine or exponential functions. Depending on the chosen boundary conditions, one obtains standing or traveling waves.

The simplest traveling wave is a plane wave $\psi_i(\mathbf{r}) = L^{-3/2}e^{i\mathbf{k}\cdot\mathbf{r}}$ with a constant electron probability distribution $\rho_i(\mathbf{r}) = \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}) = L^{-3}$ (or charge density $e\rho_i(\mathbf{r})$) for one electron inside the cubic box. The discrete quantum states i are defined through the boundary conditions. Most crystal properties are well described using the so-called *Born–von Karman boundary conditions*, which means that the wave function has to be invariant when adding to the vector $\mathbf{r} = (x, y, z)$, for example, a vector $(L, 0, 0)$: $\psi_i(x + L, y, z) = \psi_i(x, y, z)$. With the definition of ψ_i , this translates directly into the equation $e^{i\mathbf{k}_x \cdot \mathbf{L}} = 1$. This yields a quantization of k_x (y and z component accordingly)

$$k_x = \frac{2\pi n_x}{L}, \quad k_y = \frac{2\pi n_y}{L}, \quad k_z = \frac{2\pi n_z}{L}. \quad (3.8)$$

Therefore, the quantum number i introduced in Eq. (3.6) has to be replaced by (n_x, n_y, n_z) . The electronic states are labeled by the k -vector $\mathbf{k} = (k_x, k_y, k_z)$. By inserting the wave function $\psi_{\mathbf{k}}(\mathbf{r}) = L^{-3/2}e^{i\mathbf{k}\cdot\mathbf{r}}$ into the Schrödinger equation (3.7), one obtains for the energy eigenvalues

$$E(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m}, \quad (3.9)$$

where Eq. (3.8) defines the k -vectors *allowed* for electronic states. The difference Δk between two neighboring allowed k -vectors is extremely small, $\Delta k = 2\pi/L$, and we get a quasi-continuous state distribution in k -space called *electronic band*. From Eq. (3.9) we see that the characteristic k -dependence of the energy of free electrons is a parabola (*parabolic dispersion* or *parabolic band*), see also Figure 3.9.

Electrons have spin $1/2$ (Fermions) and are therefore subject to Fermi–Dirac statistics, that is, the occupation number $f(E)$ of states with energy E is given by

$$f(E) = \frac{1}{\exp\left(\frac{E-E_F}{k_B T}\right) + 1}, \quad (3.10)$$

where k_B is the Boltzmann constant and E_F is the *Fermi level*. The latter is defined by $f(E_F) = 1/2$, or, more precisely, it is identified with the electrochemical potential of the electrons, that is, the change of energy associated with removal/addition of an electron to the system.

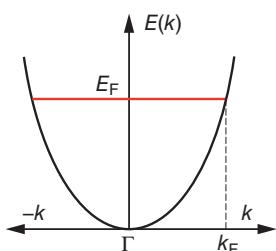


Figure 3.9 Energy $E(\mathbf{k})$ (*dispersion curve*) of the electronic states in the free electron model. $E(\mathbf{k})$ is quasi-continuous. The minimum distance between two allowed k -points is $2\pi/L$. At 0 K, the states are occupied up to the Fermi energy $E_F = (\hbar^2 k_F^2 / 2m)$. The k -vector $\Gamma = (0; 0; 0)$ is called Γ -point.

Each quantum mechanical state described by the wave vector \mathbf{k} can be occupied by two electrons (spin up and down). In a metal, for a given number of electrons N , at 0 K, all available states are filled, starting from the electronic state with the lowest energy up to the Fermi energy, which is here identical to the Fermi level E_F . This is shown in Figure 3.9 for a gas of free electrons.

The Sommerfeld model works well for the study of electronic properties of alkali metals, but for other metals or semiconductors/insulators, the correct crystal structure has to be taken into account.

3.3.2.2 Band Structure Model

So far, we have only considered a gas of noninteracting electrons within a box and Born–von Karman boundary conditions, where the charge density of the electrons is constant. Of course, in real materials (also in metals), the charge density reflects the translational symmetry of the lattice. So, for the general case of an electron in a crystal lattice, there are some modifications of the presented results [1, 2].

The wave functions of the electronic states in a crystal can be written as the so-called *Bloch functions* $\psi(\mathbf{r}) = u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$, where $u(\mathbf{r})$ has the translational symmetry of the lattice, $u(\mathbf{r} + r\mathbf{a} + s\mathbf{b} + t\mathbf{c}) = u(\mathbf{r})$ (with r, s, t being integers). Then, the charge density has the same translational symmetry as the lattice because $\rho(\mathbf{r} + r\mathbf{a} + s\mathbf{b} + t\mathbf{c}) = \rho(\mathbf{r}) = \psi_k(\mathbf{r})\psi_k^*(\mathbf{r})$. The wave function of free electrons is a special case of a Bloch function with $u(\mathbf{r})$ being a constant: $u(\mathbf{r}) = L^{-3/2}$.

By using again periodic boundary conditions and assuming a cubic primitive lattice $|\mathbf{a}| = |\mathbf{b}| = |\mathbf{c}|$, we get the same allowed k -vectors as for the free electron model with $L = N|\mathbf{a}|$, where N^3 is the total number of unit cells. If we solve the Schrödinger equation for a Bloch function, it turns out that $u(\mathbf{r})$ also depends on \mathbf{k} : $u(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})$.

Furthermore, by assuming formally large values of the lattice constants, for example, $|\mathbf{a}| \rightarrow \infty$, $k \rightarrow 0$, $\psi(\mathbf{r}) = u(\mathbf{r})$ becomes an atomic orbital, for example, the 3s orbital for the Mg metal – that is, for each atomic orbital of the unit cell, one band is obtained. Therefore, we need an additional quantum number to label the states originating from different atomic orbitals. This quantum number is called *band index* n : $E(\mathbf{k}) \rightarrow E_n(\mathbf{k})$, $u_{\mathbf{k}}(\mathbf{r}) \rightarrow u_{n,\mathbf{k}}(\mathbf{r})$, and $\psi_{\mathbf{k}}(\mathbf{r}) \rightarrow \psi_{n,\mathbf{k}}(\mathbf{r})$.

Therefore, the final form of the crystal orbital is

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (3.11)$$

That is, crystal orbitals are like atomic orbitals in the vicinity of the nuclei superimposed with plane waves, similarly to the sketch in Figure 3.5b.

If we have no primitive cubic structure, the determination of the allowed k -vectors is more complicated. One has to define the unit cell of the reciprocal lattice with basis vectors \mathbf{a}^* , \mathbf{b}^* , and \mathbf{c}^* , which are given by

$$\begin{aligned} \mathbf{a}^* &= (2\pi/V^*)\mathbf{b} \times \mathbf{c} \\ \mathbf{b}^* &= (2\pi/V^*)\mathbf{c} \times \mathbf{a} \\ \mathbf{c}^* &= (2\pi/V^*)\mathbf{a} \times \mathbf{b} \quad \text{and} \quad V^* = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}). \end{aligned} \quad (3.12)$$

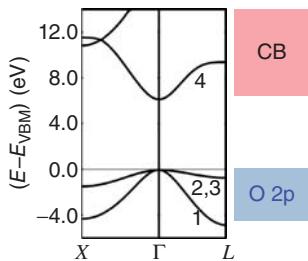


Figure 3.10 Band structure of MgO. The band structure is displayed for two directions in k -space, from $X = (0.5, 0, 0.5)$ to $\Gamma = (0, 0, 0)$ to $L = (0.5, 0.5, 0.5)$. The valence band maximum (VBM) is the zero of the energy scale. (The lattice constant is taken from Ref. [12].)

Then, the allowed k -vectors are given by

$$\mathbf{k} = \frac{n_a}{N_a} \mathbf{a}^* + \frac{n_b}{N_b} \mathbf{b}^* + \frac{n_c}{N_c} \mathbf{c}^*, \text{ with integers } n_a, n_b, n_c. \quad (3.13)$$

N_a, N_b , and N_c are the number of unit cells in the \mathbf{a}, \mathbf{b} , and \mathbf{c} directions, respectively.

The symmetry greatly simplifies the analysis of the electronic structure. From the translation symmetry of the Hamiltonian follows a translation symmetry of $E(\mathbf{k})$ in k -space, for example, $E(\mathbf{k}) = E(\mathbf{k} + \mathbf{a}^*)$, and therefore, it is sufficient to plot the band structure in the \mathbf{a}^* -direction in the range $0 \leq k \leq a^*$ or $-\frac{1}{2}a^* \leq k \leq \frac{1}{2}a^*$. By taking into account also the symmetry $E(-\mathbf{k}) = E(\mathbf{k})$, the relevant range reduces to $0 \leq k \leq \frac{1}{2}a^*$. The unit cell $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$ of the reciprocal space is called (first) *Brillouin zone* (BZ), and because of the symmetry of $E(\mathbf{k})$, it is sufficient to display the band structure for the so-called irreducible part of the BZ. Each band in the BZ can be occupied by $2N_a N_b N_c$ electrons (factor 2 for the spin) or by two electrons per unit cell as there are $N_a N_b N_c$ unit cells.

In Figure 3.10, we show as an example the band structure of rock-salt MgO. The band structure is displayed for two directions in k -space, from $X = (0.5, 0, 0.5)$ to $\Gamma = (0, 0, 0)$ to $L = (0.5, 0.5, 0.5)$ in multiples of the unit vectors $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ for the fcc lattice. We only display the highest occupied bands (O 2p, labeled 1–3) and the lowest unoccupied bands (conduction bands (CBs), the lowest unoccupied band is labeled 4). The occupied valence states are the states of the O 2p electrons. The highest occupied state is the VBM and is indicated by the chosen zero of the energy scale. The band gap between the VBM and the lowest conduction band state at the Γ -point, the CBM, is 6.2 eV. The experimental value of the band gap of MgO is about 8 eV, depending on the measurement [13, 14]. In theoretical calculations, the predicted band gaps depend sensitively on the chosen crystal potential $V_i(\mathbf{r})$. In this work, we have chosen the hybrid potential HSE06 [15]; for details, see Section 3.3.2.6.

As we consider only the 2p electrons of oxygen and the 3s electrons of Mg, we have in total six electrons per primitive unit cell (one formula unit MgO). These six electrons occupy three bands (spin-up and -down electrons in one band) marked 1–3. The upper valence band (marked 2, 3) is twofold degenerate (same energies for two bands).

We see that the dispersion curve $E(\mathbf{k})$ for a real material differs significantly from the parabolic behavior for free electrons as shown in Figure 3.9. However,

to study the electronic conductivity, the relevant electronic states are the states at the bottom of the conduction band (marked 4 in Figure 3.10). In MgO, the CBM is located at the Γ -point and the dispersion curve of the first conduction band shows almost parabolic character around Γ . A Taylor expansion of the conduction band energy $E(\mathbf{k})$ around this point gives:

$$E(\mathbf{k}) = E(\mathbf{k} = (0, 0, 0)) + \sum_{i,j} \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j} k_i k_j + \mathcal{O}(k^3). \quad (3.14)$$

In the free electron model, we have $(\partial^2 E(\mathbf{k}) / \partial k^2) = (\hbar^2 / m)$, see Eq. 3.9. In analogy, we define the so-called inverse effective mass tensor for the interacting electrons at the CBM by:

$$\left(\frac{1}{m^*} \right)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}. \quad (3.15)$$

In the isotropic case, there is only one value $\frac{1}{m^*}$ for all directions and the energy of the electronic states at the CBM for sufficiently small \mathbf{k} is described by

$$E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m^*}, \quad (3.16)$$

and formally, the free electron dispersion is recovered around the Γ -point. A band with a large (small) curvature has a small (large) effective mass for electron transport. Flat bands with no dispersion correspond to localized states; see also Sections 6.3.1 and 6.3.2.

3.3.2.3 Density of States (DOS) and Partial DOS

In addition to the band structure, the electronic states are also characterized by the DOS $n(E)$. $n(E)dE$ is equal to the number of states in an interval dE around E . If only the valence electrons are considered, the integral over $n(E)$ up to the energy of the VBM E_{VBM} is equal to the number of valence electrons N_{ve} in the unit cell

$$N_{\text{ve}} = \int^{E_{\text{VBM}}} n(E) dE. \quad (3.17)$$

For the Sommerfeld model, $E(\mathbf{k})$ depends only on the absolute value k of \mathbf{k} and the number of states for an energy E is proportional to the surface of a sphere in k -space $n(E)dE \sim k^2 dk$. As $dk \sim \frac{dE}{k}$, it follows that $n_{\text{free}}(E) \sim \sqrt{E}$.

The band structure and DOS yield the electron distribution in the reciprocal and the energy space, respectively. If we are interested in the electron distribution in real space, we have to display the charge distribution as shown in Figure 3.6b. A representation that incorporates the information on the electron distribution in both real and reciprocal spaces is the so-called *partial DOS* plot, which is obtained in the following way.

We choose spheres with radius R around the nuclei positions as indicated in Figure 3.11. Then, for each electron state $\psi_{n,k}(\mathbf{r})$, we calculate the partial charges inside these spheres, which contribute to $\psi_{n,k}(\mathbf{r})$. Additionally, we can split the partial charges according to the angular momentum $s, p_x, p_y, p_z, d_{xy}, \dots$. By using

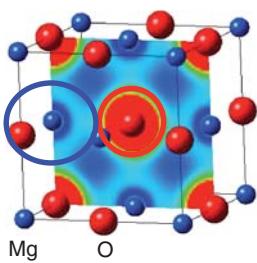


Figure 3.11 Sketch of the sphere radii for Mg (blue) and O (red) used to calculate the partial DOS of fcc MgO. Also shown is the valence charge density distribution $\rho(\)$ for one plane. The values for $\rho(\)$ are given in $e \text{ \AA}^{-3}$. Note that there is almost no valence charge density in the Mg sphere.

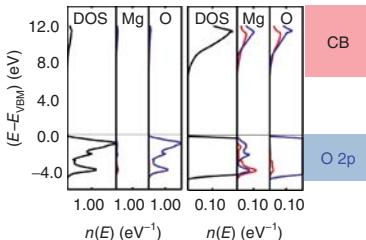


Figure 3.12 DOS and partial DOS for fcc MgO. The highest occupied state (valence band maximum, VBM) is the zero of the energy scale. The partial DOS for Mg and O are given for s (red) and p (blue) partial states, where the p partial states are the

sum over p_x, p_y , and p_z . The valence band maximum (VBM) is the zero of the energy scale. The partial DOS are given for two different scales in order to show more clearly the distribution of the partial DOS of the conduction states.

these partial charges to calculate the DOS, we illustrate the contributions of different “atomic orbitals” to the DOS. The partial DOS gives some insight into the character of the electron states as a function of energy, which is illustrated by the following example.

The first column in Figure 3.12 shows the total DOS for the highest occupied states and the lowest unoccupied states of fcc MgO. The next columns show, additionally, the s (red) and p (blue) partial DOS for Mg and O, where the first three and the last three columns only differ by the scale of $n(E)$. We see that the valence electron DOS is concentrated in the O sphere, indicating a predominantly ionic bond in accordance with the plotted valence charge density shown in Figure 3.11 and the experimental results [16]. Furthermore, the partial DOS of O below the Fermi energy (column 3 of Figure 3.12) has almost pure p character, that is, these are the 2p-like states of O. The 2p-like states of oxygen are fully occupied by six electrons because the Mg 3s-like states lie above the O 2p-like states and the valence electrons of Mg are transferred to oxygen. But we see from the last two columns in Figure 3.12 that the states of the CBs do not have pure 3s character of Mg but are more delocalized states with s and p amplitudes in both the O and Mg atomic spheres.

By combining Figures 3.10 and 3.12, we can identify which band belongs to which state. The three valence bands shown in Figure 3.10 lie in the energy range of the 2p-like states of O shown in Figure 3.12 and the lowest unoccupied states

are in principle the Mg 3s-like states and start at 6 eV above E_{VBM} . But these conduction states show a distinct admixture of other partial waves of Mg and O, and these states are distributed over the whole crystal forming extended, itinerant states.

This example shows that the partial DOS is very helpful in describing the energetic ordering of the states in oxides, although there is some ambiguousness in choosing the radii R of the spheres. We chose radii of 0.73 Å for the O spheres and 1.3 Å for the spheres around the metal site. We also see from the discussion that DOS and band structure yield complementary information; therefore, we plot them together in the following sections.

3.3.2.4 Crystal Field Splitting

Oxides $M^{IIA}O$ as MgO are the so-called *sp-compounds*, which means that the valence states can be described by s and p orbitals only. This is no longer true for transition metal oxides as the metal d states contribute to the chemical bond, see also Section 6.2.2. In this context, we encounter a simple example of a direct connection between *local* geometry and the ordering of the electronic states [3].

A free transition metal will have fivefold degenerate d states d_{xy} , d_{xz} , d_{yz} , d_{z^2} , and $d_{x^2-y^2}$, but if the metal atom is placed into a ligand field of oxygen, the degeneracy will be lifted. The octahedral symmetry is the most common and important case. By assuming a coordinate system wherein the ligands are located in the direction of the x , y , and z -axis, the d_{z^2} and $d_{x^2-y^2}$ states (now called e_g states after a two-dimensional irreducible representation of the octahedral group) are higher in energy than the d_{xy} , d_{xz} , and d_{yz} states (now called t_{2g} states), see Figure 3.13.

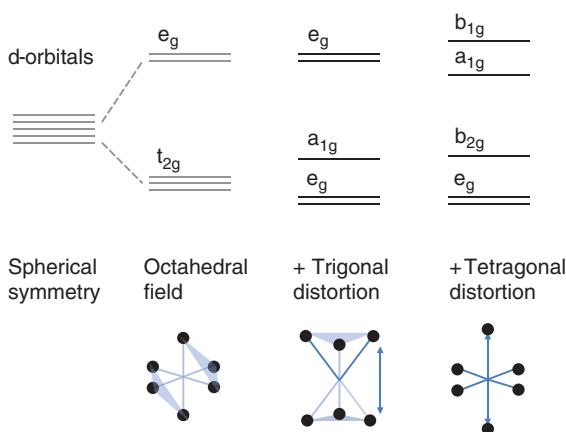


Figure 3.13 Crystal field splitting. The d orbitals split into e_g and t_{2g} states in an octahedral field. For an additional trigonal distortion (here compression), the t_{2g} states

split into e_g and a_{1g} . For a tetragonally distorted octahedral field (here, stretching), the degeneracy is lifted further.

A trigonal distortion of the octahedra, which is likely to occur if the octahedra are face-sharing, for example, in the corundum structure, leads to splitting of the t_{2g} states, see Figure 3.13 for the case of compression. A tetragonal distortion (e.g., elongation along the z -axis) lifts the degeneracy even further and both the e_g and t_{2g} level split.

For tetrahedral symmetry, the ordering of d states is inverse to the octahedral splitting, that is, t_{2g} states are higher in energy than e_g states. Usually, the splitting is less pronounced for tetrahedral symmetry than for octahedral symmetry, $\Delta_{\text{tet}} \approx 4/9\Delta_{\text{oct}}$ for the same charge of the metal ion and identical metal–ligand distances.

3.3.2.5 Exchange and Correlation

Most of the band structure investigations are performed on the basis of *density functional theory* (DFT) where the total energy of the crystal E is a functional of the electron density $\rho(\mathbf{r})$, $E = E[\rho(\mathbf{r})]$ [17]. It is assumed that $\rho(\mathbf{r})$ can be expressed as a sum over one-electron orbital probability densities $\psi_i(\mathbf{r})\psi_i^*(\mathbf{r})$: $\rho(\mathbf{r}) = \sum_i \psi_i(\mathbf{r})\psi_i^*(\mathbf{r})$. By using the variational principle, one-electron Schrödinger-like equations (called *Kohn–Sham* equations) for the orbitals $\psi_i(\mathbf{r})$ are obtained, see also Eq. 3.6. The corresponding potential acting on the i th electron,

$$V_i(\mathbf{r}) = V_{\text{ext},i}(\mathbf{r}) + V_{H,i}(\mathbf{r}) + V_{x,i}(\mathbf{r}) + V_{c,i}(\mathbf{r}), \quad (3.18)$$

includes the Coulomb potential of the nuclei $V_{\text{ext},i}(\mathbf{r})$ (called *external potential*) and the Coulomb potential of the electrons $V_{H,i}(\mathbf{r})$ (*Hartree potential*), which represents the interaction of the electron with the electron charge density $\rho(\mathbf{r})$:

$$V_{H,i}(\mathbf{r}) = \frac{e^2}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \quad (3.19)$$

The *exchange interaction* $V_{x,i}(\mathbf{r})$ takes into account the Pauli exclusion principle, and the *correlation term* $V_{c,i}(\mathbf{r})$ is due to the fact that the probability density of the electrons is dynamically correlated and cannot be described only by an effective cloud charge $\rho(\mathbf{r})$. For the calculations, the exact form of the *correlation potential* $V_{c,i}(\mathbf{r})$ is not known and one has to resort to various types of approximations [17], see Section 3.3.2.6.

For highly localized states, the strong Coulomb repulsion of the electrons results in strong electron correlations and the essential physics cannot be captured within band theory. To describe these highly correlated systems, the *Hubbard model* [18, 19] is mostly used. It is conceptually very simple but has to be described in the second quantization formalism, which we do not want to discuss here, but see also Section 6.5.1. Here, we only consider one aspect of the Hubbard model: it is found that transition metal oxides with partly filled d states can be insulators, although classical band theory predicts no optical band gap for a partly occupied band. One possible reason for this discrepancy can be described qualitatively by the Hubbard model, which is discussed next in a simple qualitative manner.

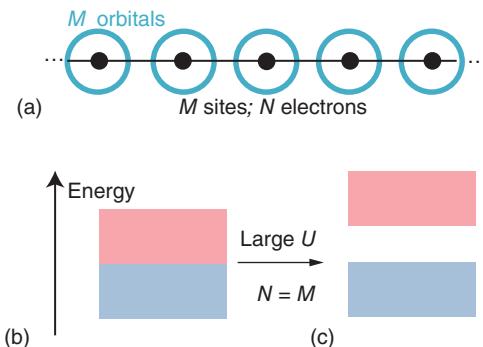


Figure 3.14 The Hubbard model, see the main text. (a) Chain of M atoms, one orbital at each site. For half-filled states, the behavior is metal-like if the kinetic energy parameter t is small as compared to the onsite

Coulomb repulsion parameter U , $t/U \gg 1$ (b), and for $t/U \ll 1$, there is an energy gap between the occupied and unoccupied states (c).

We consider a chain of M sites and only one orbital at each site, see Figure 3.14a. If the kinetic energy of the electrons, described in the Hubbard model by a parameter t , determines predominantly the energy levels of the electrons, the energy distribution of the levels is the same as shown in Figures 3.6 and 3.14b. There is no energy gap between the occupied and unoccupied states. However, for strongly correlated, localized electrons, the onsite repulsion described by the parameter U can be large, and for $U/t \gg 1$, the filling mechanism becomes different. If the number of electrons N is equal to M (half-filled states), all M orbitals are occupied by one electron. In band theory, the $(N+1)$ th electron can move to any site by taking only the Pauli principle into account. But for $U/t \gg 1$, the $(N+1)$ th electron has to overcome a large energy barrier, and therefore, in this case, the occupied and unoccupied states are separated by an energy gap called *Mott–Hubbard gap*, Figure 3.14c.

Therefore, there are two extreme limits for half-filling, that is, the number of electrons equals the number of sites:

- 1) $t \gg U$: *band limit*. The metallic behavior is recovered, as the band is only half-filled.
- 2) $U \gg t$: *atomic limit*. Hopping does not occur due to the high energy penalty and insulating behavior is exhibited.

In between these two limits, there exists a critical value of U , for which a transition occurs from a metallic to an insulating state: the *Mott transition*.

3.3.2.6 Computational Details

The expression for the exchange–correlation potential $V_{xc,i}(\mathbf{r}) = V_{x,i}(\mathbf{r}) + V_{c,i}(\mathbf{r})$ depends on the chosen functional $E[\rho(\mathbf{r})]$. There are many different empirical corrections/approximations for $V_{xc,i}(\mathbf{r})$ in the literature, which can be selected according to the system in hand [17].

In the computationally cheap local density approximation (LDA), $V_{xc,i}(\mathbf{r})$ depends only on the density $\rho(\mathbf{r})$. In this approach, the chemical bond is overestimated. This is corrected by using an additional term depending on the gradient of $\rho(\mathbf{r})$ (generalized gradient approximation, GGA), which, for example, describes metals very accurately. For studies on insulators, both approximations have the disadvantage that the size of the optical band gap is underestimated by about a factor of 2 and, therefore, the prediction of the energetic position of defect states compared to the VBM or CBM causes problems. The difference between the predicted optical band gap and the experimentally measured one is less pronounced when using orbital-dependent exchange–correlation potentials, for example, the exact exchange potential, that is, the Hartree–Fock expression for the exchange potential, hybrid potentials, GGA + U potentials, or others. A hybrid functional is a mixture of orbital dependent Hartree–Fock exchange and density-dependent exchange. In the GGA + U method, an additional orbital-dependent term with an empirical Hubbard U parameter is added in an *ad hoc* manner in the single electron Kohn–Sham equations.

Parallel or antiparallel spin alignment can be described in the single-electron *spin-polarized approach*. In this method, it is assumed that the exchange potential can be different for the spin-up and spin-down electrons ($s = \pm 1/2$), so $V_{x,i}(\mathbf{r})$ depends additionally on the spin s : $V_{x,i,s}(\mathbf{r})$.

In this work, we mainly use the screened exchange hybrid functional HSE06 [15] and an amount of exact exchange of 25% [20]. For metals, we use the GGA functional Perdew, Burke and Ernzerhof (PBE) [21], and for NbO₂, we use GGA + U [22]. All calculations were performed with the VASP code [23, 24], by using the projector-augmented wave (PAW) method [25] and an energy cutoff of 400 eV.

In the literature, the amount of exact exchange in the HSE06 functional is also varied to adjust the calculated band gap to the experimental value. We did not pursue this approach, as we are interested in the systematic change of the electronic structure, when going from one compound to another, and this is described most consistently with *one* parameter set for all calculations.

3.4

Material Classes and Characterization of the Electronic States

3.4.1

Metals

In metals (as in the free electron model), there is no band gap between the occupied and unoccupied electronic states. It follows that the electrons at the Fermi level can be excited to unoccupied states by an applied electrical field. As a result, metals display high electrical conductivities of typically 10⁵ S cm⁻¹ at room temperature.

Figure 3.15a shows the valence band structure of the simple Na metal for two directions in k -space. There is one Na atom with one valence electron per primitive

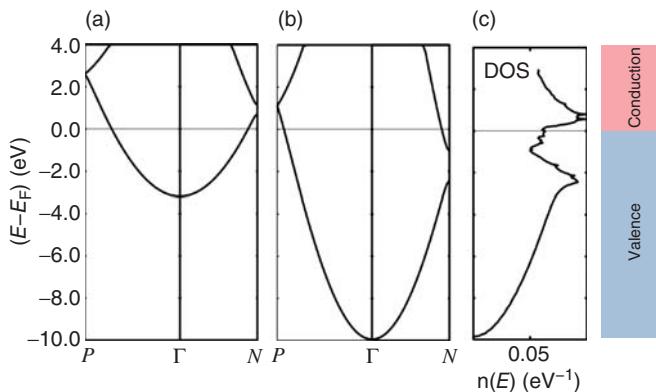


Figure 3.15 Electronic structure of metals. (a) Band structure of bcc Na metal for two directions in k -space, in reciprocal units from $P = (0.25, 0.25, 0.25)$ to $\Gamma = (0, 0, 0)$ to $N = (0, 0, 0.5)$. (b) Band structure of bcc Mg metal for the same two directions in

k -space. (c) Density of states (DOS) of bcc Mg metal (high-pressure phase). The Fermi energy is the zero of the energy scale. The lattice constants were determined experimentally [26, 27].

cell, and therefore, we have a partially filled band and the expected metallic behavior is achieved. One recognizes the deviation from the pure parabolic behavior of the free electron model only at higher energy values (at the points P and N).

Figure 3.15b shows as an example of the more involved band structure of bcc (high-pressure phase) Mg metal. As there is one Mg atom with two valence electrons in the primitive unit cell, one band should be totally occupied. However, as the different bands partly overlap, the first valence band is almost fully occupied and the second band is slightly occupied around the point N .

In metals, the electronic states show a large dispersion. For example, as shown in Figure 3.15b, the valence bandwidth (for a *single* band) is about 10 eV, whereas in oxides, the bandwidths (for a *single* valence band) are distinctly smaller, see MgO above (4 eV) and Al₂O₃ below (~ 2 eV).

Figure 3.15c shows the DOS for bcc Mg metal. For small values of E , one sees the parabolic behavior of the free electron gas ($n(E) \sim \sqrt{E}$), but for higher energies, there are distinct differences to the free electron model. The peaks in the DOS correspond to the regions in the band structure, where the band has a small slope.

The valence charge density distribution of Mg metal is shown in Figure 3.16. We see that, in the bonding region between the nuclei, the charge density is almost constant as predicted with the free electron model.

3.4.2

Semiconductors

The elements of the IVA group crystallize in an open crystal structure, for example, the diamond structure of Si. The covalent valence charge density looks different from that of metals. Figure 3.17 shows the valence charge density of Si. The valence

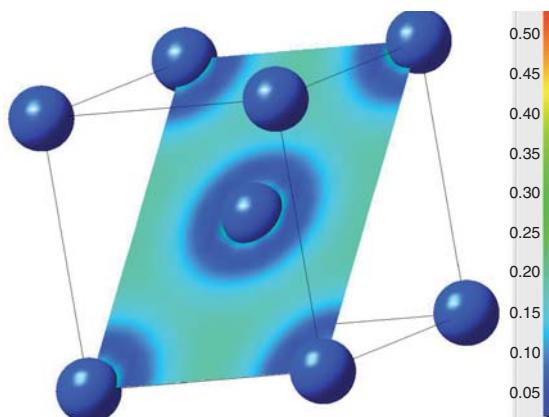


Figure 3.16 Valence charge density distribution (in $\text{e}\text{\AA}^{-3}$) of Mg metal (high-pressure phase) in the (110) plane of the bcc conventional cell. One recognizes the constant density in between the atoms.

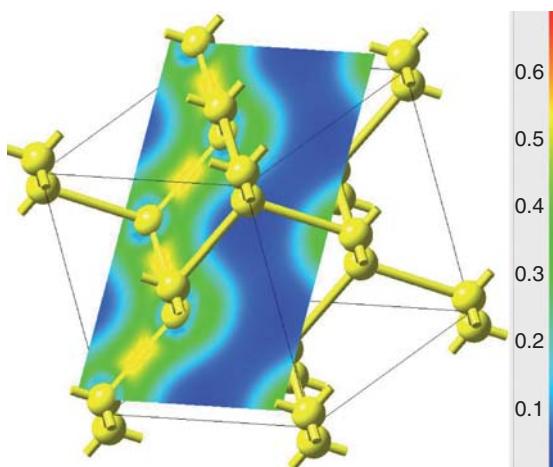


Figure 3.17 Valence charge density distribution (in $\text{e}\text{\AA}^{-3}$) of Si (diamond structure) in the (110) plane. The charge density is concentrated (yellow/green) in the nearest neighbor direction, whereas there is almost no density in the region interstitial region (deep blue).

charge density is concentrated along the nearest neighbors' directions, and there is almost no charge density in the interstitial region.

This kind of covalent bonds result in a small optical band gap E_{opt} due to the energetic separation of bonding and antibonding states, see the band structure and DOS in Figure 3.18a,b. E_{opt} is 1.1 eV for Si, 0.67 for Ge, and 1.43 for GaAs [2]. The bandwidths of the valence bands can be rather large (typically 4–8 eV), but still smaller than those in many metals.

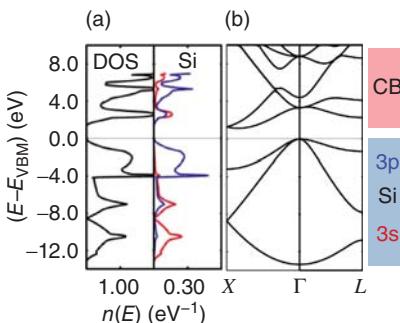


Figure 3.18 Electronic structure of Si calculated with the HSE06 functional. (a) Total DOS on the left and Si partial DOS on the right, which is subdivided into the s (red) and p (blue) contributions. (b) Band structure for two directions in k -space, in reciprocal units from $X = (0.5, 0, 0.5)$ to $\Gamma = (0, 0, 0)$

to $L = (0.5, 0.5, 0.5)$. The valence band maximum (VBM) is the zero of the energy scale. The Fermi level lies within the band gap. We can see an indirect band gap of 1.1 eV and a direct gap of 3.3 eV. The lattice constants were determined experimentally [28].

For electronic conductivity, the electrons have to be thermally excited to the lowest unoccupied band, the conduction band. In Figure 3.18b, we can observe that silicon has an *indirect band gap* of 1.1 eV (experimentally also 1.1 eV [29]), because the VBM is at the Γ -point, whereas the CBM is near the X -point. A transition of an electron from VBM to CBM necessarily has to involve phonons, which restore momentum conservation. The transition from the valence band to the conduction band at the Γ -point (the transition with zero momentum transfer $\Delta k = 0$) is much larger with a value of 3.3 eV (experimentally: 3.4 eV [30]).

For temperatures $T > 0$ K, the number of electrons in the conduction band has to be equal to the number of unoccupied states in the valence band (*electron holes*). From this and the Fermi–Dirac statistic for electrons, see Eq. 3.10, it follows that the Fermi level E_F has to lie within the band gap.

Another qualitative difference to metals is the change in the electronic properties of semiconductors due to the incorporation of dopants. Whereas in metals, impurities are sources of electron scattering and therefore deteriorate the electronic conduction, in semiconductors, they can serve as electron donors (acceptors) and improve the electronic (hole) conductivity. For example, a group IV semiconductor such as germanium (four valence electrons) can be doped with a group V element such as arsenic (five valence electrons). As only four electrons of the introduced arsenic atom will be involved in Ge–As bonds, the remaining electron will only be loosely bound to the impurity. This is due to fact that the Coulomb interaction of an additional charge in a solid is screened, that is, its value in vacuum is multiplied by $1/\epsilon_r$, with ϵ_r being the dielectric constant, for example, $\epsilon_r \approx 16$ for Ge. As a consequence, the localized state of the additional electron is located only slightly below the bottom of the conduction band (e.g., 0.013 eV for

As in Ge) [2]. It is, therefore, possible to thermally excite this electron into the conduction band already at room temperature ($k_B T \approx 0.026\text{ eV}$), thereby increasing the number of mobile negative charge carriers. This doping is called *n-type*. The Fermi level is now shifted toward the conduction band.

Analogously, germanium can be doped with group III element, for example, gallium. Then, one Ge–Ga bond is unsaturated, and an empty localized electron state is created. The energy of this hole state will be close to the top of the VBM. Now, electrons from the valence band can be thermally excited into this state, leaving behind a delocalized electron hole in the valence band. This empty state renders electron motion in the valence band possible, which can effectively be described by the motion of a positively charged electron hole in the opposite direction. Therefore, the number of mobile positive charge carriers is increased. This doping is called *p-type*, and the Fermi level is shifted closer toward the valence band.

3.4.3

Insulators

An insulator has a larger band gap than semiconductors, typically of more than 4 eV, and therefore, the intrinsic electronic conductivities at room temperature are very small (10^{-18} to $10^{-10}\text{ S cm}^{-1}$). Because of the large band gaps, these materials are sometimes called *wide-band-gap semiconductors*.

One can distinguish between several types of insulators [18]: (i) *band insulators*, (ii) *Peierls insulators*, (iii) *Anderson insulators*, and finally, (iv) *Mott insulators*.

The first two of them can be understood within the approximation of band theory. If there is no distortion or disorder in the system, the system is called a *simple band insulator*, for example, the already discussed compound MgO. If a 1D system is insulating due to a distortion along one axis as sketched in Figure 3.19, it is called a *Peierls insulator*. More generally, one can talk about electron–lattice coupling, which also includes the Jahn–Teller effect. The electron–lattice coupling can be important for narrow, half-filled d bands of transition metal oxides. A half-filled band of the undistorted structure can be split under distortion into two bands: one fully occupied and the other empty, both separated by a band gap. Thus, one gets a metal–insulator transition [31], see Section 3.5.4.1.

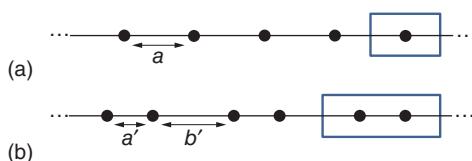


Figure 3.19 One-dimensional chain. (a) Undistorted chain with primitive unit cell. (b) Chain with dimerization: the unit cell doubles.

An *Anderson insulator* can be realized in a random alloy or an amorphous state, when the randomly distributed atoms or, in general, the structural disorder leads to more localized electron wave functions [32, 33]. This is shown in Figure 3.22b,c and discussed in Section 3.4.6.

Mott insulators are characterized by localized electronic states with strong Coulomb repulsion of the electrons (strong correlation) and can approximately be described by the Hubbard model. With the simple band theory, a Mott insulator is usually predicted to be a metal, but this can depend on the chosen model and the system. For NiO, LDA, or GGA, calculations predict only a very small gap due to the antiferromagnetic order (0.3 eV [34]). But if orbital-dependent terms are incorporated in the description, either the Hubbard U parameter in the GGA + U method [22, 35, 36] or an orbital-dependent exchange potential (HSE03 [37] or, in this work, HSE06, see Section 3.5), one obtains an insulator with a large band gap (>4 eV), which is in much better agreement with the experimental result. We want to stress that Mott insulators and Mott transitions (transitions from a metallic state to a Mott insulator) are still a subject of many investigations.

3.4.4

Point Defect States

Intrinsic point defects are missing atoms or interstitial atoms in the crystal lattice. An important example is the oxygen vacancy, which is an unoccupied oxygen lattice site. If a neutral oxygen atom is removed from the lattice, we have a neutral defect, which is written as V_O^X in Kröger–Vink notation [38]. If an O^- is removed, we have a singly positively charged oxygen vacancy, denoted as V_O^+ . In an ionic solid, often O^{2-} ions are missing, and therefore, we obtain a doubly positively charged oxygen vacancy V_O^{2+} . Further point defects such as cation vacancies/interstitials are beyond the scope of this section, and we refer the interested reader to Chapter 4 and to the corresponding literature [39, 40].

Point defects usually create localized defect states in the crystal. To describe the electronic states of an isolated point defect within the band structure approach (with periodic boundary conditions), large unit cells (*supercells*) with one defect are used for the calculation. It shall be pointed out that for the supercell calculations, we had to use, for computational reasons, a different exchange–correlation potential (GGA instead of HSE06) as compared to the small primitive cells. Therefore, the sizes of the optical band gaps shown in Figures 3.20 and 3.10 are different.

Figure 3.20 shows an isosurface of the charge density distribution of a *color center*, an O vacancy V_O^X , which is occupied by two electrons. One can clearly observe the localized state around the vacancy site. The total DOS shows the energetic position of the defect state in the band gap of the host material, see Figure 3.20. The narrow bandwidth of the defect state reflects its localized nature.

The oxygen vacancy V_O^X can act as a donor (in analogy to doping in semiconductors, see Section 3.4.2) if the introduced defect states are close to the CBM.

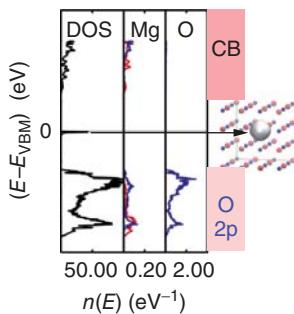


Figure 3.20 Oxygen vacancy V_O^\times in MgO (63 atom cells of $Mg_{32}O_{31}$) calculated with the GGA functional. Left: DOS and partial DOS for the cell with a vacancy V_O^\times . The highest occupied state (valence band maximum, VBM) is the zero of the energy scale. For the partial DOS, red denotes s-like state and blue denotes p-like state. The localized states of the color center lie in the band gap of the bulk phase MgO. Right: Isosurface of the charge density distribution of the two localized electrons at the vacancy position.

However, in MgO, the occupied defect states of V_O^\times lie in the middle of the large gap, and these *deep donor* states are not sources of electrons.

Introducing a defect also causes local relaxations of the lattice. In our case, for example, the nearest neighbor Mg atoms move outward by only 0.01 Å.

3.4.5

Surface States

Crystal surfaces or interfaces have been investigated in detail for many years due to their technological importance in catalysis, fuel cell applications, or resistive switching metal–insulator–metal heterostructures, which is the topic of this book, just to name a few. Here, we focus on surfaces and briefly discuss the perfect (001) surface of MgO (without steps) and its impact on the electronic structure [41, 42].

On crystal surfaces, the coordination number of the ions is smaller than that of the bulk, which is accompanied by a reduced Madelung potential [43]. As a result, various types of surface states, as well as a smaller band gap, emerge. Another important point is that on the surface, the formation energies of the defects can be very different from the bulk, leading to different concentrations of defects compared to the bulk.

Simulations of surfaces by band structure investigations are performed by the so-called slab calculations. Figure 3.21 shows one example. To study the states of the (001) surface of MgO, a tetragonal unit cell is chosen with 36 layers of MgO with a vacuum of 10 Å width, as sketched in Figure 3.21a. The corresponding tetragonal unit cell for bulk MgO is shown in Figure 3.21b: the lattice vectors \mathbf{a} and \mathbf{b} (in the x – y plane) are identical to the surface calculation. In contrast to the primitive unit cell, the tetragonal bulk cell has two formula units MgO. Therefore, there are six O 2p-like bands as shown in Figure 3.21e. The band structure of the slab calculation is shown in Figure 3.21d. Because of the long extension of the cell in the z -direction, the corresponding BZ is very narrow in the k_z -direction. By comparing Figure 3.21d and e, we can detect three effects of the surface: (i) there are several subbands, as the total symmetry of the cell is lowered; (ii) the width of the 2p-like valence states remains more or less the same; and (iii) the band gap is

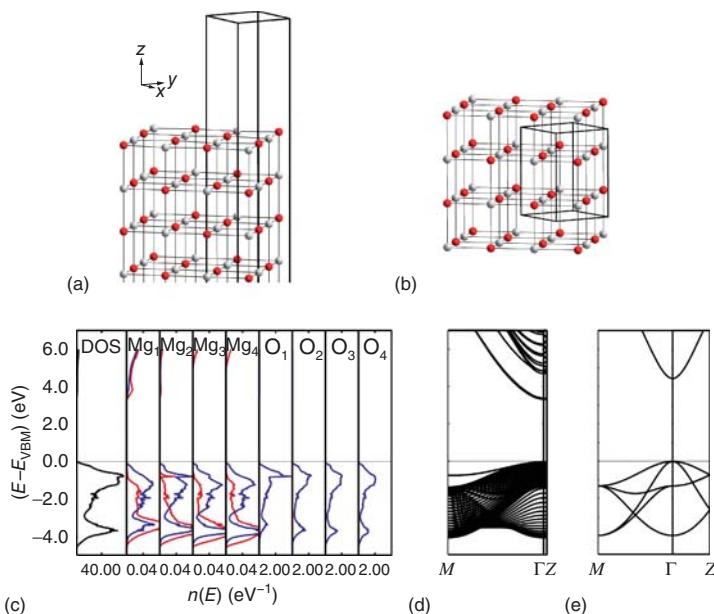


Figure 3.21 (a) Structure of the (001) surface of MgO. The tetragonal simulation cell is indicated by the thick black lines. On the surfaces, one can recognize *rumpling*. (b) Structure of bulk MgO. The tetragonal simulation cell (with the same lattice vectors and as for the surface cell) is indicated by the thick black lines. (c) DOS and partial DOS of the MgO surface cell. The partial DOS for O and Mg are given for the surface layer (O_1) and the layers below the surface

layer, where bulk behavior is recovered for the second layer of oxygen and third layer of Mg. (d) Corresponding band structure for the cell with 100 surface from $M = (0.5, 0.5, 0)$ to $\Gamma = (0, 0, 0)$ to $Z = (0, 0, 0.5)$. (e) Band structure of the tetragonal bulk cell of MgO with two formula units from $M = (0.5, 0.5, 0)$ to $\Gamma = (0, 0, 0)$ to $Z = (0, 0, 0.5)$. The valence band maximum (VBM) is the zero of the energy scale and the calculations have been performed with the GGA functional.

reduced (from 4.5 to 3.3 eV). The lowest unoccupied bands describe the surface states, and they are responsible for the smaller band gap (*band tailing*).

The surface character of the lowest conduction states can clearly be seen from the partial DOS shown in Figure 3.21c. The partial DOS are given for Mg and O atoms in the different layers perpendicular to the surface, starting from the outermost layers (1 = surface, ..., 4 = bulk). Note the different scales for the O and Mg partial DOS. Contributions to the lowest conduction states are only found for the upper two atomic layers. By analyzing the charge distribution (not shown), one finds that the highest occupied band also describes the pure surface states, but this is not readily seen from the partial DOS. The fact that the unoccupied surface states are lower in energy as compared to the bulk can be explained by simple electrostatic reasoning. The Madelung energy is smaller for surface ions, and this energy contribution is the origin of the opening of the gap between oxygen 2p states and Mg 3s states [41].

We can also observe some *rumpling* on the relaxed surface shown in Figure 3.21a, that is, the O atoms are moved slightly out of the surface, whereas the Mg atoms move into the bulk, as reported in the literature [42].

3.4.6

Amorphous States

Chemical or structural disorder has significant effects on the electronic structure of a material. Point defects or surfaces introduce some modifications of the band structure, as seen earlier. In the extreme case, that is, in an amorphous material without any long range order, we do not have a band structure any more, as a Bloch wave, Eq. 3.11, is linked to the translational symmetry of a system. Nevertheless, notions such as “band gap” or “band structure” are still applied, but in a broader sense, as compared to crystalline materials. However, still, the charge density and the partial DOS can characterize the electronic states.

As an example, we consider amorphous MgO. Theoretically, an amorphous system can be studied using the supercell approach by performing a molecular dynamics (MD) simulation at high temperatures (above the melting temperature) and then abruptly quenching the system. In this way, the amorphous MgO system shown in Figure 3.22c has been created. For the obtained configuration, the resulting DOS and partial DOS for a few ions are displayed in Figure 3.22a. Note again the different scales for the O and Mg partial DOS. In contrast to the crystalline bulk phase, the partial DOS are different for all atoms and the gap between filled and unfilled states has become distinctly smaller than the band gap for the bulk system, see Figure 3.21e for comparison.

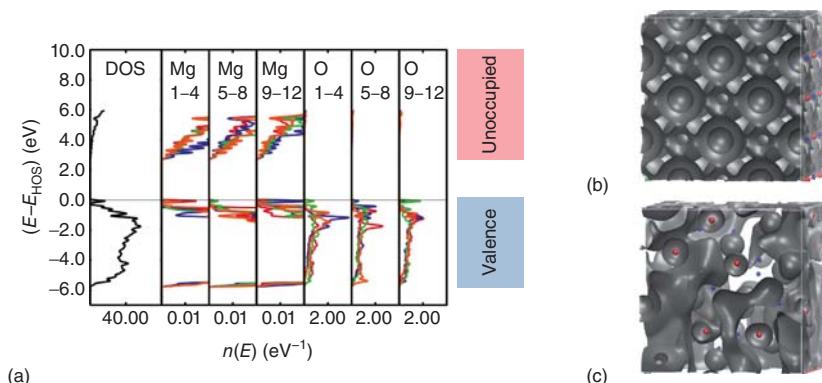


Figure 3.22 Amorphous MgO. (a) DOS and partial DOS for simulated amorphous MgO. The highest occupied state is the zero of the energy scale. The partial DOS are shown for 12 different O and Mg spheres distinguished by the different colors. (b) Charge

density of the first unoccupied states (CBM) in crystalline MgO. One clearly sees that the density is uniformly distributed over the whole crystal (c) charge density of the first unoccupied states in amorphous MgO. The charge is distributed in a nonuniform way.

By comparing the charge densities of the first unfilled states for crystalline and amorphous MgO in Figure 3.22b and c, we clearly see that the charge is distributed nonuniformly in the amorphous state. Localization of the wave function due to disorder, as can be partly seen in Figure 3.22c, is a common phenomenon in amorphous materials (Anderson localization [32]).

3.5

Electronic Structure of Selected Oxides

In this section, we present the electronic structure of some of the oxides listed in Table 3.1. All oxides discussed in this section are investigated extensively in the literature with various different methods, both theoretically and experimentally. Our theoretical results are deduced from DFT band structure calculations using mainly the hybrid exchange–correlation potential HSE06, see Section 3.3.2.6 on computational details, to study the systematic variations between the systems. All band structures and DOS were calculated for the primitive cells of the compounds.

3.5.1

Nontransition Metal Oxides

We have already discussed the electronic structure of MgO, also in the context of defects, surfaces, and amorphous structure. In this section, we discuss other non-transition metal oxides: Al_2O_3 , SrO, and ZnO. Al_2O_3 (α -phase) has the corundum structure, SrO, similarly to MgO, crystallizes in the NaCl structure and ZnO in the wurtzite structure, see Table 3.1. We consider the ordering of the electronic states in the energy range from the 2s-like states of O to the lower conduction bands.

3.5.1.1 Al_2O_3

As another example for a *simple band insulator*, we show the band structure and DOS of $\alpha\text{-Al}_2\text{O}_3$ (corundum) in Figure 3.23. The primitive cell of alumina contains four Al atoms and six O atoms. The ordering of the states is very similar to that of the band insulator MgO. The valence states in the range from 0 to -7 eV are oxygen 2p-like states, see the DOS in Figure 3.23a, and the Al has transferred its 3p and 3s valence electrons to oxygen, resulting mainly in an ionic bond with fully occupied oxygen 2p orbitals. Upon counting, these partly degenerate valence bands are 18 in total, which is the number of 2p orbitals (p_x, p_y, p_z) times the number of oxygen atoms in the cell. The 18 valence bands result in a total width of the valence bands of 7.3 eV , which compares well with the experimentally determined width of 8 eV [45]. The lowest unoccupied states are in principle the Al 3s-like states. But these conduction states, as can be seen in the DOS in Figure 3.23a, show a distinct admixture of other partial waves of Al and O, and these states are more itinerant rather than being metal 3s states, similar to MgO, see above page 64. The calculated band gap of $\alpha\text{-Al}_2\text{O}_3$ is 8.1 eV , which is in good agreement with the experimental value of 8.8 eV [46].

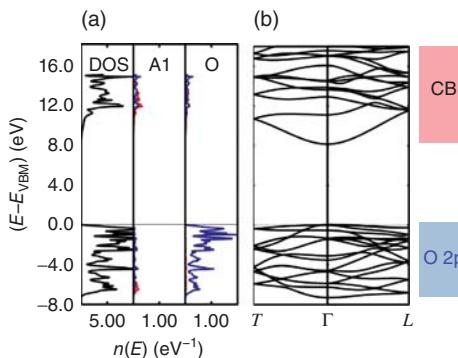


Figure 3.23 Electronic structure of Al_2O_3 (corundum structure) calculated with the HSE06 functional. (a) Total DOS on the left and Al and O partial DOS on the right, which are subdivided into the s (red) and p (blue) contributions. We show only one Al and one O, as the other Al and O in the cell are symmetrically equivalent to the

displayed ones. The lowest displayed states are the valence bands with oxygen 2p character. (b) Corresponding band structure. The direct band gap is 8.1 eV. The valence band maximum (VBM) is the zero of the energy scale. (The lattice constants were determined experimentally [44].)

3.5.1.2 SrO

SrO crystallizes in the same (fcc) structure as MgO . The band structure and DOS of SrO are shown in Figure 3.24. We see the following differences of SrO as compared to MgO and also Al_2O_3 : (i) In SrO , we have 4d-like states at the bottom of the conduction band, see the green partial DOS for Sr in Figure 3.24a. (ii) SrO has an *indirect band gap* of 4.8 eV from Γ to X , see Figure 3.24b. The direct gaps at Γ and X are 5.3 and 5.0 eV and slightly larger. Our calculated direct gaps at Γ and X are smaller than the corresponding experimental values: 6.1 and 5.8 eV [48]. There are reports in the theoretical literature with regard to both lowest direct and indirect band gap [49]. (iii) The total width of the valence bands is roughly 2 eV and, therefore, considerably smaller than that for MgO and Al_2O_3 .

3.5.1.3 ZnO

Bands considering the structure and electronic structure, there are differences between ZnO and the oxides discussed earlier. First, the coordination numbers are different, see Table 3.1, and second, for the $\text{M}^{\text{IIB}}\text{O}$ oxide ZnO , the fully occupied 3d shell distinctly influences the chemical bond. There are two formula units in the primitive unit cell, and therefore, there are two 2s-like bands and six 2p-like bands of oxygen. In Figure 3.25, we have plotted the DOS and band structure of ZnO . The ordering of the states from the bottom is $E(\text{Zn 3d}) < E(\text{O 2p}) < E(\text{CB})$. There is only a small band gap between the Zn 3d-like states around -6 eV and mainly O 2p-like states in the region from -5 eV to the VBM, see the DOS in Figure 3.25a. Furthermore, there is some hybridization of the Zn 3d- and O 2p-like states, indicating that there is a distinct covalent bond character in ZnO . The bands at the CBM are a mixture of Zn 3d-like and O 2p-like states. The calculated

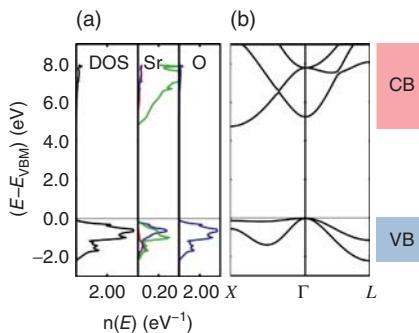


Figure 3.24 Electronic structure of SrO (rock-salt structure) calculated with the HSE06 functional. (a) Total DOS on the left and O and Sr partial DOS on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. Notice: the scaling of the Mg partial DOS and that of O partial DOS differ by a factor of 10. The

valence band has oxygen 2p character, and at the CBM, we have Sr 3d-like states. (b) Corresponding band structure. The indirect band gap is 4.8 eV. The valence band maximum (VBM) is the zero of the energy scale. The experimental lattice constant is used [47].

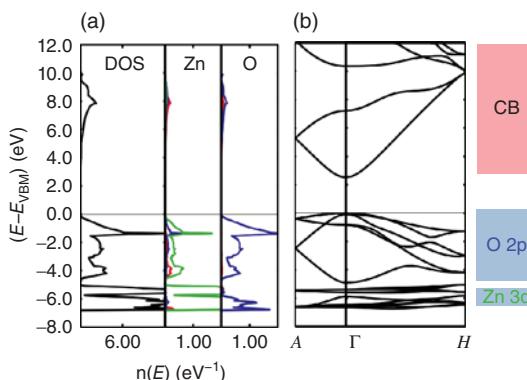


Figure 3.25 Electronic structure of ZnO (wurtzite structure) calculated with the HSE06 functional. (a) Total DOS on the left and Zn and O partial DOS on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. The lowest displayed states are the Zn 3d states.

The top of the valence band has mainly oxygen 2p character. (b) Corresponding band structure. The direct band gap is 2.5 eV. The VBM is the zero of the energy scale. The experimental lattice constant is used [50].

direct band gap is 2.5 eV, which is somewhat smaller than the experimental value of 3.4 eV [51], but it compares well with other theoretical calculations [52]. For hybrid functional studies on defects in ZnO, the amount of exact exchange (in our case, 25%, see Section 3.3.2.6) has been increased in the literature to a value of 37.5% to match the experimental band gap [53].

3.5.2

Titanates

We want to demonstrate the trends in the electronic structure of different oxides Ti_mO_n , as these compounds are among those that are of importance for resistive switching applications. Therefore, we have included a hypothetic ideal defect-free fcc TiO , although this compound only exists in a defective phase. Even for its “stoichiometric” composition, there are around 15% of both oxygen and titanium vacancies [3]. In addition to the 4s valence electrons, Ti has two electrons in the 3d shell and exists in different oxidation states: 2+ (TiO), 3+ (Ti_2O_3), and 4+ (TiO_2 , SrTiO_3). It follows that the 3d-like states are partially occupied for Ti^{2+} and Ti^{3+} . We will, therefore, also discuss Ti_2O_3 , TiO_2 , and the perovskite SrTiO_3 to show systematic changes of the electronic structure by changing both the crystal structure and the number of 3d electrons in the compound.

3.5.2.1 TiO

Figure 3.26 shows the DOS and band structure of defect-free hypothetical TiO with fcc structure and one formula unit in the primitive unit cell. Because of the 3d states, the ordering of states in TiO (a d^2 -compound) is different from the already discussed oxides. As for MgO , there are fully occupied 2p-like states. Also, the widths of these bands are similar to that of MgO . However, the highest occupied states are not the O 2p-like states. Above the O 2p-like states, we find the Ti 3d-like states, which are partly occupied, and there is a gap between the O 2p- and the Ti 3d-like states of about 4 eV. The 3d-like bands are very broad, see Figure 3.26b, and the total bandwidth ΔE_{3d} is about 9 eV. The different 3d states due to the crystal field splitting (t_{2g} , e_g) overlap, and there is no band gap between the occupied and unoccupied states. But the states at the Fermi energy do not have itinerant character and are more localized Ti 3d-like states. Our calculations agree with other calculations in the literature [55, 56].

3.5.2.2 Ti_2O_3

The ordering of the electronic states in Ti_2O_3 (two formula units in the primitive cell) is the same as in TiO , see Figure 3.27a. First, we have the oxygen 2p-like bands from -5 to -10 eV, followed by the 3d-like states of Ti. However, in corundum-type Ti_2O_3 (we have Ti^{3+} , so it is a d^1 -compound), there are trigonally distorted face-sharing octahedra along the c -axis, which causes additional splitting of the partially filled t_{2g} band into an a_{1g} (two states) and an e_g (four states) band [3], see also the section on crystal field splitting. Due to the small distance of the Ti-atoms along the c -axis, a metal–metal bond forms in Ti_2O_3 with bonding and antibonding states. As there are two electrons per bond, the bonding states are low in energy and filled, and the antibonding states are empty and higher in energy. So, there is a small gap of 0.4 eV (experimental value: 0.1 eV [58, 59]) opening in the d band due to the formation of a Ti–Ti bond in Ti_2O_3 .

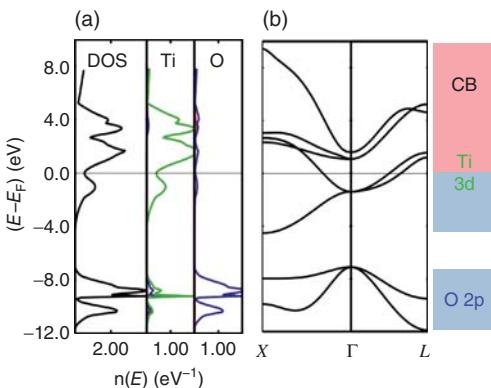


Figure 3.26 Electronic structure of TiO (rock-salt structure) calculated with the HSE06 functional. (a) Total DOS on the left and partial DOS of Ti and O on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. The lowest displayed states are the O 2p-like

states. Above these states lie the Ti 3d-like states. These states are occupied up to the VBM, which is the zero of the energy scale. The actual sp-conduction states are located above the Ti 3d-like states. (b) Corresponding band structure. The experimental lattice constant is used [54].

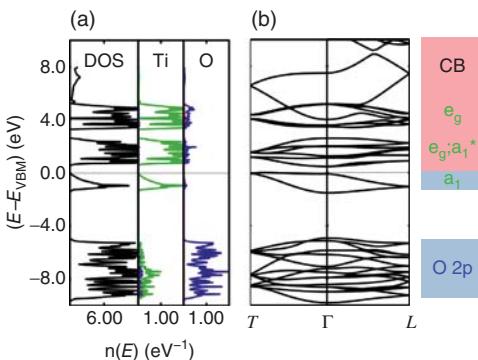


Figure 3.27 Electronic structure of Ti_2O_3 (corundum structure) calculated with the HSE06 functional. (a) Total on the left and partial DOS of Ti and O on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. One clearly sees

the occupied a_{1g} states from -1 eV to the VBM. (b) Corresponding band structure. The VBM is the zero of the energy scale. The experimental lattice constant is used (low-temperature phase) [57].

We clearly see in the partial DOS of Ti (Figure 3.27a) the occupied bonding a_{1g} states from -1 eV to the VBM, which are d_{z^2} -orbitals, that is, oriented along the z -direction (c -axis) of the corundum structure, as expected from the dimerization. Higher in energy and unoccupied are the e_g states (the remaining states from split t_{2g} band) and the antibonding states, indicated on the very right of the plot. From 4 to 5 eV, we have the upper e_g states from the trigonally distorted octahedral field.

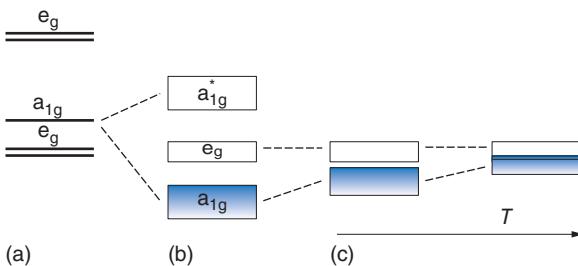


Figure 3.28 Metal–metal bonding in Ti_2O_3 .
 (a) Due to the trigonally distorted octahedra in Ti_2O_3 , the crystal field splitting yields low lying a_{1g} and e_g levels. (b) The metal–metal bonding of the Ti along the c -axis is reflected in bonding and antibonding a_{1g} levels. (c) With increasing temperature

T , the lattice parameter increases and the metal–metal overlap decreases. Therefore, the bands become narrower, and the bonding level is shifted upward and finally overlaps with the e_g level. The material becomes metallic.

In the band structure shown in Figure 3.27b, we see that there is an indirect gap between a point near T and Γ . We have two a_{1g} bands as we have four Ti in the primitive cell. Our results agree well with the calculations in the literature [60], which have been performed with another functional.

The above explained scenario only holds for temperatures below 390 K [3]; above this temperature, a gradual metal–insulator transition (MIT) is observed. The lattice expansion reduces the metal–metal overlap, and therefore, for higher temperatures, the Ti–Ti distance is considerably larger than that for low temperatures [61]. The bonding states increase in energy until there is a band-crossing of the a_{1g} and empty e_g states and Ti_2O_3 becomes metallic for $T > 390$ K, see the schematic plot in Figure 3.28. So, crystal field splitting alone is not responsible for the gap due to the bandwidth of the a_{1g} and e_g bands. The additional lowering of the energy of the a_{1g} band due to the formation of a metal–metal bond is at the origin of the small band gap (*Zandt–Honig–Goodenough model* [62]).

3.5.2.3 TiO_2

In rutile TiO_2 (two formula units in the primitive cell), the ordering of the states is the same as for TiO and Ti_2O_3 , as can be seen in Figure 3.29. But again the occupation of the states differs. In TiO_2 , the oxidation state is +4 and all four valence electrons of Ti are formally transferred to O; see also Section 6.4.2. Therefore, the 3d-like states of Ti are unoccupied (TiO_2 is a d^0 compound) and are the lowest conduction states, see Figure 3.29a. Nevertheless, we can see some hybridization of Ti 3d states and O 2p states in the valence band, which is also confirmed experimentally [3, 64]. We find a *direct band gap* of 3.2 eV, which is in good agreement with the literature [65, 66], as the energy of the conduction band at the R point, which is not shown here, is slightly higher than at the Γ point. Our calculated gap is close to the experimentally determined gap of 3 eV [67]. The total valence bandwidth is 6.2 eV, which is in good agreement with the experimental value of 5–6 eV determined by XPS [68, 69] and XES [70]. In the partial DOS of Ti shown

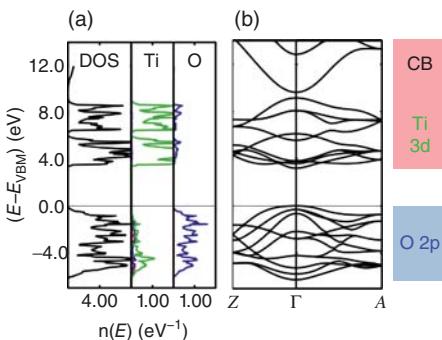


Figure 3.29 Electronic structure of TiO₂ (rutile structure) calculated with the HSE06 functional. (a) Total DOS and partial DOS of Ti and O, which are subdivided into the s (red), p (blue), and d (green) contributions. The partial DOS for the different O are very similar. Therefore, we have plotted only one

O. The lowest displayed states are the O 2p-like states between –6 and 0 eV. The conduction band is mainly of Ti 3d character and the calculated band gap is 3.2 eV. (b) Corresponding band structure. The VBM is the zero of the energy scale. (The experimental lattice constant is used [63].)

in Figure 3.29a, we can clearly see two blocks in the d states due to the crystal field splitting [71]: from 3 to 6 eV are the t_{2g}-derived bands, which are lower in energy as compared with the e_g-derived bands from roughly 7 to 9 eV. As the symmetry is not a full octahedral symmetry, the bands are split further [72].

3.5.2.4 SrTiO₃

In SrTiO₃, Sr has the oxidation state +2 as in SrO and Ti +4 as in TiO₂. Therefore, one might expect that the partial DOS for SrTiO₃ is some kind of a mixture of partial DOS for SrO and that for TiO₂. Indeed, one gets for SrTiO₃, the ordering of states expected from the study of SrO and TiO₂, see Figure 3.30a. First, we want to mention that the Sr 4p-like states below the O 2p-like states (not shown in Figure 3.30) are slightly shifted to higher energy values in SrTiO₃ as compared to SrO. These states correspond to a broad structure in XPS data in this energy range [74]. The width of mainly the oxygen 2p-like valence states between –5 and 0 eV is almost the same as in TiO₂ and also consistent with the experimental data [75]. In Figure 3.30b, we find an *indirect* minimal band gap between the oxygen 2p band at R and the Ti 3d band at Γ , which is in agreement with the literature [49, 76]. Our calculated band gap value of 3.3 eV (experimental value also 3.3 eV [77, 78]) is again very close to the value for rutile TiO₂. The Ti 3d states are followed by the Sr 4d states above 8 eV, as can also be expected from the larger band gap of SrO as compared to TiO₂.

3.5.3

Magnetic Insulators

Until now, we have dealt with nontransition metal oxides (MgO, Al₂O₃, SrO, ZnO) or transition metal oxides with a small number of d electrons, such as

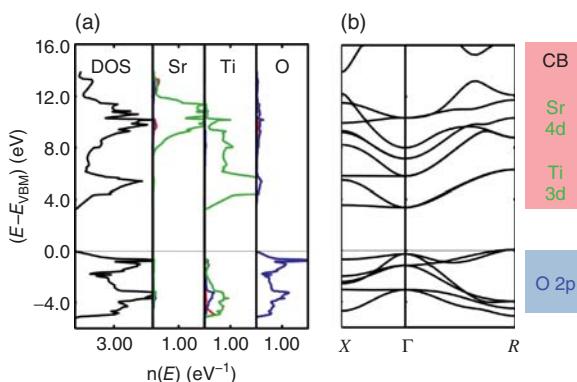


Figure 3.30 Electronic structure of SrTiO₃ (perovskite structure) calculated with the HSE06 functional. (a) Total DOS and partial DOS of Sr, Ti, and the three oxygen, which are subdivided into the s (red), p (blue), and d (green) contributions. The lowest

displayed states are the O 2p-like states, followed by the unoccupied Ti 3d-like states in CBM. The calculated band gap is 3 eV. (b) Corresponding band structure. The VBM is the zero of the energy scale. (The experimental lattice constant is used [73].)

titanates. Increasing the number of d electrons in a partially filled d band is equivalent to increasing the complexity of the electronic system and entering the field of strongly correlated systems, see Section 3.3.2.5. As examples we will consider MnO and NiO in more detail after making some more general remarks on *magnetic insulators*.

MnO, defective FeO_{1-x}, CoO, and NiO are called *magnetic insulators*. This is due to the fact that they are semiconductors with a pronounced band gap and that they show *antiferromagnetic order*, which can be explained by a mechanism called *superexchange* [19]. This mechanism is partly linked to the rock-salt structure: metal atoms with localized d orbitals and oxygen atoms with p orbitals are located alternately along lines in this structure, see Figure 3.31. As there is some overlap between the orbitals, the electrons can be transferred from the oxygen to the metal. This virtual hopping of electrons lowers the energy. Only if the two neighboring cations have an electron configuration with different spin directions (due to the Pauli exclusion principle), it is possible for the electrons in the filled p orbital of the oxygen to hop to both neighboring metal d orbitals. Therefore, antiferromagnetic order is energetically favored, see also Figure 3.31. MnO, FeO_{1-x}, CoO, and NiO display the so-called type II magnetic ordering, which is characterized by parallel spins in the (111) planes of the lattice and spins in adjacent sheets are antiparallel [3].

Due to orbital contraction with higher charge Z of the nucleus in the 3d series, the repulsion between the localized electrons becomes more and more important, and this interaction cannot be properly described by LDA or GGA. This Coulomb repulsion is the origin of the behavior of highly correlated electrons, which can be captured by the Hubbard model, as already mentioned. There is a gap between filled and unfilled d states corresponding to the excitation reaction (e.g., for NiO,

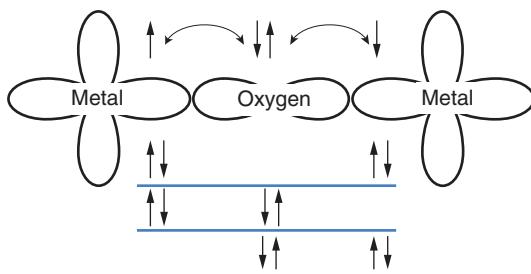


Figure 3.31 Schematic plot of superexchange. The up and down arrows represent the up and down spin electrons, respectively, and the upper line of arrows corresponds to the *antiferromagnetic order*. Based on the Pauli exclusion principle, only in this

spin configuration, the electrons can be transferred between the orbitals (without spin flip) to yield the spin configurations in the lines below. These additional configurations lower the energy of the system and stabilize the antiferromagnetic order.

where only the occupancy of the d orbitals is shown, that is, d^8 is the abbreviation for eight electrons in the d orbitals):



So, two Ni atoms with eight electrons in the d shell react to one Ni atom with seven and one Ni atom with nine electrons. It is essentially the d^9 configuration, which has the highest energy cost, as it incorporates the additional Coulomb repulsion (Hubbard U) of the localized electrons, which is therefore responsible for the gap between occupied and unoccupied d states [3, 34, 36, 37].

3.5.3.1 NiO

In Figure 3.32, we show the DOS and band structure of antiferromagnetic NiO (d^8 compound) with two formula units and two nonequivalent Ni atoms in the primitive cell. We have performed *spin polarized* calculations, that is, each band can be occupied by just one electron with spin up or spin down. In Figure 3.32a, we see the DOS and partial DOS for O and Ni. For Ni, the partial DOS are shown for the spin-up and down states. One obtains antiferromagnetic ordering: there are more spin-up states occupied than spin-down states for the first Ni atom Ni_1 , whereas the opposite holds for Ni_2 . A strong hybridization of Ni 3d and O 2p states at the top of the valence band can be seen, which has been reported in many different theoretical calculations. Magnetic insulators with considerable 2p character in the valence band as in NiO are also called *charge transfer insulators* [80]. The total width of the valence band is 7.4 eV, which compares well with the experimentally determined width of 8–8.5 eV as deduced from XES data [81]. The unoccupied d states in the conduction band are e_g states [37], as expected from crystal field splitting. We find an indirect minimal band gap from T to Γ of 4.6 eV, see Figure 3.32b, which is in reasonable agreement with the experimental gap of 4.0 eV [81]. Other theoretical calculations [37, 82] also report indirect minimal gaps, for example, with the HSE03 functional. However, in some calculations, the

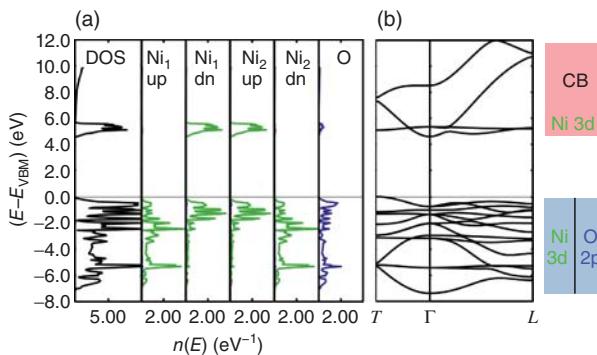


Figure 3.32 Electronic structure of NiO (rock-salt structure with antiferromagnetic ordering of type II) calculated with the HSE06 functional. Here, we perform a *spin polarized* calculation, that is, each band can be occupied by just one electron. (a) Total DOS and partial DOS of Ni (spin up and down) and oxygen, which are subdivided into the s (red), p (blue), and d (green) con-

tributions. The lowest displayed states are the valence bands formed by a considerable hybridization of oxygen 2p and Ni 3d states. The calculated band gap is 4.6 eV and the net magnetic moment on the Ni is $1.6 \mu_B$. (b) Corresponding band structure. The VBM is the zero of the energy scale. (The experimental lattice constant is used [79].)

CBM is at the K -point, whereas in our calculation, we find the energy of the lowest conduction band at this point to be 0.1 eV higher than that at the Γ -point. The magnetic moment on the Ni atom in our calculation is $1.6 \mu_B$, which is also in good agreement with the experimental value of $1.6\text{--}1.9 \mu_B$ [83–85].

In principle, antiferromagnetic order could also lead to a band insulator, as it doubles the unit cell, and therefore, a previously half-filled band can be split into two bands: one filled and one empty with a gap in between, in analogy to the arguments brought forward in Section 3.4.3 and the opening of a gap for linear chain of atoms in a dimer-like configuration. Nevertheless, for example, in NiO, the antiferromagnetic order is not a prerequisite for a gap as the band gap remains essentially the same above the *Néel temperature* in the paramagnetic phase, which can also be described by a spin-polarized calculation; see also Section 6.5.3.

3.5.3.2 MnO

In Figure 3.33, we show the DOS and band structure of antiferromagnetic MnO (d^5 compound) with two formula units and two nonequivalent Mn atoms in the primitive cell; see also Section 6.5.2. As for NiO, *spin polarized* calculations have been performed. In Figure 3.33a, we see the DOS and partial DOS (also for spin up and down for Ni) and see that – as we have less d electrons than in NiO – for a given Mn atom, for example, Mn₁, one spin channel is completely occupied (spin up) and the other completely empty (spin down). With the antiferromagnetic order, the situation is inverse for Mn₂. This is in accordance to *Hund's rule*, where electrons tend to align their spin when they are placed in the same orbitals, and as five electrons fill the d band to 50%, we obtain a maximal net magnetic moment

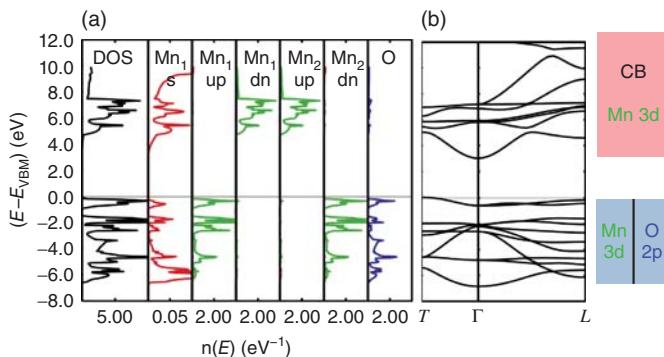


Figure 3.33 Electronic structure of MnO (rock-salt structure with antiferromagnetic ordering of type II) calculated with the HSE06 functional. Here we perform a spin polarized calculation, that is, each band can be occupied by just one electron. (a) Total DOS and partial DOS of Mn (spin up and down) and oxygen, which are subdivided

into the s (red), p (blue), and d (green) contributions. The calculated band gap is 3 eV and the magnetic moment on the Ni is $4.5 \mu_B$. (b) Corresponding band structure. The VBM is the zero of the energy scale. (The experimental lattice constant is used [86].)

of $4.5 \mu_B$ in our calculation. This is in very good agreement with the experimental value of $4.58 \mu_B$ [84]. The top of the valence band is of e_g character (as also expected from the crystal field splitting) [37] with some admixture of oxygen 2p states. The bottom of the conduction band is of Mn 4s character; see the partial DOS and the parabolic conduction band around Γ in Figure 3.33b. We calculate an *indirect* band gap from T to Γ of 3.0 eV, which is smaller than the experimentally determined value of about 3.9–4.1 eV [81, 87]. The band structure agrees well with other theoretical calculations, which were performed with different functionals [37, 82].

3.5.4

M^{VB} Metal Oxides

3.5.4.1 Metal–Insulator Transitions: NbO_2 , VO_2 , and V_2O_3

We have described the gradual metal–insulator transition in Ti_2O_3 with increasing temperature, in Section 3.5.2.2. There are other famous examples for MITs in the field of transition metal oxides, the most prominent being VO_2 and V_2O_3 (which are also discussed in Sections 6.6.2 and 6.6.4), but also NbO_2 . The bandwidth for higher orbital momenta (d and f) follows the trend $4f < 5f < 3d < 4d \sim 5d$, that is, 4f electrons are the most localized ones and 4d and 5d are the least localized [3]. Correspondingly, correlation effects for the 4d electrons of Nb do not play as strong a role as for the 3d electrons in VO_2 and V_2O_3 . Therefore, NbO_2 can be described by band theory and the occurring MIT is explained by an *embedded Peierls instability* [31], see Section 3.4.3 and the following discussion. VO_2 [88]

and V_2O_3 [89, 90] have been calculated with LDA and DMFT (dynamical mean-field theory) in order to treat the correlations more explicitly. Calculations of VO_2 with the hybrid functional HSE06 fail to predict the correct ground state [91].

VO_2 (d^1 compound) is a challenging system, as correlations play an important role in combination with a structural transition, so it has been described as a *correlation-assisted Peierls transition* [88] or an *orbital-assisted collaborative Mott–Peierls transition* [92]. There is still an ongoing debate on the dominant cause for the MIT, but apparently, it is established that both correlations and structural rearrangements are needed for the transition to take place [93]. The high-temperature paramagnetic metallic rutile phase transforms into a low-temperature insulating monoclinic phase (denoted as M_1) at $T_c(VO_2)=340\text{ K}$ [94]. The M_1 phase shows dimerization of the metal atoms along the tetragonal c -axis (characteristic of a Peierls transition) and a zigzag-like shift of the vanadium atoms perpendicular to this axis [95]. To complicate matters, there are two additional phases, a triclinic phase (T) and a monoclinic phase (M_2), which appear under uniaxial stress or by doping with, for example, Cr [95, 96].

V_2O_3 (d^2 compound, corundum structure), in particular Cr-doped V_2O_3 , is considered as a prototype of a correlation-driven *Mott transition*, see Section 3.3.2.5. Particularly, $(V_{0.989}\text{Cr}_{0.011})_2\text{O}$ was reported to have a temperature-induced *isostructural* phase transition from a paramagnetic insulator (PI) to a paramagnetic metal (PM) [97], in addition to transition from an antiferromagnetic insulator (AFI) to a PM at lower temperatures. But transitions can also be observed by doping or application of pressure [3], which were shown to be nonequivalent (for a review, see [90]).

As for VO_2 , the MIT in NbO_2 (also d^1 compound) is accompanied by a structural transition, that is, the metallic rutile phase at high temperatures undergoes a transition ($T_c(NbO_2)=1081\text{ K}$ [98, 99]) to an insulating monoclinic phase. Here, we also find dimerization of the Nb atoms and zigzag-like displacements of the metal atoms in a body-centered tetragonal lattice, which can be described with 32 formula units per cell [100]. Figure 3.34 shows the DOS of the low-temperature and high-temperature phase of NbO_2 . In both phases, the valence band is composed of mainly oxygen 2p-like states, as for all the oxides discussed so far. In Figure 3.34b, the Fermi energy lies within the Nb 4d band, resulting in metallic behavior in the rutile phase. As shown in Figure 3.34a, the DOS is noisier due to the lower symmetry of the compound and we observe a band gap of 0.8 eV, which is in good agreement with the experimental band gap of 0.5 eV at room temperature [103]. The opening of the gap can be explained by metal–metal bonding (dimerization along the c -axis) and splitting of the Nb 4d states, which are parallel to the c -axis in bonding and antibonding states [31].

3.5.4.2 Tantalum Oxides TaO_x

The atomic structure of tantalum oxides is discussed controversially in the literature. Amorphous, metastable crystal structures and defect structures are reported, and the discussion about the structure of tantalum oxide seems to be still in progress. The technologically interesting low-temperature Ta_2O_5 phase

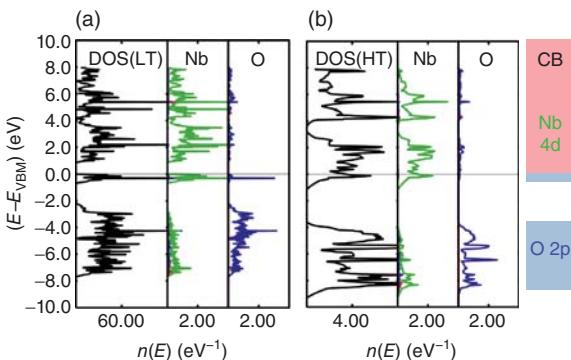


Figure 3.34 Electronic structure of NbO_2 calculated with the GGA + U functional [22] with $U - J = 3$. (a) Monoclinic NbO_2 (low-temperature phase). Total DOS and partial DOS of Nb and O, which are subdivided into the s (red), p (blue), and d (green) contributions. We have an insulator with a band gap of 0.8 eV. The VBM is the zero of the energy scale. The lattice constants were determined

experimentally [101]. (b) Rutile NbO_2 (high-temperature phase). Total DOS and partial DOS of Nb and O, which are subdivided into the s (red), p (blue), and d (green) contributions. We observe metallic behavior, as the Fermi energy lies within the Nb 4d states. The Fermi energy is the zero of the energy scale. (The experimental lattice constant is used [102].)

crystallizes in an orthorhombic structure. Early investigations on anodic oxide films of tantalum show amorphous tantalum oxide, which can be crystallized by heating to form a disordered structure, $\beta\text{-Ta}_2\text{O}_5$ [104]. Later, a superstructure of Ta_2O_5 ($\text{L-Ta}_2\text{O}_5$) was reported, an orthorhombic defect structure with 22 Ta and 58 O in the ideal unit cell [105], where three O sites are nonoccupied. Further, a new $\beta\text{-Ta}_2\text{O}_5$ structure with two formula units per unit cell [106] and a hexagonal structure, $\delta\text{-Ta}_2\text{O}_5$ [107], were proposed. Other structures were suggested from theoretical studies of the electronic properties of tantalum oxide. One example is a simplified β form with one formula unit per unit cell [108], called α model, and from MD simulations, a structure is found [109] with two formula units per unit cell (space group $Pbam$), called λ model. The MD simulations show that Ta_2O_5 in this structure is lower in energy than other structures mentioned, and it is proposed that this structure is hidden in the structural disorder found experimentally [109]. We also want to mention that, by oxidizing tantalum metal, an fcc (NaCl) structure of TaO is found [110], and a rutile structure is found by reducing Ta_2O_5 to TaO_2 by shock-loading processes [111].

The electronic states of tantalum(V) oxides are investigated extensively using the different suggested structures and using different exchange–correlation functionals. Details are given in recent papers [108–110]. As an example, we show in Figure 3.35, the band structure and DOS of $\lambda\text{-Ta}_2\text{O}_5$. The ordering of the valence states is $2p(\text{O}) < 5d(\text{Ta}), 6s(\text{Ta})$, and an optical band gap E_g between the occupied 2p-like states of O and the unoccupied 5d-like states is found. Our HSE06 value of E_g of 4 eV agrees with the experimental value [112] and the calculated value using the so-called screened exchange hybrid functional [109, 113].

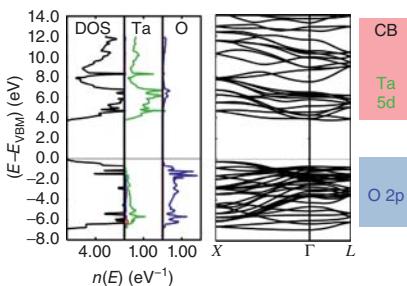


Figure 3.35 Electronic structure of Ta_2O_5 (λ phase, orthorhombic structure, space group $Pbam$ [109]) calculated with the HSE06 functional. (a) Total DOS on the left and Ta and O partial DOS on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. The lowest displayed states are the O 2p-like states and the unoc-

cupied Ta 5d-like states. The top of the valence band has mainly oxygen 2p character. (b) Corresponding band structure. The direct band gap is 4.0 eV. The VBM is the zero of the energy scale. (The lattice constants were determined by theoretical investigations [109].)

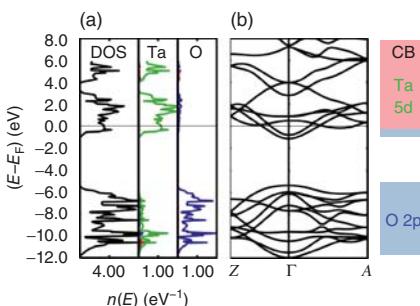


Figure 3.36 Electronic structure of Ta_2O_4 (rutile structure) calculated with the HSE06 functional. (a) Total DOS on the left and Ta and O partial DOS on the right, which are subdivided into the s (red), p (blue), and d (green) contributions. The lowest displayed states are the O 2p-like states and the partly occupied Ta 5d-like states. (b) Corresponding band structure. The direct band gap is 4.0 eV. The Fermi energy is the zero of the energy scale. (The lattice constants were determined experimentally [111].)

states are the O 2p-like states and the partly occupied Ta 5d-like states. (b) Corresponding band structure. The direct band gap is 4.0 eV. The Fermi energy is the zero of the energy scale. (The lattice constants were determined experimentally [111].)

Most interesting is the change in the electronic structure for structures with an O deficit. An example is $\text{Ta}_2\text{O}_{5-x}$, which has been studied extensively by experimental and theoretical methods [110, 113–116]. By assuming first the ordered rutile structure for $x=1$, one obtains the band structure as shown in Figure 3.36. The ordering of the states is the same as in Ta_2O_5 , but the Fermi energy is shifted to the now partly occupied 5d-like states, which are unoccupied in Ta_2O_5 .

For nonstoichiometric phases, the $\text{Ta}_2\text{O}_{5-x}$ structure and the electronic structure of the defects are complicated [110, 114], and the energetic position of the defect states E_{def} depends on the structure, the local relaxation of the atomic arrangement around the defects, and the charge of the defects. E_{def} ranges from

the conduction bands, as shown in Figure 3.36, over shallow states of about 0.5 eV below the CBM to deep states in the middle of the band gap, as in the states described in Section 3.4.4. Isolated defects in the conduction bands or directly below the CBM are localized states, which become more and more itinerant-like states with increasing number of defects [110].

3.6

Ellingham Diagram for Binary Oxides

We want to recall that a lowering in the oxidation state as considered at the end of the previous section takes place for thermodynamic reasons by an increase in temperature T . This can be seen from the Ellingham diagram shown in Figure 3.37. For binary oxides, the standard free reaction enthalpy $\Delta G^0(T)$ for the reaction $yM + \frac{1}{2}O_2 \leftrightarrow M_yO$ is always negative and increases in magnitude with increasing T .

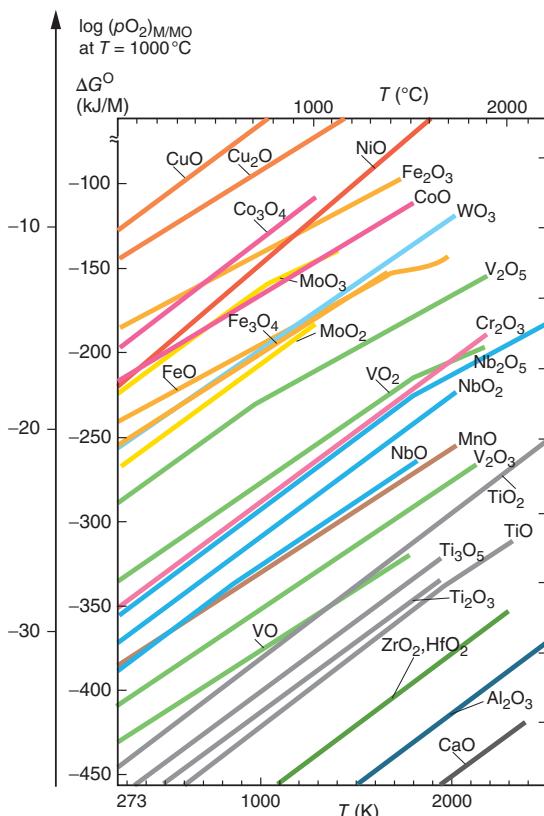


Figure 3.37 Temperature dependence of the standard free reaction enthalpy $\Delta G^0(T)$ for the reaction $yM + \frac{1}{2}O_2 \leftrightarrow M_yO$ for various transition metal oxides M_yO . The

energy axis is also shown in equivalent (formal) equilibrium oxygen partial pressures. The color code shows phases of identical cations, from Ref. [117].

The temperature- caused lowering of valence should also be accompanied by structural and electronic changes, which can change the resistivity behavior, and this effect might play a role in the mechanism of redox-based resistivity switching [117].

Acknowledgments

We want to acknowledge the Jülich Supercomputing Center (JSC) for providing the computation time and the High-Performance Computing Cluster at RWTH Aachen University. We thank Volker Eyert at Materials Design® and Michael Stephan (Jülich) for important discussions and support, and we thank Renate Eibler (Wien) and Barbara Albert (Darmstadt) for their helpful comments.

References

1. Kittel, C. (2004) *Introduction to Solid State Physics*, John Wiley & Sons, Inc..
2. Ashcroft, N.W. and Mermin, N.D. (1976) *Solid State Physics*, Saunders College.
3. Cox, P.A. (1992) *Transition Metal Oxides: An Introduction to Their Electronic Structure and Properties*, Oxford University Press.
4. de Groot, F. and Kotani, A. (2008) *Core Level Spectroscopy of Solids*, CRC Press.
5. O'Keeffe, M. (1977) *Acta Crystallogr., Sect. A*, **33**, 924.
6. Rohrer, G.S. (2001) *Structure and Bonding in Crystalline Materials*, Cambridge University Press.
7. Magnéli, A. (1948) *Acta Chem. Scand.*, **2**, 501.
8. Schwingenschlögl, U. and Eyert, V. (2004) *Ann. Phys. (Leipzig)*, **9**, 475.
9. Andersson, S., Collen, B., Kuylenstierna, U., and Magnéli, A. (1957) *Acta Chem. Scand.*, **11**, 1641.
10. Andersson, G. (1954) *Acta Chem. Scand.*, **8**, 1599.
11. Canadell, E., Doublet, M.-L., and Iung, C. (2012) *Orbital Approach to the Electronic Structure of Solids*, Oxford University Press.
12. Bredow, T. and Gerson, A.R. (2000) *Phys. Rev. B*, **61**, 5194.
13. Roessler, D.M. and Walker, W.C. (1967) *Phys. Rev.*, **159**, 733.
14. Whited, R.C., Flaten, C.J., and Walker, W.C. (1973) *Solid State Commun.*, **13**, 1903.
15. Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003) *J. Chem. Phys.*, **118**, 8207.
16. Zuo, J.M., O'Keeffe, M., Rez, P., and Spence, J.C.H. (1997) *Phys. Rev. Lett.*, **78**, 4777.
17. Engel, E. and Dreizler, R.M. (2011) *Density Functional Theory: An Advanced Course*, Springer.
18. Gebhard, F. (1997) *The Mott Metal-Insulator Transition: Models and Methods*, Springer.
19. Koch, E. (2012) in *Reihe Modeling and Simulation*, vol. 2 (eds E. Pavarini, E. Koch, F. Anders, and M. Jarrell), Forschungszentrum Jülich, Chapter 7.
20. Perdew, J.P., Ernzerhof, M., and Burke, K. (1996) *J. Chem. Phys.*, **105**, 9982.
21. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) *Phys. Rev. Lett.*, **77**, 3865.
22. Dudarev, S.L., Botton, G.A., Savrasov, S.Y., Humphreys, C.J., and Sutton, A.P. (1998) *Phys. Rev. B*, **57**, 1505.
23. Kresse, G. and Furthmüller, J. (1996) *Phys. Rev. B*, **54**, 11169.
24. Kresse, G. and Joubert, D. (1999) *Phys. Rev. B*, **59**, 1758.
25. Blöchl, P.E. (1994) *Phys. Rev. B*, **50**, 17953.
26. Berliner, R., Fajen, O., Smith, H.G., and Hitterman, R.L. (1989) *Phys. Rev. B*, **40**, 12086.

27. Olijnyk, H. and Holzapfel, W.B. (1985) *Phys. Rev. B*, **31**, 4682.
28. Hubbard, C.R., Swanson, H.E., and Mauer, F.A. (1975) *J. Appl. Crystallogr.*, **8**, 45.
29. Bludau, W., Onton, A., and Heinke, W. (1974) *J. Appl. Phys.*, **45**, 1846.
30. Zucca, R.R.L. and Shen, Y.R. (1970) *Phys. Rev. B*, **1**, 2668.
31. Evert, V. (2002) *Europhys. Lett.*, **58**, 851.
32. Anderson, P.W. (1958) *Phys. Rev.*, **109**, 1492.
33. Mott, N.F. and Davis, E.A. (2012) *Electronic Processes in Non-Crystalline Materials*, Oxford University Press.
34. Terakura, K., Williams, A.R., Oguchi, T., and Kübler, J. (1984) *Phys. Rev. Lett.*, **52**, 1830.
35. Anisimov, V.I., Zaanen, J., and Andersen, O.K. (1991) *Phys. Rev. B*, **44**, 943.
36. Anisimov, V.I., Solovyev, I.V., Korotin, M.A., Czyzyk, M.T., and Sawatzky, G.A. (1993) *Phys. Rev. B*, **48**, 16929.
37. Rödl, C., Fuchs, F., Furthmüller, J., and Bechstedt, F. (2009) *Phys. Rev. B*, **79**, 235114.
38. Kröger, F.A. and Vink, H.J. (1956) *Solid State Phys.*, **3**, 307.
39. De Vita, A., Gillan, M.J., Lin, J.S., Payne, M.C., Stich, I., and Clarke, L.J. (1992) *Phys. Rev. B*, **46**, 12964.
40. Alfè, D. and Gillan, M.J. (2005) *Phys. Rev. B*, **71**, 220101 (R).
41. Schönberger, U. and Aryasetiawan, F. (1995) *Phys. Rev. B*, **52**, 8788.
42. Sushko, P.V., Shluger, A.L., and Catlow, C.R.A. (2000) *Surf. Sci.*, **450**, 153.
43. Henrich, V.E. and Cox, P.A. (1994) *The Surface Science of Metal Oxides*, Cambridge University Press.
44. Dondi, M., Matteuci, F., Baldi, G., Barzanti, A., Cruciani, G., Zama, I., and Bianchi, C.L. (2008) *Dyes Pigm.*, **76**, 179.
45. Houston, J.E. (1975) *J. Vac. Sci. Technol.*, **12**, 255.
46. French, R.H. (1990) *J. Am. Ceram. Soc.*, **73**, 477.
47. Ghebouli, M.A., Ghebouli, B., Bouhemadou, A., Fatmi, M., and Bouamama, K. (2011) *J. Alloys Compd.*, **509**, 1440.
48. Kaneko, Y. and Koda, T. (1988) *J. Cryst. Growth*, **86**, 72.
49. Cappellini, G., Bouette-Russo, S., Amadon, B., Noguera, C., and Finocchi, F. (2000) *J. Phys. Condens. Matter*, **12**, 3671.
50. Karzel, H., Potzel, W., Köfferlein, M., Schiessl, W., Steiner, M., Hiller, U., Kalvius, G.M., Mitchell, D.W., Das, T.P., Blaha, P., Schwarz, K., and Pasternak, M.P. (1996) *Phys. Rev. B*, **53**, 11425.
51. Reynolds, D.C., Look, D.C., Jogai, B., Litton, C.W., Cantwell, G., and Harsch, W.C. (1999) *Phys. Rev. B*, **60**, 2340.
52. Wróbel, J., Kurzydłowski, K.J., Hummer, K., Kresse, G., and Piechota, J. (2009) *Phys. Rev. B*, **80**, 155124.
53. Oba, F., Togo, A., Tanaka, I., Paier, J., and Kresse, G. (2008) *Phys. Rev. B*, **77**, 245202.
54. Bartkowski, S., Neumann, M., Kurmaev, E.Z., Fedorenko, V.V., Shamin, S.N., Cherkashenko, V.M., Nemonov, S.N., Winiarski, A., and Rubie, D.C. (1997) *Phys. Rev. B*, **56**, 10656.
55. Ern, V. and Switendick, A.C. (1965) *Phys. Rev.*, **137**, A1927.
56. Neckel, A., Rastl, P., Eibler, R., Weinberger, P., and Schwarz, K. (1976) *J. Phys. C: Solid State Phys.*, **9**, 579.
57. Taylor, D. (1984) *Br. Ceram. Trans. J.*, **83**, 92.
58. Honig, J.M. and Reed, T.B. (1968) *Phys. Rev.*, **174**, 1020.
59. Shin, S.H., Chandrashekhar, G.V., Loehman, R.E., Honig, J.M., and Sladek, R.J. (1973) *Phys. Rev. B*, **8**, 1364.
60. Guo, Y., Clark, S.J., and Robertson, J. (2012) *J. Phys. Condens. Matter*, **24**, 325504.
61. Rice, C.E. and Robinson, W.R. (1977) *Acta Crystallogr. Sect. B*, **33**, 1342.
62. Zandt, L.L.V., Honig, J.M., and Goodenough, J.B. (1968) *J. Appl. Phys.*, **39**, 594.
63. Tobaldi, D.M., Tucci, A., Skapin, A.S., and Esposito, L. (2010) *J. Eur. Ceram. Soc.*, **30**, 2481.
64. Grunes, L.A., Leapman, R.D., Wilker, C.N., Hoffmann, R., and Kunz, A.B. (1982) *Phys. Rev. B*, **25**, 7157.
65. Janotti, A., Varley, J.B., Rinke, P., Umezawa, N., Kresse, G., and

- Van de Walle, C. (2010) *Phys. Rev. B*, **81**, 085212.
66. Landmann, M., Rauls, E., and Schmidt, W.G. (2012) *J. Phys. Condens. Matter*, **24**, 195503.
67. Amtout, A. and Leonelli, R. (1995) *Phys. Rev. B*, **51**, 6842.
68. Göpel, W., Anderson, J.A., Frankel, D., Jaehnig, M., Phillips, K., Schäfer, J.A., and Rocker, G. (1984) *Surf. Sci.*, **139**, 333.
69. Hüfner, S. and Wertheim, G.K. (1973) *Phys. Rev. B*, **8**, 4857.
70. Fischer, D.W. (1972) *Phys. Rev. B*, **5**, 4219.
71. Paxton, A.T. and Thiên-Nga, L. (1998) *Phys. Rev. B*, **57**, 1579.
72. Glassford, K.M. and Chelikowsky, J.R. (1992) *Phys. Rev. B*, **46**, 1284.
73. Schmidbauer, M., Kwasniewski, A., and Schwarzkopf, J. (2012) *Acta Crystallogr. Sect. B*, **68**, 8.
74. Kowalczyk, S.P., McFeely, F.R., Ley, L., Gritsyna, V.T., and Shirley, D.A. (1977) *Solid State Commun.*, **23**, 161.
75. Brookes, N.B., Law, D.S.-L., Padmore, T.S., Waeburton, D.R., and Thornton, G. (1988) *Solid State Commun.*, **57**, 473.
76. Piskunov, S., Heifets, E., Eglitis, R.I., and Borstel, G. (2004) *Comput. Mater. Sci.*, **29**, 165.
77. Tezuka, Y., Shin, S., Ishii, T., Ejima, T., Suzuki, S., and Sato, S. (1994) *J. Phys. Soc. Jpn.*, **63**, 347.
78. van Benthem, K., Elsässer, C., and French, R.H. (2001) *J. Appl. Phys.*, **90**, 6156.
79. Smith, N. (1936) *J. Am. Chem. Soc.*, **58**, 173.
80. Kunes, J., Anisimov, V.I., Skornyakov, S.L., Lukyanov, A.V., and Vollhardt, D. (2007) *Phys. Rev. Lett.*, **99**, 156404.
81. Kurmaev, E.Z., Wilks, R.G., Moewes, A., Finkelstein, L.D., Shamin, S.N., and Kunes, J. (2008) *Phys. Rev. B*, **77**, 165127.
82. Gillen, R. and Robertson, J. (2013) *J. Phys. Condens. Matter*, **25**, 165502.
83. Alperin, H.A. (1962) *J. Phys. Soc. Jpn. Suppl. B*, **17**, 12.
84. Cheetham, A.K. and Hope, D.A.O. (1983) *Phys. Rev. B*, **27**, 6964.
85. Fender, B.E.F., Jacobson, A.J., and Wedgwood, F.A. (1968) *J. Chem. Phys.*, **48**, 990.
86. Trukhanov, S.V., Troyanchuk, I.O., Bobrikov, I.A., Simkin, V.G., and Balagurov, A.M. (2007) *Poverkhnostnye Fizika, Khimiya, Mekhanika*, **12**, 11.
87. van Elp, J., Potze, R.H., Eskes, H., Berger, R., and Sawatzky, G.A. (1991) *Phys. Rev. B*, **44**, 1530.
88. Biermann, S., Poteryaev, A., Lichtenstein, A.I., and Georges, A. (2005) *Phys. Rev. Lett.*, **94**, 026404.
89. Held, K., Keller, G., Eyert, V., Vollhardt, D., and Anisimov, V.I. (2001) *Phys. Rev. Lett.*, **86**, 5345.
90. Hansmann, P., Toschi, A., Sangiovanni, G., Saha-Dasgupta, T., Lupi, S., Marsi, M., and Held, K. (2013) *Phys. Status Solidi B*, **250**, 1251.
91. Grau-Crespo, R., Wang, H., and Schwingenschlögl, U. (2012) *Phys. Rev. B*, **86**, 081101 (R).
92. Haverkort, M.W., Hu, Z., Tanaka, A., Reichelt, W., Streletsov, S.V., Korotin, M.A., Anisimov, V.I., Hsieh, H.H., Lin, H.-J., Chen, C.T., Khomskii, D.I., and Tjeng, L.H. (2005) *Phys. Rev. Lett.*, **95**, 196404.
93. Tomczak, J.M. and Biermann, S. (2007) *J. Phys. Condens. Matter*, **19**, 365206.
94. Morin, F.J. (1959) *Phys. Rev. Lett.*, **3**, 34.
95. Eyert, V. (2011) *Phys. Rev. Lett.*, **107**, 016401.
96. Eyert, V. (2002) *Ann. Phys. (Leipzig)*, **11**, 9.
97. Mc Whan, D.B., Rice, T.M., and Remeika, J.P. (1969) *Phys. Rev. Lett.*, **23**, 1384.
98. Jannink, R.F. and Whitmore, D.H. (1966) *J. Phys. Chem. Solids*, **27**, 1183.
99. Sakata, T., Sakata, K., and Nishida, I. (1967) *Phys. Status Solidi*, **20**, K155.
100. Eyert, V., Horny, R., Höck, K.-H., and Horn, S. (2000) *J. Phys. Condens. Matter*, **12**, 4923.
101. Arakcheeva, A.V., Grinevich, V.V., Shamrai, V.F., Meyer, M., and Chapuis, G. (1999) *Crystallogr. Rep.*, **44**, 2.
102. Goldschmidt, V.M., Barth, T., Holmsen, D., Lunde, G., and Zachariasen, W.H. (2002) *Golden Book Phase Transitions*, 1, 1.

103. Adler, D. (1968) *Rev. Mod. Phys.*, **40**, 714.
104. Calvert, L.D. and Draper, P.H.G. (1962) *Can. J. Chem.*, **40**, 1943.
105. Stephenson, N.E. and Roth, R.S. (1971) *Acta Crystallogr. B*, **27**, 1037.
106. Aleshina, L. and Loginova, S. (2002) *Crystallogr. Rep.*, **47**, 415.
107. Fujikawa, H. and Taga, Y. (1994) *J. Appl. Phys.*, **75**, 2538.
108. Nashed, R., Hassan, W.M.I., Ismail, Y., and Allam, N.K. (2013) *Phys. Chem. Chem. Phys.*, **15**, 1352.
109. Lee, S.-H., Kim, J., Kim, S.-J.K.S., and Park, G.-S. (2013) *Phys. Rev. Lett.*, **110**, 235502.
110. Yang, Y., Nahm, H.-H., Sugino, O., and Ohno, T. (2013) *AIP Adv.*, **3**, 042101.
111. Syono, Y., Kikuchi, M., Goto, T., and Fukuoka, K. (1983) *J. Solid State Chem.*, **50**, 133.
112. Chun, W.-J., Ishikawa, A., Fujisawa, H., Takata, T., Kondo, J.N., Hara, M., Kawai, M., Matsumoto, Y., and Domen, K. (2003) *J. Phys. Chem. B*, **107**, 1798.
113. Guo, Y. and Robertson, J. (2014) *Appl. Phys. Lett.*, **104**, 112906.
114. Rampasad, R. (2003) *J. Appl. Phys.*, **94**, 5609.
115. Park, G.-S., Kim, Y.B., Park, S.Y., Li, X.S., Heo, S., Lee, M.-J., Chang, M., Kwon, J.H., Kim, M., Chung, U.-I., Dittmann, R., Waser, R., and Kim, K. (2013) *Nat. Commun.*, **4**, 2382.
116. Prakash, A., Jana, D., and Maikap, S. (2013) *Nanoscale Res. Lett.*, **8**, 418.
117. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21**, 2632.

4

Defect Structure of Metal Oxides

Giuliano Gregori

4.1

Definition of Defects

Compared with the ideal structure of a perfect crystal (cf. Chapter 3), a real solid contains a variety of defects, which can be classified according to their dimensionality.

4.1.1

Zero-Dimensional Defects

These are point defects (Figure 4.1), namely cation or anion vacancies and interstitials, which are in thermodynamic equilibrium (cf. also Section 4.2 and Figure 4.3) and normally present in all solids at temperatures above 0 K. A vacancy is an unoccupied regular lattice site, while an interstitial is an ion displaced from its lattice position into an interstitial site. If impurities are present in the material, they can stay on the lattice sites of the host material or occupy interstitial positions. Point defects can be classified as intrinsic (native) and extrinsic (impurities/dopants) and are usually defined using the Kröger–Vink notation, which describes their type, electrical charge, and lattice position relative to the perfect lattice. For example, the intrinsic point defects in a solid of type MX (with M^+ being the cation and X^- the anion) are expressed as M_i^{\cdot} , $X_i^{\prime\prime}$, $V_M^{\prime\prime}$, $V_X^{\prime\prime}$, where the subscripts indicate the lattice site (i stands for the interstitial site), the superscripts indicate the charge of the defect with respect to the ideal lattice (the dot indicates a positive charge and the bar indicates a negative charge), and V denotes a vacancy. Extrinsic defects are typically aliovalent cations (or anions) that are accidentally present in the solid (impurities) or were deliberately added (dopants) to it.

4.1.2

One-Dimensional Defects

The one-dimensional or line defects are dislocations, which can be classified into edge, screw, mixed, and partial dislocations. Such line defects are strongly related

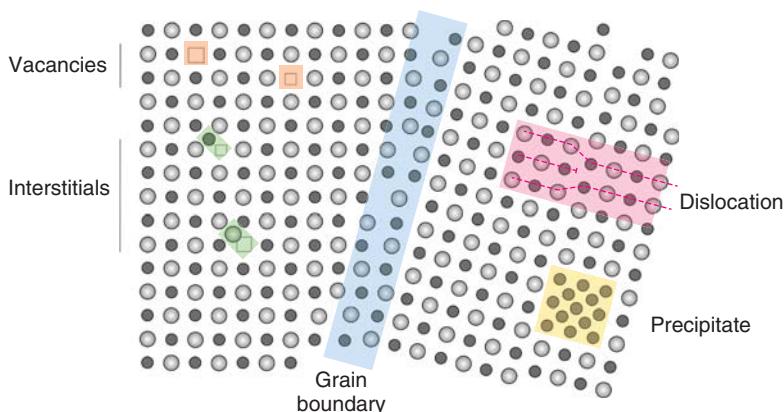


Figure 4.1 Schematic two-dimensional representation of some of the defects present in a solid of the type MX .

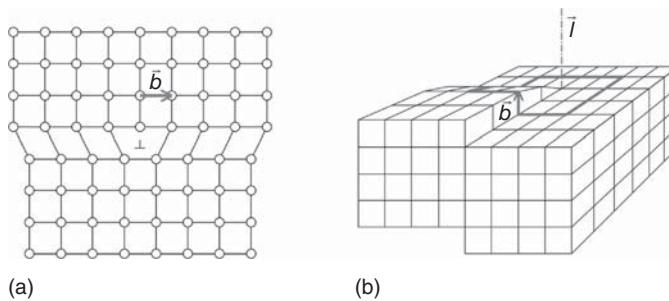


Figure 4.2 Simplified representation of (a) an edge dislocation with $\vec{\ell} \perp \vec{b}$ and (b) a screw dislocation with $\vec{\ell} \parallel \vec{b}$.

to the ability of deforming materials, and thus, they are extremely relevant for the mechanical properties, particularly of metals and, at high temperatures, of ceramics (e.g., creep). Figure 4.2 displays a schematic of edge and screw dislocations. The first can be thought of as consisting of an extra half crystallographic plane inserted (or extracted) into (from) the regular crystalline lattice. This extra plane does not extend throughout the crystal, and its bottom edge (defined as dislocation line) represents the edge dislocation while the volume surrounding it is the dislocation core. Despite the different geometry, analogous considerations apply also to screw dislocations. The quantities used to describe these line defects are the dislocation line vector $\vec{\ell}$ (usually depicted with the symbol \perp) and the Burgers vector \vec{b} , which is the vector required to complete a clockwise loop around the

dislocation counting an equal number of ions in each direction. In the case of the edge dislocation, the Burgers vector and the dislocation line identify the slip plane along which the dislocation can move.

Quite importantly, as dislocations break the local lattice symmetry (e.g., see Ref. [1–3]), they can result to be electrically charged and thus affect the local concentration of mobile point defects such as vacancies and interstitials (cf. Section 4.5.1.4).

4.1.3

Two-Dimensional Defects

These surface defects include interfaces between (i) a solid and the environment (surface), (ii) two solids (e.g., interfaces in multilayered thin films), (iii) two phases (e.g., same composition with different crystal structures), and (iv) regions having same composition and crystal structure but different orientations (i.e., grain boundaries – including tilt and twist boundaries – twin boundaries and stacking faults). For the scope of this chapter, grain boundaries are considered. Their formation – starting from two adjacent and equally oriented grains – can be envisaged as resulting from the rotation of one grain with respect to the other: if the axis of rotation is perpendicular to the boundary, then a twist boundary is obtained; if the axis lies in the plane of the boundary, then a tilt boundary results. The degree of rotation is important because it affects the coherency of the final grain boundary. Interestingly, low rotation angle tilt boundaries can be treated as an ensemble of edge dislocations, while low rotation angle twist boundaries can be described as an aggregation of screw dislocations.

As for dislocations, also for grain boundaries, a core can be identified: this is the region in which, in order to accommodate grains with different orientations, the lattice symmetry is broken and a deviation from the bulk cation-to-anion ratio may occur [4–8]. This obviously can result in the formation of a local electrostatic charge, which – as pointed out in the case of dislocations – can deeply modify the mobile defect concentrations (cf. Section 4.5.2.3) in the surrounding. Moreover, it is worth emphasizing that even if at the grain boundaries, the bulk cation-to-anion ratio is maintained, the local distortion of the lattice can result in a modification of the defect formation energy (cf. Section 4.2), which can lead to a preferential segregation of defects at the grain boundary.

4.1.4

Three-Dimensional Defects

Three-dimensional defects are volume defects which differ from the hosting solid in orientation, composition, and structure. These include pores, inclusions, precipitates, and second phases.

4.2

General Considerations on the Equilibrium Thermodynamics of Point Defects

In the following, a brief description of the main aspects of the equilibrium thermodynamics of point defects is presented for the case of small concentrations of defects (diluted solutions) [9, 10]. Starting from the ideal case of a perfect crystal (with no defects), the formation of defects in a real solid corresponds to a change in its total Gibbs free energy, which reads

$$\Delta_d G = (\Delta_d H - T\Delta_d S_{\text{vib}}) - T\Delta_d S_{\text{conf}} = N_d(\Delta_d \epsilon - T\Delta_d s_{\text{vib}}) - T\Delta_d S_{\text{conf}}, \quad (4.1)$$

where N_d is the number of defects and the subscripts *vib* and *conf* refer to the vibrational and configurational contributions, respectively. $\Delta_d \epsilon$ is the enthalpy of formation of a single defect, while Δs_{vib} refers to the entropy change associated with the modified vibrations in the surrounding of a single defect. The configurational term is expressed as

$$T\Delta_d S_{\text{conf}} = T k_B \ln \Omega = k_B T \ln \left(\frac{(N)!}{N_d!(N-N_d)!} \right), \quad (4.2)$$

from which, by using Stirling's approximation and assuming the number of defects to be much smaller than the total number of available lattice sites ($N_d \ll N$),

$$\begin{aligned} T\Delta_d S_{\text{conf}} &= k_B T \ln \Omega = k_B T N \ln \left(\frac{N}{N - N_d} \right) - k_B T N_d \ln \left(\frac{N_d}{N - N_d} \right) \\ &= -k_B T N_d \ln \left(\frac{N_d}{N} \right) \end{aligned} \quad (4.3)$$

results. Thus, for a real solid, the total free energy reads

$$G_{\text{real}} = G_{\text{perfect}} + \Delta_d G = G_{\text{perfect}} + N_d(\Delta_d \epsilon - T\Delta_d s_{\text{vib}}) + N_d k_B T \ln \left(\frac{N_d}{N} \right). \quad (4.4)$$

From this, the chemical potential of a defect ($\mu_d = \partial G / \partial n_d$, with $n_d = N_d/N_A$, N_A being Avogadro's number) can be determined:

$$\mu_d = \frac{\partial G_{\text{real}}}{\partial n_d} = \frac{N_A \partial G_{\text{real}}}{\partial N_d} = N_A(\Delta_d \epsilon - T\Delta_d s_{\text{vib}}) + RT \ln x_d = \mu_d^* + RT \ln x_d \quad (4.5)$$

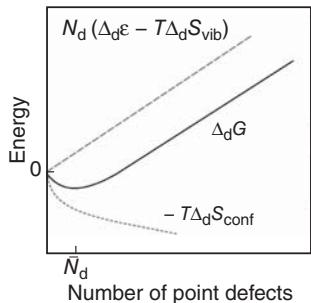


Figure 4.3 The change of free energy $\Delta_d G$ expressed as a function of the number of point defects exhibits a minimum, which explains why defects are always present in a solid at nonzero temperatures.

(note that $x_d = N_d/N = n_d/n$ and $\mu_d^* = N_A(\Delta_d \varepsilon - T\Delta s_{\text{vib}})$). As at equilibrium, $\partial G_{\text{real}}/\partial N_d = 0$,

$$x_d = \frac{N_d}{N} = e^{-\mu_d^*/RT} = e^{-N_A(\Delta_d \varepsilon - T\Delta s_{\text{vib}})/RT} = e^{-(\Delta_d \varepsilon - T\Delta s_{\text{vib}})/k_B T} \quad (4.6)$$

results. Consequently, mass action laws can be used to determine the defect concentrations for given defect reactions. Finally, it is worth noting that the free energy change $\Delta_d G$ expressed as a function of the number of defects N_d exhibits a minimum (see Figure 4.3) owing to the configurational entropic term in Eq. (4.1). This explains why at the thermodynamic equilibrium defects are always present in solids at nonzero temperatures.

4.3

Definition of Point Defects

4.3.1

Intrinsic Defects

For the definition of intrinsic defects and their formation equilibrium constants, the prototypical case of an oxide with the general formula MO_2 is considered.

4.3.1.1 Frenkel Defects

The Frenkel disorder is characterized by the displacement of cations from their regular lattice sites into interstitial positions. This results in the formation of a stoichiometric amount of interstitial ions and vacancies in the cation sublattice, which is described by the following reaction:



with

$$K_F = [\text{M}_i^{\cdots\cdots}][V_M^{////}] = e^{(T\Delta_F S^\circ - \Delta_F H^\circ)/RT}, \quad (4.8)$$

where $\Delta_F S^\circ$ and $\Delta_F H^\circ$ are the vibrational entropic term and the enthalpy of formation, respectively. Note that in Eq. (4.7), the activity of the regular ions M_M^X is considered to be constant. This means that – at a given temperature – the concentration of native cation interstitials and vacancies is given by:

$$[\text{M}_i^{\cdots\cdots}] = [V_M^{////}] = \sqrt{K_F} = e^{\Delta_F S^\circ / 2R} e^{-\Delta_F H^\circ / 2RT}. \quad (4.9)$$

4.3.1.2 Anti-Frenkel Defects

The anti-Frenkel disorder involves the anion sublattice: here, oxygen ions are displaced from their regular lattice sites into interstitial positions according to

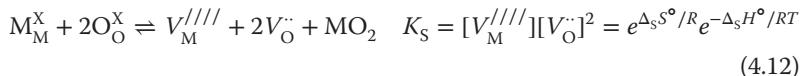


Note that, similarly to the case of Frenkel disorder, the following expression can be written:

$$[O_i^{++}] = [V_O^{\cdot\cdot}] = \sqrt{K_F} = e^{\Delta_F S^\circ / 2R} e^{-\Delta_F H^\circ / 2RT}. \quad (4.11)$$

4.3.1.3 Schottky Defects

The Schottky disorder corresponds to the formation of a stoichiometric amount of cation and anion vacancies (Schottky pair) in the solid, and it is defined as follows:



The process is accompanied by expansion of the samples' volume.

4.3.1.4 Anti-Schottky Defects

The anti-Schottky disorder is the counterpart of the Schottky reaction: in this case, oppositely charged interstitial defects are formed:



4.3.1.5 Electron Band–Band Transfer

In addition to the ionic disorder, the intrinsic electronic disorder reaction also needs to be taken into account, which results from the promotion of an electron from the valence band into the conduction band (with the corresponding formation of a hole in the valence band):



with

$$n = p = \sqrt{K_B} = e^{\Delta_B S^\circ / 2R} e^{-\Delta_B H^\circ / 2RT}. \quad (4.15)$$

4.3.2

Extrinsic Defects

4.3.2.1 Reactions with the Environment

Let us now consider the equilibrium reactions with the environment. It is worth noting that although, in principle, there is only one reaction describing the interaction of the oxide with the surrounding, it is convenient to write it depending on the majority defects present in the solid.

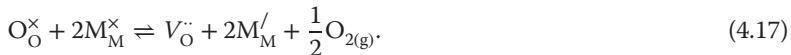
Oxygen Excorporation The oxygen excorporation reactions with participation of electrons and the corresponding law of mass action are expressed as



where

$$K_{O,n} = e^{\Delta_o S^\circ / R} e^{-\Delta_o H^\circ / RT}.$$

If the electrons are localized at the cations (e.g., as in cerium dioxide), then Eq. (4.16) can be rewritten as follows:



As the reaction entropic term associated with the formation of oxygen in the gas phase is positive, the concentration of oxygen vacancies becomes increasingly energetically favorable at high temperatures.

Oxygen Incorporation The insertion of oxygen into the MO_2 lattice results in the formation of oxygen interstitials and holes, which is expressed as follows:



The corresponding law of mass action reads

$$K_{\text{O},p} = [\text{O}_\text{i}^{\prime\prime\prime}] \cdot p^2 \cdot p\text{O}_2^{-1/2} = K_{\hat{F}} \cdot K_B^2 \cdot K_{\text{O},n}^{-1}.$$

However, if oxygen vacancies are present in the oxide (e.g., to compensate acceptor dopants), the oxygen incorporation reaction is



with

$$K_{\text{O},\hat{p}} = [V_\text{O}^{\cdot\cdot}]^{-1} \cdot p^2 \cdot p\text{O}_2^{-1/2} = K_{\text{O},p} \cdot K_{\hat{F}}^{-1}.$$

Metal Oxidation For those situations in which the majority defects are $V_\text{M}^{\prime\prime\prime\prime\prime}$, the interaction with the environment can be written as



Metal Excess Finally, metal excess in an oxide can occur not only from the oxygen deficiency but also from the formation of cation interstitials as described by



4.3.2.2 The Brouwer Diagram

From the aforementioned reactions and by taking into account also the formation reactions of the intrinsic defects (Eqs. (4.7)–(4.15)), the concentration of the native point defects can be determined as a function of oxygen partial pressure and temperature. For this purpose, the condition of charge neutrality needs to be fulfilled. For the MO_2 solid considered here, this reads:

$$4[\text{M}_i^{\cdot\cdot\cdot}] + 2[V_\text{O}^{\cdot\cdot}] + h^{\cdot} = 4[V_\text{M}^{\prime\prime\prime\prime\prime}] + 2[\text{O}_\text{i}^{\prime\prime\prime}] + e^{\prime\prime}. \quad (4.22)$$

Since the defect concentrations can differ from each other even by several orders of magnitude, only the majority defects are usually considered in Eq. (4.22). This results in simple power laws describing the dependence of the

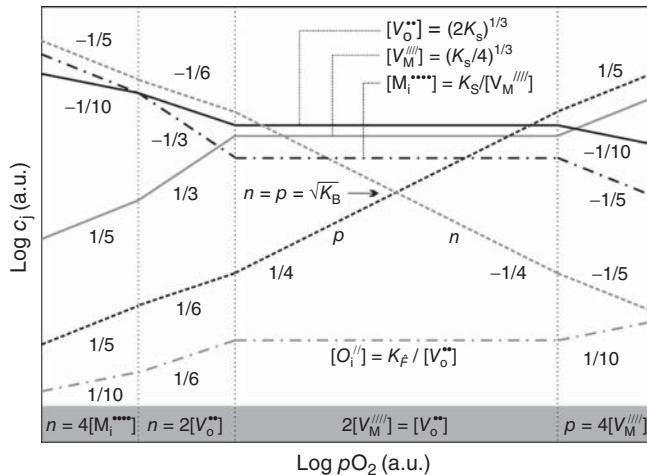


Figure 4.4 Brouwer diagram determined at an arbitrary temperature for an oxide of the type MO_2 in which $K_S \gg K_B \gg K_F$.

defect concentration on the oxygen partial pressure. Figure 4.4 illustrates the Brouwer diagram of a MO_2 oxide (for which $K_S \gg K_B \gg K_F$ holds) depicted for an arbitrary temperature. The gray field shows the electroneutrality conditions valid at various pO_2 . As an example of how the defect concentrations can be determined, let us now consider the intermediate pO_2 range, in which Eq. (4.22) becomes $[V_O^{\bullet}] = 2[V_M^{///}]$. From this and by knowing K_S (see reaction (4.12)), both the concentrations of oxygen vacancies and metal vacancies can be promptly calculated. The equilibrium constants of the Frenkel and anti-Frenkel reactions allow for the determination of the metal and the oxygen interstitials, respectively. As for the concentration of holes, reactions (4.12) and (4.19) must be taken into account, from which (together with $[V_O^{\bullet}] = 2[V_M^{///}]$)

$$p = K_{O,\hat{p}}^{1/2} \cdot (2K_S)^{1/6} \cdot pO_2^{1/4} \quad (4.23)$$

results. As for the pO_2 dependence of the electrons, the band–band transfer reaction needs to consider Eq. (4.14), yielding $n = (K_B/p) \propto pO_2^{-1/4}$.

4.3.2.3 Impurities and Dopants

The presence of aliovalent cations or anions in the form of impurities or dopants modifies the intrinsic electroneutrality condition and can result in a significant variation in the concentration of the native defects. The addition of hypovalent cations to an oxide results in the formation of negatively charged defects A_M^{\prime} (acceptors as, e.g., magnesium in alumina, Mg_{Al}^{\prime}), while the insertion of hypervalent cations results in the formation of positively charged defects D_M^{\prime} (donors as, e.g., lanthanum in strontium titanate, La_{Sr}^{\prime}). It is worth noting that, although less employed in real applications, the formation of acceptors or donors can be

achieved also through the addition of aliovalent anions, for example, N_O^+ (acting as acceptors) or F_O^- (acting as donors).

Furthermore, it is important to emphasize that impurities can be redox-active (as, e.g., in the case of Fe in SrTiO_3) [11], meaning that the effective concentration of acceptors (or donors) can vary with temperature and oxygen partial pressure. Therefore, in the case of acceptor doping, the following reaction needs to be considered (obviously an analogous reaction can be written also for donor doping):



$$K_\text{A} = \frac{[\text{A}_\text{M}^/]^2 \cdot [\text{V}_\text{O}^\cdot] \cdot p\text{O}_2^{1/2}}{[\text{A}_\text{M}^\times]^2} \quad (4.25)$$

with

$$[\text{A}_\text{M}^/] + [\text{A}_\text{M}^\times] = [\text{A}_\text{M}^{\text{Total}}]. \quad (4.26)$$

4.4

Space-Charge Effects

The defect chemistry considerations made in the preceding section are valid for the bulk situation (single crystal), where at equilibrium, the electroneutrality condition takes into account the electrostatic charge compensation between differently charged point defects. The situation is remarkably different in proximity of a charged interface (e.g., grain boundary), which induces the redistribution of the mobile charge carriers within the so-called space-charge layer (SCL). As shown by direct electron microscopy analyses [4–8] and predicted by theoretical models [9, 12] as well as computer simulations [13, 14], such excess charge arises from the deviation of the bulk cation-to-anion ratio (including segregation of impurities or dopants) due to microstructural conditions, which are related to the necessity of accommodating grains with different orientations. In such a situation, the total interface core charge is compensated by the total SCL charge:

$$\Sigma_{\text{core}} = -2\Sigma_{\text{SCL}}. \quad (4.27)$$

To quantitatively evaluate the charge carrier concentration variation within the SCL compared with the bulk, the electrochemical potential $\tilde{\mu}_j$ of a charge carrier j needs to be considered [15–17]. As in a one-dimensional case, at equilibrium, $\tilde{\mu}_{j,x} = \tilde{\mu}_{j,\infty}$ (with x indicating the position within the SCL and ∞ the bulk situation, far from the boundaries), one can write

$$\mu_{j,x}^0 + k_\text{B} T \ln c_{j,x} + z_j e \phi_x = \mu_{j,\infty}^0 + k_\text{B} T \ln c_{j,\infty} + z_j e \phi_\infty, \quad (4.28)$$

where $c_{j,x}$ and $c_{j,\infty}$ are the concentrations of the charge carrier j within the SCL and in the bulk, respectively. From this,

$$\frac{c_{j,\text{SCL}}}{c_{j,\infty}} = \exp\left(-\frac{z_j e (\phi_{\text{SCL}} - \phi_\infty)}{k_\text{B} T}\right) = \exp\left(-\frac{z_j e \Delta \phi}{k_\text{B} T}\right) \quad (4.29)$$

and in particular

$$\frac{c_{j,0}}{c_{j,\infty}} = \exp\left(-\frac{z_j e \Delta\phi_0}{k_B T}\right). \quad (4.30)$$

follows, where $c_{j,0}$ is the concentration of the charge carrier at $x = 0$ and $\Delta\phi_0 = \phi_0 - \phi_\infty$ (ϕ_0 being the electrostatic potential arising from the excess core charge at $x = 0$ and ϕ_∞ the electrostatic potential of the bulk). Note that if one takes $\phi_\infty = 0$, then the exponential term of Eq. (4.30) becomes $-z_j e \phi_0 / k_B T$. It is noteworthy here that the local charge density ρ resulting from the deviations of the charge carrier concentrations from the bulk values are linked to the electrostatic potential ϕ through the Poisson equation:

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{\rho(x)}{\epsilon_0 \epsilon_r}, \quad (4.31)$$

where ϵ_0 is the vacuum dielectric constant and ϵ_r the relative dielectric constant of the material. By combining Eqs. (4.29) and (4.31), the following Poisson–Boltzmann differential equation results

$$\frac{\partial^2 \phi(x)}{\partial x^2} = \frac{-e}{\epsilon_0 \epsilon_r} \sum_j z_j c_{j,\infty} \exp\left(\frac{-z_j e}{k_B T} \Delta\phi(x)\right). \quad (4.32)$$

Such expressions can be analytically solved in two cases, which are briefly summarized in the following sections.

4.4.1

Mott–Schottky Situation

This is the case of a material with an immobile, homogeneously distributed dopant (whose concentration is $c_{\text{dop},\infty}$ and its charge number z_{dop}) and a mobile majority charge carrier, which is depleted in the space-charge region. As in this case, $\rho \approx z_{\text{dop}} e c_{\text{dop},\infty}$, Eq. (4.32) can be written as

$$\frac{\partial^2 \phi(x)}{\partial x^2} = \frac{-z_j e c_{j,\infty}}{\epsilon_0 \epsilon_r} = \frac{-z_{\text{dop}} e c_{\text{dop},\infty}}{\epsilon_0 \epsilon_r}. \quad (4.33)$$

This is integrated by assuming the semi-infinite boundary conditions

$$\frac{\partial \phi(\lambda^*)}{\partial x} = 0 \text{ and } \phi(\lambda^*) = \phi_\infty, \quad (4.34)$$

yielding

$$\phi(x) - \phi_\infty = \Delta\phi(x) = \frac{-z_{\text{dop}} e c_{\text{dop},\infty}}{2\epsilon_0 \epsilon_r} (x - \lambda^*)^2, \quad (4.35)$$

where

$$\lambda^* = \sqrt{\frac{2\epsilon_0 \epsilon_r \Delta\phi_0}{z_{\text{dop}} e c_{\text{dop},\infty}}} \quad (4.36)$$

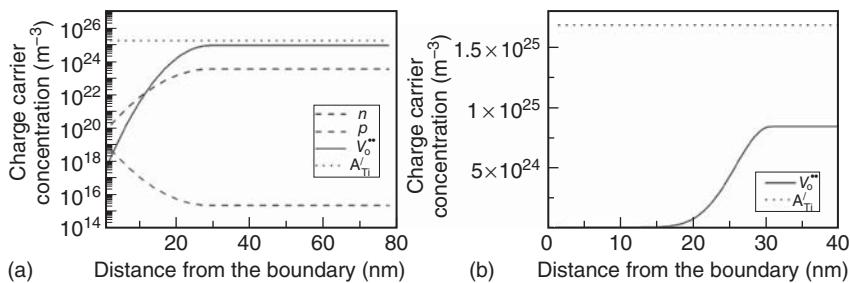


Figure 4.5 Charge concentration profiles calculated for SrTiO_3 as a function of the distance from a grain boundary ($x=0$) in the Mott–Schottky situation (acceptor concentration equal to 0.1 mol%): (a) log-scale and (b) linear-scale representation

(electrons and holes are here out of scale). Note that due to their higher charge, oxygen vacancies are stronger depleted than holes. The parameter used are $T=600\text{ }^\circ\text{C}$, $\epsilon_r=200$, $\Delta\phi_0=0.7\text{ V}$, $c_{\text{dop},\infty}=0.1\text{ mol\%}$.

is the extent of the space-charge region. The grain boundary core charge density is given by

$$\Sigma_{\text{core}} = \sqrt{8\epsilon_0\epsilon_r z_{\text{dop}} e c_{\text{dop},\infty} \Delta\phi_0} = 2\lambda^* z_{\text{dop}} e c_{\text{dop},\infty}. \quad (4.37)$$

Figure 4.5 illustrates the mobile charge carrier concentration profiles for the case of SrTiO_3 with an acceptor concentration of 0.1 mol% ($T=600\text{ }^\circ\text{C}$ and $\Delta\phi_0=0.7\text{ V}$).

4.4.2

Gouy–Chapman Situation

In the Gouy–Chapman (GC) situation, all charge carriers are mobile and can rearrange under the influence of the electrostatic field present in the space-charge region. It is noteworthy that Eq. (4.32) can be analytically solved only in a very specific case, namely the symmetrical case consisting of two intrinsic defects with opposite charge numbers and same bulk concentration (i.e., $z_1=-z_2$ and $c_{1,\infty}=c_{2,\infty}$). Then, solving Eq. (4.32) leads to

$$\phi(x) = \frac{2k_B T}{z_j e} \ln \left(\frac{1 + \Theta \exp(-x/\lambda)}{1 - \Theta \exp(-x/\lambda)} \right) \quad (4.38)$$

with

$$\lambda = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2z_j^2 e^2 c_{j,\infty}}} \quad (4.39)$$

being the Debye length and

$$\Theta = \tanh \left(\frac{z_j e \Delta\phi_0}{4k_B T} \right). \quad (4.40)$$

In this case, the grain boundary core charge density can be determined according to

$$\Sigma_{\text{core}} = \sqrt{8k_B T \epsilon_0 \epsilon_r c_{j,\infty}} \sinh \left(\frac{z_j e \Delta \phi_0}{2k_B T} \right). \quad (4.41)$$

A series of analytical solutions for various cases of SCL concentration profiles of charged defects can be found in the literature [18, 19].

Finally, it is worth summarizing the main properties of the space-charge zones: (i) the extent of the SCL is proportional to $(\epsilon_r / c_{j,\infty})^{1/2}$, meaning that large values of the dielectric constant and low impurity or dopant concentrations yield a large SCL; (ii) in the Mott–Schottky (MS) situation, $\lambda^* > \lambda$ by a factor that depends on the square root of the space-charge potential; (iii) if $\Delta \phi_0$ does not significantly increase with increasing temperature, the ratio $c_{j,0}/c_{j,\infty}$ approaches unity with increasing temperature (see Eq. (4.30)); and (iv) the space-charge potential is related to the excess charge density in the grain boundary core through Eqs. (4.37) and (4.41) for the Mott–Schottky and the Gouy–Chapman situation, respectively.

4.5

Case Studies

4.5.1

Titanium Oxide (Rutile)

Owing to the high technological relevance of titanium oxide (TiO_2), its defect chemistry has been extensively investigated over the years [20–28]. Firstly, the case of nominally pure TiO_2 is considered (Figure 4.6).

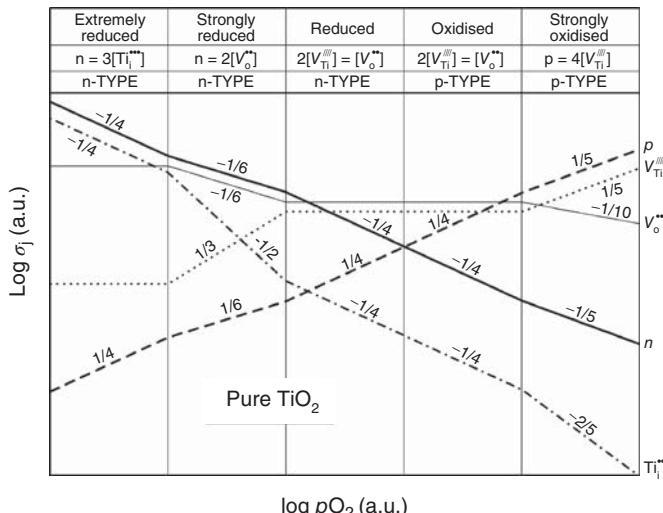
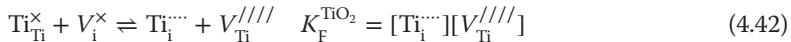


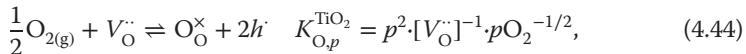
Figure 4.6 Brouwer diagram of nominally pure TiO_2 . (Reprinted with permission from Ref. [28]. © (2008) American Chemical Society.)

4.5.1.1 Nominally Pure TiO₂

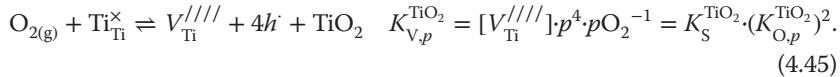
The formation reactions of intrinsic defects (Frenkel and Schottky disorder) are expressed as



- 1) Under oxidizing conditions, the electronic disorder reaction (4.14) together with the oxygen incorporation reaction shall be considered



which, if combined with the Schottky disorder reaction, leads to



The electroneutrality condition reads (note that the electronic defect concentration under oxidizing conditions is much lower than the ionic defect concentration)

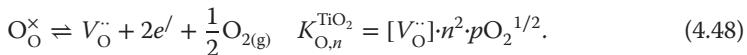
$$2[V_{\text{Ti}}^{\cdots\cdots\cdots}] = [V_{\text{O}}^{\cdot\cdot}] \quad (4.46)$$

yielding

$$p = (K_{\text{V},p}^{\text{TiO}_2})^{1/4} \left(\frac{K_{\text{S}}^{\text{TiO}_2}}{4} \right)^{-1/12} \quad p\text{O}_2^{1/4} \quad (4.47)$$

and correspondingly $n = (K_{\text{B}}^{\text{TiO}_2}/p) \propto p\text{O}_2^{-1/4}$.

- 2) In a reducing environment, the oxygen excorporation reaction should be considered



As long as the electroneutrality condition $2[V_{\text{O}}^{\cdot\cdot}] = n$ holds, the majority defects are electrons, and their concentration is determined by

$$n = (2K_{\text{O},n}^{\text{TiO}_2})^{1/3} \cdot p\text{O}_2^{-1/6}. \quad (4.49)$$

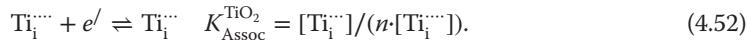
- 3) However, if the atmosphere is highly reducing, then the formation of titanium interstitials with a perceptible mobility at elevated temperature also plays a crucial role. It has been indeed observed that – at least at high temperature – $[\text{Ti}_{\text{i}}^{\cdots\cdots}]$ is slightly greater than $[V_{\text{O}}^{\cdot\cdot}]$ [27], meaning that the following should be taken into account:



In such a situation, the electroneutrality is fulfilled by $4[\text{Ti}_{\text{i}}^{\cdots\cdots}] = n$, from which

$$n = (4K_{\text{Ti},n}^{\text{TiO}_2})^{1/5} \cdot p\text{O}_2^{-1/5} \quad (4.51)$$

is obtained. For large concentrations of titanium interstitials and electrons, the possible association between electrons and $[Ti_i^{...}]$ needs to be considered:



In this case, the charge neutrality condition reads $3[Ti_i^{...}] = n$, leading to

$$n = (3K_{Assoc}^{TiO_2} K_{Ti,n}^{TiO_2})^{1/4} \cdot pO_2^{-1/4}. \quad (4.53)$$

4.5.1.2 Acceptor-Doped TiO_2

Acceptors are often present in TiO_2 as impurities (e.g., Fe^{3+}). At high pO_2 and for $[A_{Ti}^/] > 4[V_{Ti}^{///}]$, the electroneutrality condition is expressed as $[A_{Ti}^/] = 2[V_O^-]$, meaning that the oxygen-vacancy concentration is fixed by the concentration of acceptors. In this case, under oxidizing conditions, the oxygen incorporation reaction (4.44) must be considered, from which

$$p = \left(\frac{K_{O,p}^{TiO_2} \cdot [A_{Ti}^/]}{2} \right)^{1/2} \cdot pO_2^{1/4} \quad (4.54)$$

and $(n = (K_B^{TiO_2}/p) \propto pO_2^{-1/4})$ follows. As pointed out in Section 4.3.2.3, if the acceptor concentration is not pO_2 -invariant and changes instead according to Eq. (4.24), then the pO_2 dependence of the charge carrier concentrations is more complex and cannot be expressed by a simple analytical relationship. With decreasing pO_2 , the electron concentration increases up to $n \geq [A_M^/]$, for which the condition of charge neutrality reads $2[V_O^-] = n$. Under such conditions (and at even lower pO_2 values), the concentration of the various defects can be determined using Eqs. (4.49)–(4.53).

4.5.1.3 Donor-Doped TiO_2

Let us first consider the simple case, in which donors (here, assumed to be pentavalent cations) are compensated by electrons $[D_{Ti}^-] = n$. In such a situation, n is pO_2 -independent (as well as p , owing to the band–band transfer reaction). If, however, the conditions become more reducing to the extent that $[V_O^-] > [D_{Ti}^-]$, then electroneutrality is fulfilled by $2[V_O^-] = n$. From here on, if the oxygen partial pressure is further decreased, then the same considerations made for undoped titania under strongly reducing atmosphere are valid also in this case (Eqs. (4.49)–(4.53)).

On the other hand, with increasing pO_2 , two different situations can occur:

$$\text{for } [D_{Ti}^-] < 2[V_O^-], \quad 2[V_O^-] = 4[V_{Ti}^{///}] \quad (4.55)$$

or

$$\text{for } [D_{Ti}^-] > 2[V_O^-], \quad [D_{Ti}^-] = 4[V_{Ti}^{///}]. \quad (4.56)$$

For the first situation, the same relations written for undoped titania under oxidizing conditions hold (see Eqs. (4.46) and (4.47)). If instead Eq. (4.56) is valid,

Table 4.1 Thermodynamic data (see also Sections 4.3.1 and 4.3.2) of the most relevant defect chemistry reactions in TiO_2 [28].

Equilibrium constant	Δ° (eV)	Δ° (meV K $^{-1}$)	References
$K_{\text{O},n}^{\text{TiO}_2}$	5.13	1.1	[28]
$K_{\text{Ti},n}^{\text{TiO}_2}$	10.66	2.48	[28]
	10.67	—	[29]
	9.6	—	[30]
$K_{\text{B}}^{\text{TiO}_2}$	2.31	0.46	[28]
	3.7	—	[29]
	3.05	—	[31]
	3.3	—	[30]
$K_{\text{V},p}^{\text{TiO}_2}$	3.68	-2.1	[28]
$K_{\text{Assoc}}^{\text{TiO}_2}$	1.52	0.5	[28]
	1.43	—	[29]

then the following expressions are obtained:

$$p = (K_{\text{V},p}^{\text{TiO}_2})^{1/4} \cdot \left(\frac{4}{[\text{D}_{\text{Ti}}]} \right)^{1/4} \cdot p_{\text{O}_2}^{1/4} = (K_{\text{O},p}^{\text{TiO}_2})^{1/2} \cdot \left(\frac{4K_{\text{S}}^{\text{TiO}_2}}{[\text{D}_{\text{Ti}}]} \right)^{1/4} \cdot p_{\text{O}_2}^{1/4} \quad (4.57)$$

and $n = K_{\text{B}}^{\text{TiO}_2}/p \propto p_{\text{O}_2}^{-1/4}$. The thermodynamic data of the most relevant defect chemistry reactions in titania are listed in Table 4.1.

4.5.1.4 The Role of Dislocations

Recently, the relevance of dislocations with regard to the electrical transport properties and, hence, the point defect chemistry of TiO_2 has been proven on a set of single crystals, which had been plastically deformed under compression in order to induce the formation of dislocations [32]. As shown in Figure 4.7, electrical measurements performed along the same direction of the dislocation slip plane [001] revealed a significant increase of the transport properties under oxidizing conditions.

Quite remarkably, upon mechanical deformation, the conductivity was not only improved but also was switched from p-type (slope 1/4 of the square symbols in the log–log plot of Figure 4.7) to ionic (oxygen vacancies as well as titanium interstitials). Such findings could be explained in the framework of the space-charge model, according to which an excess negative charge (i.e., negative potential $\Delta\phi_0$) lies in the dislocation core (due to titanium vacancies), leading to an accumulation of the positively charged mobile defects. It is worth noting that such an enrichment effect is expected to be more pronounced for oxygen vacancies and particularly titanium interstitials than for holes due to the larger charge of the ionic defects (cf. Eq. (4.30)).

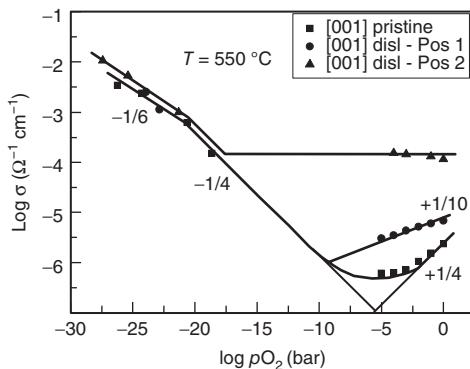


Figure 4.7 Electrical conductivity as a function of oxygen partial pressure of a pristine TiO_2 single crystal (squares) compared with a plastically deformed crystal (triangles and circles) upon compression along the [001]

direction. Positions 1 and 2 refer to different locations in the deformed crystal, which contained a different density of dislocations. (Reproduced from Ref. [32] by permission of John Wiley & Sons, Inc.)

4.5.2

Strontium Titanate

Strontium titanate is an ABO_3 perovskite with Sr occupying the A sites and Ti ions on the B sites. Sr and O form a close-packed cubic structure, in which Ti is located in 1/4 of the octahedral voids (in the center of the cube). This compact atomic arrangement has obviously a great impact on the point defect chemistry of this compound, resulting in an easier formation of vacancies (Schottky defects) rather than interstitials (Frenkel defects). In contrast to TiO_2 , the mobility of cation defects is extremely low in perovskites. Indeed, together with oxygen incorporation Eq. (4.19), oxygen excorporation Eq. (4.16), and the electronic disorder reaction (4.14), the partial Schottky disorder is also taken into account [11]



$$K_S^{\text{STO}}(T) = [V_{\text{Sr}}^{/\!/}][V_{\text{O}}^{\cdot\cdot}] = K_S^{\text{STO},0}(T) \exp\left(\frac{-\Delta_S H^\circ}{k_B T}\right), \quad (4.59)$$

in which the activity of SrO is assumed to be invariant [33].

4.5.2.1 Acceptor-Doped SrTiO_3

Acceptor doping is usually achieved by substituting Ti with a trivalent cation. Under oxidizing atmosphere, the oxygen incorporation reaction together with the electroneutrality condition $[\text{A}_{\text{Ti}}^{\text{/}}] = 2[V_{\text{O}}^{\cdot\cdot}]$ yields

$$p = \left(\frac{K_{\text{O},p}^{\text{STO}} \cdot [\text{A}_{\text{Ti}}^{\text{/}}]}{2} \right)^{1/2} \cdot p\text{O}_2^{+1/4}, \quad (4.60)$$

while, under reducing conditions, the oxygen exorporation reaction (together with $[A'_{Ti}] = 2[V'_O]$) results in

$$n = (2K_{O,n}^{\text{STO}})^{1/2} \cdot [A'_{Ti}]^{-1/2} \cdot pO_2^{-1/4}. \quad (4.61)$$

As in the case of TiO_2 , if the acceptor concentration is not pO_2 -independent (see reaction (4.24)), then the oxygen partial pressure dependence of the concentration of the mobile charge carriers becomes rather complex. For more severely reducing conditions, under which $n = 2[V'_O]$ holds,

$$n = (2K_{O,n}^{\text{STO}})^{1/3} \cdot pO_2^{-1/6}$$

is expected.

4.5.2.2 Donor-Doped SrTiO_3

In contrast to the acceptor case, donor doping can be achieved by adding either a trivalent cation occupying the A site (e.g., $\text{La}^{+}_{\text{Sr}}$) or a pentavalent cation occupying the B lattice site (e.g., $\text{Nb}^{+5}_{\text{Ti}}$). For this reason, the donor concentration is written in the following as $[D']$ without specifying the host lattice site. For an exhaustive treatment of the donor-doped case, the reader is referred to Ref. [33].

While, under highly reducing conditions (for which the electroneutrality condition is $n = 2[V'_O]$), the considerations made above also hold for the case of donor doping, for moderately reducing and oxidizing environments, the electroneutrality conditions read $n = [D']$ and $2[V''_{\text{Sr}}] = [D']$, respectively. Under a moderately reducing atmosphere, while n and p are fixed by the donor concentration,

$$[V'_O] = K_{O,n}^{\text{STO}} \cdot [D']^{-2} \cdot pO_2^{-1/2} \quad (4.62)$$

and $[V''_{\text{Sr}}] = (K_S^{\text{STO}} / [V'_O]) \propto pO_2^{1/2}$ result.

For $2[V''_{\text{Sr}}] = [D']$, the concentration of the oxygen vacancies is given by

$$[V'_O] = \frac{2K_S^{\text{STO}}}{[D']}, \quad (4.63)$$

whereas

$$p = (2K_{O,p}^{\text{STO}} \cdot K_S^{\text{STO}})^{1/2} \cdot [D']^{-1/2} \cdot pO_2^{1/4} \quad (4.64)$$

and $n = (K_B^{\text{STO}} / p) \propto pO_2^{-1/4}$.

Note that the involvement of low mobile strontium vacancies leads to much larger equilibration times than with oxygen vacancies in the acceptor-doped case [34, 35]. Table 4.2 summarizes the thermodynamic data of the the most important defect chemistry reactions in strontium titanate.

4.5.2.3 Grain Boundaries in SrTiO_3

Over the years, several contributions have demonstrated the blocking character of grain boundaries of nominally pure or acceptor-doped SrTiO_3 for holes and oxygen-vacancy migration (under oxidizing atmosphere), pointing toward the boundaries to be positively charged [36–40]. In particular, DC-biased impedance

Table 4.2 Equations and constants of the most relevant defect chemistry reactions in SrTiO₃ [11, 33].

Equation	Constant values	References
$K_S^{\text{STO}} = K_S^{\text{STO},0} e^{-\frac{\Delta_S H^\circ}{k_B T}}$	$K_i^{\text{STO},0} = 3 \times 10^{44} \text{ cm}^{-6}$ $\Delta_S H^\circ = 2.5 \text{ eV}$	[33]
$K_B^{\text{STO}} = K_B^{\text{STO},0}$ $= N_C(T)N_V(T)e^{-\frac{(\Delta_B H^\circ - \beta T)}{k_B T}}$	$K_B^{\text{STO},0} = 7.67 \times 10^{42} \text{ cm}^{-6}$ $\Delta_B H^\circ(0 \text{ K}) = 3.3 \text{ eV}$ $\beta = 6.0 \times 10^{-4} \text{ eV K}^{-1}$ $N_C(T) = 4.1 \times 10^{16} \text{ cm}^{-3} (T/\text{K})^{3/2}$ $N_V(T) = 3.5 \times 10^{16} \text{ cm}^{-3} (T/\text{K})^{3/2}$ $\Delta_B H^\circ(0 \text{ K}) = 3.17 \text{ eV}$ $\beta = 5.66 \times 10^{-4} \text{ eV K}^{-1}$	[11]
$K_{O,n}^{\text{STO}} = K_{O,n}^{\text{STO},0} e^{-\frac{\Delta_R H^\circ}{k_B T}}$	$K_{O,n}^{\text{STO},0} = 5 \times 10^{71} \text{ cm}^{-9} \text{ bar}^{1/2}$ $\Delta_R H^\circ = 6.1 \text{ eV}$	[33]
$K_{O,p}^{\text{STO}} = K_{O,p}^{\text{STO},0} e^{-\frac{\Delta_O H^\circ}{k_B T}}$	$K_{O,p}^{\text{STO},0} = 1.02 \times 10^{23} \text{ Pa}^{-1/2} \text{ cm}^{-3}$ $\Delta_O H^\circ = 1.63 \text{ eV}$	[11]

spectroscopy studies performed on SrTiO₃ bicrystals clearly proved the presence of positively charged interfaces [39, 40].

Due to the rather large dielectric constant of this oxide, the SCL extent can be rather broad (see Eqs. (4.36) and (4.39)), particularly in the case of slightly doped compositions. The investigation of nanocrystalline nominally undoped SrTiO₃ (with a residual acceptor concentration on the order of 100 ppm) revealed the intriguing situation of samples in which the average grain size was smaller than the extent of the SCL (mesoscopic situation) [41–43]. Figure 4.8 shows the electrical conductivity of mesoscopic SrTiO₃ with an average grain size of 30 and 50 nm (in which $2\lambda^*$ is on the order of 80 nm) [42]. Remarkably, compared with the bulk transport properties of a sample with the same nominal composition, the p-type conductivity (characterized by the positive slope under oxidizing conditions) decreases by almost 4 orders of magnitude, while under reducing atmosphere, the n-type conductivity (negative slope) is increased by almost 3 orders of magnitude. Moreover, in the nanocrystalline samples, the p–n transition is shifted up to 12 orders of magnitudes in terms of oxygen partial pressure [42]. It is noteworthy that all these effects can be explained by a change in the concentration profiles due to space-charge effects occurring in proximity of the grain boundaries.

As far as line defects are concerned, electron microscopy studies have convincingly shown that the core of edge dislocations in SrTiO₃ exhibits a lower occupancy of the oxygen sites, which point toward the local oxygen-vacancy concentration being larger than in the bulk [1–3]. This obviously suggests the core to be electrically charged. It has been speculated that this oxygen nonstoichiometry is related to a fast oxygen diffusion along dislocations (or network of

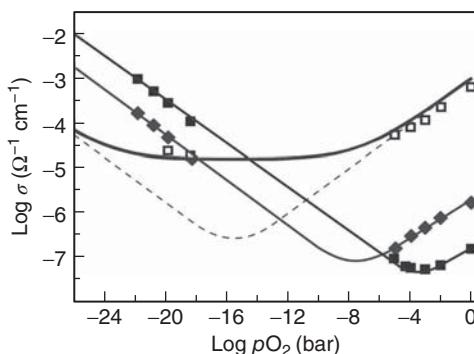


Figure 4.8 Electrical conductivity as a function of the oxygen partial pressure p_{O_2} of (i) the bulk contribution of a microcrystalline sample (open squares), (ii) a nanocrystalline sample with an average grain size of 50 nm (gray diamonds), and (iii) a nanocrystalline sample with an average grain size of 30 nm (solid squares) measured at 544 °C. All the samples were prepared from the

same batch of SrTiO_3 powder containing about 100 ppm Fe impurities. The solid gray line represents a plateau at p_{O_2} ranging between 10^{-20} and 10^{-10} atm corresponding to the conductivity calculated from the defect chemistry model of Ref. [11] for a concentration of Fe impurities of 100 ppm. (Reproduced from Ref. [42] by permission of John Wiley & Sons, Inc.)

dislocations), which under reducing environment can quickly create electrically conducting pathways along dislocation networks [44, 45]. Although the ability to modify the local electrical conduction properties has been shown [45] and can be employed for resistive-switching applications [46], there are no indications of enhanced oxygen diffusivity along the dislocations cores in SrTiO_3 as discussed in Chapter 5.

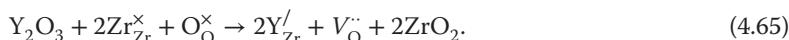
4.5.3

Zirconium and Hafnium Oxide

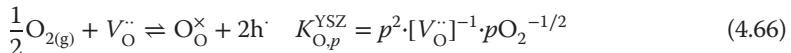
4.5.3.1 Zirconium Oxide

Zirconium oxide (ZrO_2) is an oxide of great technological relevance characterized by three polymorphic phases: monoclinic between room temperature and ~ 1170 °C, tetragonal up to 2370 °C, and finally, cubic up to the melting point [47]. The addition of hypovalent cations (e.g., Ca^{2+} , Y^{3+} , Yb^{3+} , Sc^{3+}) stabilizes the two high-temperature phases down to room temperature. For this, the acceptor concentration is decisive: for example, 7 at% Y-doped ZrO_2 (or 3.5 mol% Y_2O_3) is tetragonal while 16 at% Y-doped ZrO_2 (or 8 mol%, also known as 8YSZ) is cubic. Remarkably, the cubic phase provides the high ionic (oxygen vacancies) conductivity at high temperatures, which has made this oxide the material of choice as an electrolyte for various applications [48–54].

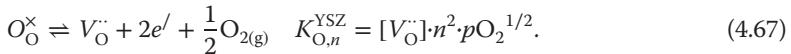
The addition of hypovalent cations leads to the formation of oxygen vacancies according to the following reaction, which is written for the case of Y^{3+} :



The other relevant defect chemistry reactions are, at high oxygen partial pressure, the oxygen incorporation reaction



and, at low oxygen partial pressures, the oxygen excorporation reaction:



Moreover, the electronic disorder reaction (4.14) must also be taken into account. For high $p\text{O}_2$ values, the global electroneutrality condition can be written as

$$[Y_{\text{Zr}}^{\prime}] = 2[V_{\text{O}}^{\cdot}] + p. \quad (4.68)$$

However, for the typical acceptor concentrations used for technological applications (e.g., 8YSZ), $[V_{\text{O}}^{\cdot}] \gg p$, and thus, $[Y_{\text{Zr}}^{\prime}] = 2[V_{\text{O}}^{\cdot}]$ holds. A similar reasoning also applies for the low $p\text{O}_2$ range, for which $[Y_{\text{Zr}}^{\prime}] \gg n$ and hence $[Y_{\text{Zr}}^{\prime}] = 2[V_{\text{O}}^{\cdot}]$ results. In the light of this situation, the total conductivity of heavily acceptor-doped zirconia is ionic over a broad range of oxygen partial pressures.

It is noteworthy that, in heavily doped zirconia, due to the high oxygen vacancy concentration, the activities a_i should be considered in the mass action laws. Thus, as indicated in, for example, Ref. [55], the oxygen incorporation and excorporation reactions should be rewritten as

$$\overline{K}_{\text{O},p}^{\text{YSZ}} = a_p^2 \cdot (a_{V_{\text{O}}^{\cdot}})^{-1} \cdot p\text{O}_2^{-1/2} = p^2 \cdot (a_{V_{\text{O}}^{\cdot}})^{-1} \cdot p\text{O}_2^{-1/2} = p^2 \cdot (f_{V_{\text{O}}^{\cdot}} [V_{\text{O}}^{\cdot}])^{-1} \cdot p\text{O}_2^{-1/2}, \quad (4.69)$$

and

$$\overline{K}_{\text{O},n}^{\text{YSZ}} = f_{V_{\text{O}}^{\cdot}} [V_{\text{O}}^{\cdot}] \cdot n^2 \cdot p\text{O}_2^{1/2}, \quad (4.70)$$

respectively, in which the activity coefficient $f_{V_{\text{O}}^{\cdot}}$ depends on $[Y_{\text{Zr}}^{\prime}]$, but it is practically independent of the minority charge carrier concentrations. Moreover, $f_{V_{\text{O}}^{\cdot}}$ is $p\text{O}_2$ -invariant within the range of oxygen partial pressure, in which $2[V_{\text{O}}^{\cdot}] = [Y_{\text{Zr}}^{\prime}]$ holds, while $f_{V_{\text{O}}^{\cdot}}$ is temperature-dependent particularly in the low-temperature range, at which defect associations become relevant [55].

Early electrical transport studies on YSZ already revealed a change in the temperature dependence of the conductivity below temperatures around (depending on the specific composition) 550 °C [51–54, 56]. The correspondingly greater measured activation energies are due to an additional enthalpic term stemming from the energy of association between defects. Similarly, defect association and clustering are also responsible for the drop in conductivity when the acceptor concentration exceeds a certain optimum value, which for YSZ single crystals occurs for 8 mol% Y_2O_3 [54]. With increasing dopant concentration, the migration barrier of oxygen moving through Zr–Y or a Y–Y nearest neighbor pair is larger than that in the case of Zr–Zr [57, 58].

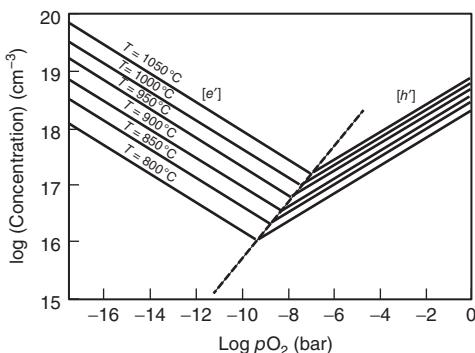


Figure 4.9 Concentration of the minority charge carriers as a function of the oxygen partial pressure. Note that the oxygen-vacancy concentration in 8YSZ is on the

order of $2 \cdot 10^{21} \text{ cm}^{-3}$. (Reproduced from Ref. [63] by permission from ECS – The Electrochemical Society.)

In this context, it is worth emphasizing that defect interactions involve both cation–vacancy and vacancy–vacancy interactions. However, according to computer simulation studies, while cation–vacancy interactions can be minimized by choosing a trivalent dopant with an ionic radius comparable with the one of zirconium (e.g., scandium), vacancy–vacancy interactions appear to be an intrinsic property of the material [59, 60].

Quite remarkably, the trivalent dopant distribution (and, consequently, the oxygen-vacancy distribution) in a nominally homogeneously doped zirconia is not random, but it forms a rather complex structure (cf. Ref. [61] and references therein). Interestingly, not only cation clusters tend to trap oxygen vacancies and keep them immobile even at high temperatures but, with increasing dopant concentration, but also single and pairs of vacancies tend to align along preferential lattice directions.

As far as the minority charge carriers (holes and electrons) are concerned, oxygen permeation and Hebb–Wagner polarization has been employed to determine their concentrations and mobilities [62–64]. Figure 4.9 illustrates the concentration of the minority charge carriers (electrons and holes) as a function of the oxygen partial pressure and different temperatures for 8 mol% Y_2O_3 -doped ZrO_2 [62].

Obviously, heavily doped zirconia also can contain impurities, which (particularly if redox-active) can affect the defect chemistry of the minority charge carriers; the reader is referred to Ref. [55] (and references therein) for a detailed treatment.

4.5.3.2 The Role of Grain Boundaries and Dislocations

There is a vast literature showing that grain boundaries tend to hinder the ionic transport of acceptor-doped zirconia. The results of conductivity measurements performed on 8YSZ at temperatures ranging between 250 and 500 °C can be interpreted in terms of the space-charge model, meaning that the lower ionic

conductivity at the grain boundary results from a depletion of oxygen vacancies due to an excess positive charge lying in the grain boundary core [65, 66]. On the other hand, dislocations arising from the plastic deformation of YSZ single crystals seem not to improve the electrical transport of the material significantly [67].

4.5.3.3 Hafnium Oxide

Hafnium oxide (HfO_2) has attracted much interest in the past years because, owing to its large (as compared to SiO_2) relative dielectric constant ($\epsilon_r = 26$), it has been considered a promising candidate for gate dielectric material in metal-oxide-semiconductor transistors [68].

The ionic radii and the ionization energies of Zr and Hf are almost exactly identical. As a consequence, all redox and defect-related properties are very similar. Only properties that relate to the different mass of the ions, such as the phonon spectra and, hence, also the phase transformations, are somewhat different. Analogously to ZrO_2 , HfO_2 also undergoes two phase transformations, which, however, occur at higher temperatures compared to zirconia: the monoclinic-to-tetragonal transformation starts at about $1540\text{ }^\circ\text{C}$, while the transition from tetragonal into cubic is observed at around $2600\text{ }^\circ\text{C}$ [69–71]. Also, the cubic fluorite-type structure of hafnia can be stabilized at room temperature by the addition of divalent (e.g., Ca^{2+}) or trivalent cations (e.g., Yb^{3+} , Y^{3+} , Sm^{3+}). Obviously, the insertion of acceptors results in the formation of oxygen vacancies as described by Eq. (4.65) for the case of ZrO_2 .

Several studies performed on acceptor-doped HfO_2 at high temperatures have revealed a predominant oxygen-vacancy conductivity over a broad range of oxygen partial pressures, whose extent depends on the acceptor concentration [72–75]. Under comparable operation conditions and acceptor concentrations, the ionic conductivity of HfO_2 is, however, lower than the ionic conductivity of ZrO_2 . Under oxidizing conditions, a positive partial pressure dependence can be detected, indicating p-type conductivity (see, e.g., Figure 4.4).

A systematic study on the role of different dopants and dopant concentrations in the ionic conductivity of HfO_2 revealed that Yb doping yields the largest conductivity, while for three different dopants (Yb_2O_3 , Y_2O_3 , and Sm_2O_3), the maximum conductivity corresponds to an acceptor concentration of about 10 mol% [75].

Calorimetric data suggests the tendency of strong short-range ordering between oxygen vacancies and cations in cubic acceptor-doped HfO_2 . The ionic radius of the dopant seems to play the key role in inducing the vacancy to be located close to Hf (large radius) or the aliovalent cation (small radius) [76].

4.5.4

Aluminum Oxide

Aluminum oxide (in particular, corundum, $\alpha\text{-Al}_2\text{O}_3$) is a material of great technological relevance for a large number of different applications. It has driven, over the years, a vibrant research activity aiming at understanding in particular its high-temperature properties and processes, which are related to its ionic transport and

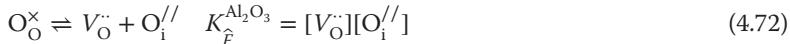
point defect chemistry. Most of the studies devoted to this scope can be grouped into (i) computational methods on the one hand [77–83] and (ii) diffusion [84–88] and electrical transport [89–91] investigations on the other.

The intrinsic point defect chemistry of $\alpha\text{-Al}_2\text{O}_3$ is defined by the following reactions as well as the electronic disorder (4.14):

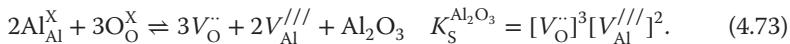
Frenkel disorder



Anti-Frenkel disorder



Schottky disorder



It is worth noting that the formation energy of such intrinsic defects has been estimated to be rather high: [77, 79–81]. For Schottky defects, it ranges between 4.18 and 5.86 eV per defect; for Frenkel defects, between 5.22 and 10.0 eV; and for anti-Frenkel defects, between 3.79 and 8.27 eV. It is evident that these rather broad and overlapping energy ranges make it difficult to predict the majority native defects (intrinsic case). Such high formation energies have further relevant consequences: (i) Al_2O_3 starts to exhibit a perceptible electrical conductivity only at very high temperatures ($T > 1500^\circ\text{C}$) and (ii) its defect chemistry – and hence its transport properties – are dominated by the presence of even small amounts (i.e., a few atomic ppm) of impurities. It shall be noticed that here the association between charged defects is omitted for the sake of simplicity, but, particularly for intermediate- and low-temperature situations, this association should also be taken into account. Please refer to, for example, Refs. [80] and [81] for a description of the possible associates.

4.5.4.1 Acceptor-Doped Alumina

In the case of acceptor-doped alumina (e.g., in the presence of $\text{Mg}_{\text{Al}}^{\prime\prime}$), the global charge neutrality condition is fulfilled by

$$[\text{Mg}_{\text{Al}}^{\prime\prime}] = 3[\text{Al}_i^{\cdot\cdot\cdot}] \quad (4.74)$$

or

$$[\text{Mg}_{\text{Al}}^{\prime\prime}] = 2[V_\text{O}^{\cdot\cdot}]. \quad (4.75)$$

Both electrical transport [89] and oxygen diffusion studies [84] indicate the second condition to be valid. In this case, the oxygen incorporation reaction reads



$$K_{\text{O},A}^{\text{Al}_2\text{O}_3} = [\text{Mg}_{\text{Al}}^X]^2 \cdot [V_\text{O}^{\cdot\cdot}]^{-1} \cdot [\text{Mg}_{\text{Al}}^{\prime\prime}]^{-2} \cdot p\text{O}_2^{-1/2}.$$

$$\text{For } [\text{Mg}_{\text{Al}}^X] = [\text{Mg}_{\text{Total}}] > [\text{Mg}_{\text{Al}}^{\prime\prime}] = 2[V_\text{O}^{\cdot\cdot}] \quad (4.77)$$

$$[V_O^{\cdot}] \propto pO_2^{-1/6}; [Mg_{Al}^{\cdot}] \propto pO_2^{-1/6}, \quad (4.78)$$

and from

$$Mg_{Al}^{\times} \rightleftharpoons Mg_{Al}^{\prime} + h^{\cdot} \quad (4.79)$$

$p \propto pO_2^{1/6}$ results.

The conductivity is p-type at high oxygen partial pressures (and T exceeding 1500 °C) and ionic at low oxygen partial pressure.

4.5.4.2 Donor-Doped Alumina

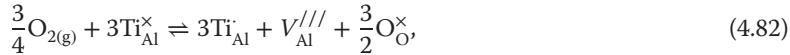
In the case of donor-doped alumina (e.g., Ti_{Al}^{\cdot}), two situations that fulfill the global charge neutrality condition can be considered:

$$[Ti_{Al}^{\cdot}] = 3[V_{Al}^{\prime\prime\prime}] \quad (4.80)$$

or alternatively,

$$[Ti_{Al}^{\cdot}] = 2[O_i^{\prime/}]. \quad (4.81)$$

In the first case [90], the oxidation reaction can be written as



which together with the dopant oxidation reaction

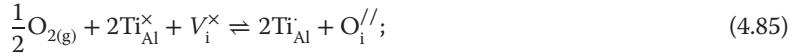


yields

$$K_{O,D}^{Al_2O_3} = [Ti_{Al}^{\cdot}]^3 \cdot [V_{Al}^{\prime\prime\prime}] \cdot [Ti_{Al}^{\times}]^{-3} \cdot pO_2^{-3/4} = [V_{Al}^{\prime\prime\prime}] \cdot n^{-3} \cdot pO_2^{-3/4}, \quad (4.84)$$

from which $n \propto pO_2^{-1/4}$ is obtained.

If Eq. (4.81) holds instead [91], the oxidation reaction reads



$$K_{O,D}^{Al_2O_3} = [Ti_{Al}^{\cdot}]^2 \cdot [O_i^{\prime/}] \cdot [Ti_{Al}^{\times}]^{-2} \cdot pO_2^{-1/2} \propto [O_i^{\prime/}] \cdot n^{-2} \cdot pO_2^{-1/2},$$

from which $n \propto pO_2^{-1/4}$ also results. Therefore, in both situations, n-type conductivity is expected at high oxygen partial pressure (and at high temperatures, $T > 1500$ °C), while the ionic transport dominates at low pO_2 .

Both models were found to describe rather well the conductivity data acquired from Ti-doped single crystal, with only a slightly better agreement with the $[Ti_{Al}^{\cdot}] = 3[V_{Al}^{\prime\prime\prime}]$ case [90]. Moreover, recent tracer diffusion experiments carried out on Ti-doped Al_2O_3 single crystals [92] revealed the aluminum diffusion coefficient to be orders of magnitude larger than the oxygen one indicating that $[Ti_{Al}^{\cdot}] = 3[V_{Al}^{\prime\prime\prime}]$ should be considered. Although scattered within 1 order of magnitude, the oxygen diffusion coefficients are in fairly good agreement with the defect chemistry and the electroneutrality conditions illustrated above.

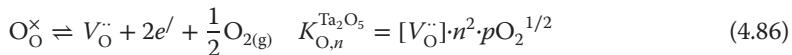
Rather, the experimental activation energies (ranging between 1 and 2.5 eV) are not consistent with the values predicted by the computational studies (about 5 eV). Finally, it is worth mentioning that such discrepancies between theoretical predictions and experimental findings – which have also been described as a “conundrum in corundum” [85] – recently suggested the (controversial) existence of “AlO” vacancies [87, 88].

4.5.5

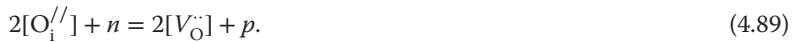
Tantalum Oxide

Tantalum oxide (Ta_2O_5) exhibits two stable phases at high temperature: the tetragonal (α - Ta_2O_5) and the orthorhombic [93] or monoclinic [94] (β - Ta_2O_5) phase. The transition from β to α occurs at a temperature from about 1320 °C [93, 95] to 1360 °C [94, 96]. Between room temperature and about 320 °C, a monoclinic phase has also been observed [94]. Similarly to the case of hafnia, β - Ta_2O_5 has also been intensively investigated for its dielectric properties for applications as high-k-dielectric in metal-oxide-semiconductor transistors. The α - Ta_2O_5 phase attracted attention in the past because of its interesting ionic transport properties.

Following high-temperature conductivity measurements performed on β - Ta_2O_5 over a broad range of oxygen partial pressures, the defect chemistry of such a compound can be described by the following reactions [97–102]:



in addition to the electroneutrality condition, which in the intrinsic situation reads



At low $p\text{O}_2$, Eq. (4.89) can be rewritten as $n = 2[V_\text{O}^\cdot]$, meaning that the electron concentration varies with the oxygen partial pressure according to

$$n = (2K_{\text{O},n}^{\text{Ta}_2\text{O}_5})^{1/3} p\text{O}_2^{-1/6}. \quad (4.90)$$

If, instead, Ta_2O_5 contains acceptor impurities ($\text{A}_{\text{Ta}}^{//}$ = e.g., Al^{3+} , Cr^{3+} , Fe^{3+} , Sc^{3+}), then the electron concentration in moderately reducing atmosphere ($[\text{A}_{\text{Ta}}^{//}] > n$) can be expressed in the following form:

$$n = \left(\frac{K_{\text{O},n}^{\text{Ta}_2\text{O}_5}}{[\text{A}_{\text{Ta}}^{//}]} \right)^{1/2} p\text{O}_2^{-1/4}, \quad (4.91)$$

while under strongly reducing conditions, the intrinsic situation (Eq. (4.90)) still holds. At higher $p\text{O}_2$ values, the conductivity is $p\text{O}_2$ -invariant, meaning that the

oxygen vacancies are the majority mobile charge defects, whereas under oxidizing conditions, the increasing conductivity with increasing pO_2 point toward p-type conductivity resulting from:

$$p = K_B^{Ta_2O_5} \left(\frac{K_{O,n}^{Ta_2O_5}}{\left[A_{Ta}^{|||} \right]} \right)^{-1/2} p O_2^{1/4}. \quad (4.92)$$

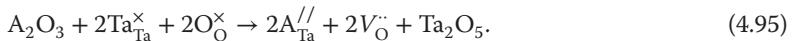
In principle, holes responsible for the observed p-type conductivity can also be compensated by $O_i^{|||}$ or $V_{Ta}^{|||||}$. However, due to the estimated large enthalpy of formation of such defects, it is expected that even a small concentration of acceptor impurities (the solubility of aliovalent cations in Ta_2O_5 is rather limited) [103] exceeds the intrinsic concentrations of oxygen interstitials and tantalum vacancies [97, 104].

As pointed out previously [104], for a more precise treatment of the defect chemistry, the concentration of aliovalent cations cannot be assumed to be constant over the whole range of temperatures and pO_2 considered, owing to association/dissociation, segregation, or dissolution reactions. Thus, by taking into account the temperature dependence of the acceptor concentration, the following quantities were determined [104] for nominally undoped Ta_2O_5 containing various acceptor impurities [98]

$$K_{O,n}^{Ta_2O_5} = 2 \times 10^{72} \exp \left(\frac{-\Delta_O H^\circ}{k_B T} \right) \text{ cm}^{-9} \text{ Pa}^{1/2}, \quad \Delta_O H^\circ = 5.6 \text{ eV} \quad (4.93)$$

$$K_{sol}^{Ta_2O_5} = 4 \times 10^{42} \exp \left(\frac{-\Delta_{sol} H^\circ}{k_B T} \right) \text{ cm}^{-6}, \quad \Delta_{sol} H^\circ = 1.54 \text{ eV}. \quad (4.94)$$

It is important to note that the dopants A are treated here as a second phase or precipitate (A_2O_3) and $K_{sol}^{Ta_2O_5}$ is the equilibrium constant of the dissolution reaction of the precipitate



As far as the high-temperature phase is concerned, α - Ta_2O_5 can be stabilized at lower temperatures by adding moderate amounts (2–5 mol%) of different oxides such as SnO_2 , Ga_2O_3 , Cr_2O_3 , MgO , and Sc_2O_3 [105]. However, the effect of the addition of hypovalent cations seems to be controversial. At first, it was observed that the insertion of acceptors leads to an enhanced ionic transport (oxygen vacancies), which in the case of 2 mol% Sc_2O_3 doping dominates the electrical properties of the material down to $pO_2 \sim 10^{-15}$ atm at 750 °C. In this case, the bulk ionic conductivity of stabilized α - Ta_2O_5 exceeded the ionic conductivity of β - Ta_2O_5 by 2 orders of magnitude under the same experimental conditions [106]. The corresponding bulk activation energy was calculated to be about 1 eV, while the grain boundaries seemed to hinder the ionic transport [106]. The observed discrepancy between the dopant concentration and the resulting ionic conductivity was interpreted in terms of a local rearrangement of the α - Ta_2O_5 constituting polyhedra owing to the

Table 4.3 Relevant thermodynamic data for $\alpha\text{-Ta}_2\text{O}_5$ and $\beta\text{-Ta}_2\text{O}_5$.

Phase	Δ_o ° (eV)	Δ_B ° (eV)	References
$\alpha\text{-Ta}_2\text{O}_5$	6.2	—	[106]
	4.87–5.27	—	[107]
$\beta\text{-Ta}_2\text{O}_5$	5.7	3.6	[98]
	5.13	—	[99]
	4.82	—	[100]
	5.0	—	[108]
	5.63	—	[109]
	—	3.58	[97]
	—	2.9	[110]

presence of the dopant, which could induce a disadvantageous variation of the local anion-to-cation ratio. As expected, the n-type conductivity became predominant at lower $p\text{O}_2$, and its dependence on the partial pressure followed Eq. (4.91). Interestingly, the n-type conductivity was observed to be 1 order of magnitude larger than for $\beta\text{-Ta}_2\text{O}_5$ under the same conditions. At very high temperatures (e.g., $T = 1440^\circ\text{C}$), indication of p-type conductivity for $\log(p\text{O}_2) > -3$ atm was also observed [106].

Nonetheless, in a later study, $\alpha\text{-Ta}_2\text{O}_5$ was found to be a good intrinsic ionic conductor irrespective of the nature and concentration of the dopants [107]. Moreover, longer exposure to 700–800 °C (aging) was found to result in a significant decrease in the ionic conductivity down to values comparable to those of $\beta\text{-Ta}_2\text{O}_5$. Such an effect has been interpreted in terms of local ordering or clustering of defects. In the aged state, consistently with the anisotropic nature of the $\alpha\text{-Ta}_2\text{O}_5$ crystal structure, the conductivity along the [100] direction was observed to be 2 orders of magnitude larger than the conductivity along the [001] direction [107].

Table 4.3 summarizes some of the relevant thermodynamic data associated with the defect formation reactions in tantalum oxide.

References

- Zhang, Z., Sigle, W., and Rühle, M. (2002) *Phys. Rev. B*, **66**, 094108.
- Jia, C.L., Thust, A., and Urban, K. (2005) *Phys. Rev. Lett.*, **95**, 225506.
- Jia, C.L., Houben, L., and Urban, K. (2006) *Philos. Mag. Lett.*, **86** (11), 683–690.
- (a) Klie, R.F. and Browning, N.D. (2000) *Appl. Phys. Lett.*, **77**, 3737–3739;
(b) Klie, R.F., Ito, Y., Stemmer, S., and Browning, N.D. (2001) *Ultramicroscopy*, **86**, 289–302.
- Jia, C.L. and Urban, K. (2004) *Science*, **303**, 2001–2004.
- von Alffhan, S., Benedek, N.A., Chen, L., Chua, A., Cockayne, D., Dudeck, K.J., Elsässer, C., Finnis, M.W., Koch, C.T., Rahmati, B., Rühle, M., Shih, S.-J., and Sutton, A.P. (2010) *Annu. Rev. Mater. Res.*, **40**, 557–599.

7. Wang, Z., Saito, M., McKenna, K.P., Gu, L., Tsukimoto, S., Shluger, A.L., and Ikuhara, Y. (2011) *Nature*, **479**, 380–383.
8. An, J., Koh, A.L., Park, J.S., Sinclair, R., Gür, T.M., and Prinz, F.B. (2013) *J. Phys. Chem. Lett.*, **4**, 1156–1160.
9. Kofstad, P. (1972) *Nonstoichiometry, Diffusion, and Electrical Conductivity in Binary Metal Oxides*, John Wiley & Sons, Inc., New York.
10. Maier, J. (2004) *Physical Chemistry of Ionic Materials*, John Wiley & Sons, Ltd, Chichester.
11. Denk, I., Münch, W., and Maier, J. (1995) *J. Am. Ceram. Soc.*, **78** (12), 3265–3272.
12. De Souza, R.A. (2009) *Phys. Chem. Chem. Phys.*, **11**, 9939–9969.
13. Lee, H.-S., Mizoguchi, T., Mistui, J., Yamamoto, T., Kang, S.-J.L., and Ikuhara, Y. (2011) *Phys. Rev. B*, **83**, 104110.
14. Nyman, B.J., Helgee, E.E., and Wahnström, G. (2012) *Appl. Phys. Lett.*, **100**, 061903.
15. Maier, J. (1985) *Ber. Bunsen Ges. Phys. Chem.*, **89**, 355–362.
16. Maier, J. (1986) *Ber. Bunsen Ges. Phys. Chem.*, **90**, 26–33.
17. Maier, J. (1995) *Prog. Solid State Chem.*, **23**, 171–263.
18. Kim, S., Fleig, J., and Maier, J. (2003) *Phys. Chem. Chem. Phys.*, **5**, 2268–2273.
19. Göbel, M.C., Gregori, G., and Maier, J. (2014) *Phys. Chem. Chem. Phys.*, **16** (21), 10214–10231.
20. Hurlen, T. (1959) *Acta Chem. Scand.*, **13** (2), 365–376.
21. Kofstad, P. (1962) *J. Phys. Chem. Solids*, **23**, 1559–1586.
22. Greener, E.H., Barone, F.J., and Hirthe, W.M. (1965) *J. Am. Ceram. Soc.*, **48** (12), 623–627.
23. Blumenthal, R.N., Baukus, J., and Hirthe, W.M. (1967) *J. Electrochem. Soc.*, **114** (2), 172–176.
24. Marucco, J.-F., Gautron, J., and Lemasson, P. (1981) *J. Phys. Chem. Solids*, **42**, 363–367.
25. Balachandran, U. and Eror, N.G. (1988) *J. Mater. Sci.*, **23**, 2676–2682.
26. (a) Bak, T., Nowotny, J., Rekas, M., and Sorrell, C.C. (2003) *J. Phys. Chem. Solids*, **64**, 1043–1056;(b) Bak, T., Nowotny, J., Rekas, M., and Sorrell, C.C. (2003) *J. Phys. Chem. Solids*, **64**, 1057–1067;(c) Bak, T., Nowotny, J., Rekas, M., and Sorrell, C.C. (2003) *J. Phys. Chem. Solids*, **64**, 1069–1087.
27. Lee, D.-K. and Yoo, H.-J. (2006) *Solid State Ionics*, **177**, 1–9.
28. Nowotny, M.K., Sheppard, L.R., Bak, T., and Nowotny, J. (2008) *J. Phys. Chem. C*, **112**, 5275–5300.
29. Blumenthal, R.N., Coburn, J., Baukus, J., and Hirthe, W.M. (1966) *J. Phys. Chem. Solids*, **27** (4), 643–654.
30. Lee, D.K., Jeon, J.L., Kim, M.H., Choi, W., and Yoo, H.I. (2005) *J. Solid State Chem.*, **178** (1), 185–193.
31. Cronemeyer, D.C. (1952) *Phys. Rev.*, **87** (5), 876–886.
32. Adeppali, K.K., Kelsch, M., Merkle, R., and Maier, J. (2013) *Adv. Funct. Mater.*, **23**, 1798–1806.
33. Moos, R. and Härdtl, K.H. (1997) *J. Am. Ceram. Soc.*, **80** (10), 2549–2563.
34. Meyer, R., Waser, R., Helmbold, J., and Borchardt, G. (2003) *Phys. Rev. Lett.*, **90** (10), 105901.
35. Gömann, K., Borchardt, G., Schulz, M., Gömann, A., Maus-Friedrichs, W., Lesage, B., Kaïtasov, O., Hoffmann-Eifert, S., and Schneller, T. (2005) *Phys. Chem. Chem. Phys.*, **7**, 2053–2060.
36. Vollman, M. and Waser, R. (1994) *J. Am. Ceram. Soc.*, **77** (1), 235–243.
37. Waser, R. (1995) *Solid State Ionics*, **75**, 89–99.
38. Denk, I., Claus, J., and Maier, J. (1997) *J. Electrochem. Soc.*, **144** (10), 3526–3536.
39. De Souza, R.A., Fleig, J., Maier, J., Kienzle, O., Zhang, Z.L., Sigle, W., and Rühle, M. (2003) *J. Am. Ceram. Soc.*, **86** (6), 922–928.
40. De Souza, R.A. and Maier, J. (2007) *Faraday Discuss.*, **134**, 235–245.
41. (a) Balaya, P., Jamnik, J., Fleig, J., and Maier, J. (2006) *Appl. Phys. Lett.*, **88**, 062109; (b) Balaya, P., Ahrens, M., Kienle, L., Maier, J., Rahmati, B., Lee, S.B., Sigle, W., Pashkin, A., Kuntscher, C., and Dressel, M. (2006) *J. Am.*

- Ceram. Soc.*, **89** (9), 2804–2811; (c) Balaya, P., Jamnik, J., Fleig, J., and Maier, J. (2007) *J. Electrochem. Soc.*, **154**, P69–P76.
42. Lupetin, P., Gregori, G., and Maier, J. (2010) *Angew. Chem. Int. Ed.*, **49** (52), 10123–10126.
 43. Gregori, G., Lupetin, P., Heinze, S., Habermeier, H.-U., and Maier, J. (2013) *J. Mater. Sci.*, **48** (7), 2790–2796.
 44. Szot, K., Speier, W., Carius, R., Zastrow, U., and Beyer, W. (2002) *Phys. Rev. Lett.*, **88** (7), 075508.
 45. Szot, K., Speier, W., Bihlmayer, G., and Waser, R. (2006) *Nat. Mater.*, **5**, 312–320.
 46. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21**, 2632–2663.
 47. Green, D.J., Hannink, R.H.J., and Swain, M.V. (1989) *Transformation Toughening of Ceramics*, CRC Press, Boca Raton, FL.
 48. Nernst, W. (1899) *Z. Elektrochem.*, **6** (2), 41–43.
 49. Wagner, C. (1943) *Z. Elektrochem.*, **23–24**, 265–268.
 50. Strickler, D.W. and Carlson, W.G. (1963) *J. Am. Ceram. Soc.*, **47** (3), 122–127.
 51. Bauerle, J.E. and Hrizo, H. (1969) *J. Phys. Chem. Solids*, **30**, 565–570.
 52. Kilner, J.A. and Brook, R.J. (1982) *Solid State Ionics*, **6**, 237–252.
 53. Badwal, S.P.S. and Swain, M.V. (1985) *J. Mater. Sci. Lett.*, **4**, 487–489.
 54. Ikeda, S., Sakurai, O., Uematsu, K., Mizutani, N., and Kato, M. (1985) *J. Mater. Sci.*, **20**, 4593–4600.
 55. Sasaki, K. and Maier, J. (2004) *Solid State Ionics*, **134**, 303–321.
 56. Gibson, I.R. and Irvine, J.T.S. (1996) *J. Mater. Chem.*, **6** (5), 895–898.
 57. Shimojo, F., Okabe, T., Tachibana, F., Kobayashi, M., and Okazaki, H. (1992) *J. Phys. Soc. Jpn.*, **61** (8), 2848–2857.
 58. Krishnamurthy, R., Yoon, Y.-G., Srolovitz, D.J., and Car, R. (2004) *J. Am. Ceram. Soc.*, **87** (10), 1821–1830.
 59. Bogicevic, A., Wolverton, C., Crosbie, G.M., and Stechel, E.B. (2001) *Phys. Rev. B*, **64**, 014106.
 60. Marrocchelli, D., Madden, P.A., Norberg, S.T., and Hull, S. (2011) *Chem. Mater.*, **23**, 1365–1373.
 61. Goff, J.P., Hayes, W., Hull, S., Hutchings, M.T., and Clausen, K.N. (1999) *Phys. Rev. B*, **59**, 14202–14219.
 62. Kleitz, M., Fernandez, E., Foultier, J., and Fabry, P. (1981) in *Advances in Ceramics*, vol. 3 (eds A.H. Heuer and L.W. Hobbs), The American Ceramic Society, Columbus, OH, p. 349.
 63. Park, J.-H. and Blumenthal, R.N. (1989) *J. Electrochem. Soc.*, **136** (10), 2867–2876.
 64. Burke, L.D., Rickert, H., and Steiner, R. (1971) *Z. Phys. Chem.*, **74**, 146–167.
 65. Guo, X. and Maier, J. (2001) *J. Electrochem. Soc.*, **148** (3), E121–E126.
 66. Guo, X., Sigle, W., Fleig, J., and Maier, J. (2002) *Solid State Ionics*, **154–155**, 555–561.
 67. Otsuka, K., Matsunaga, K., Nakamura, A., Ji, S., Kuwabara, A., Yamamoto, T., and Ikuhara, Y. (2004) *Mater. Trans.*, **45** (7), 2042–2047.
 68. Houssa, M., Pantisano, L., Ragnarsson, L.-A., Degraeve, R., Schrama, T., Pourtoisa, G., De Gendta, S., Groesenekken, G., and Heyns, M.M. (2006) *Mater. Sci. Eng. R*, **51** (4–6), 37–85.
 69. Baun, W.L. (1963) *Science*, **140**, 1330–1331.
 70. Trubelja, M.F. (1987). PhD thesis, Ionic conductivity in the hafnia-rare earth systems and phase equilibria in the system hafnia-zirconia-yttria Pennsylvania State University, University Park, PA.
 71. Wang, J., Li, H.P., and Stevens, R. (1992) *J. Mater. Sci.*, **27**, 5397–5430.
 72. Johansen, H.A. and Cleary, J.G. (1964) *J. Electrochem. Soc.*, **111** (1), 100–103.
 73. Smith, A.W., Meszaros, F.W., and Amata, C.D. (1966) *J. Am. Ceram. Soc.*, **49** (5), 240–244.
 74. Schieltz, J.D., Patterson, J.W., and Wilder, D.R. (1971) *J. Electrochem. Soc.*, **118** (8), 1257–1261.
 75. Trubelja, M.F. and Stubican, V.S. (1991) *J. Am. Ceram. Soc.*, **74** (10), 2489–2494.
 76. Navrotsky, A., Simoncic, P., Yokokawa, H., Chen, W., and Lee, T. (2007) *Faraday Discuss.*, **134**, 171–180.

77. Dienes, G.J., Welch, D.O., Fischer, C.R., Hatcher, R.D., Lazareth, O., and Samberg, M. (1975) *Phys. Rev. B*, **11**, 3060.
78. Catlow, C.R.A., James, R., Mackrodt, W.C., and Steward, R.F. (1982) *Phys. Rev. B*, **25** (2), 1006–1026.
79. Jacobs, P.W.M. and Kotomin, E.A. (1993) *Philos. Mag. A*, **68** (4), 695–709.
80. Grimes, R.W. (1994) *J. Am. Ceram. Soc.*, **77** (2), 378–384.
81. Lagerlöf, K.P.D. and Grimes, R.W. (1998) *Acta Mater.*, **46** (16), 5689–5700.
82. Atkinson, K.J.W., Grimes, R.W., Levy, M.R., Coull, Z.L., and English, T. (2003) *J. Eur. Ceram. Soc.*, **23**, 3059–3070.
83. Matsunaga, K., Tanaka, T., Yamamoto, T., and Ikuhara, Y. (2003) *Phys. Rev. B*, **68**, 085110.
84. Lagerlöf, K.P.D., Mitchell, T.E., and Heuer, A.H. (1989) *J. Am. Ceram. Soc.*, **72** (11), 2159–2171.
85. Heuer, A.H. and Lagerhöf, K.P.D. (1999) *Philos. Mag. Lett.*, **79** (8), 619–627.
86. Heuer, A.H. (2008) *J. Eur. Ceram. Soc.*, **28**, 1495–1507.
87. Doremus, R.H. (2004) *J. Appl. Phys.*, **95** (6), 3217–3222.
88. Doremus, R.H. (2006) *J. Appl. Phys.*, **100**, 101301.
89. Mohapatra, S.K. and Kröger, F.A. (1977) *J. Am. Ceram. Soc.*, **60** (3–4), 141–148.
90. Mohapatra, S.K. and Kröger, F.A. (1977) *J. Am. Ceram. Soc.*, **60** (9–10), 381–387.
91. Kröger, F.A. (1984) *Solid State Ionics*, **12**, 189–199.
92. Fielitz, P., Borchardt, G., Ganschow, G., Bertram, R., and Markwitz, A. (2008) *Solid State Ionics*, **179** (11–12), 373–379.
93. Lagergren, S. and Magneli, A. (1952) *Acta Chem. Scand.*, **6**, 444–446.
94. Waring, J.L. and Roth, R.S. (1968) *J. Res. Natl. Bur. Stand.*, **72A** (2), 175–183.
95. Wasilewski, R.J. (1953) *J. Am. Chem. Soc.*, **75** (4), 1001–1002.
96. Reisman, A., Holtzberg, F., Berkenblit, M., and Berry, M. (1956) *J. Am. Chem. Soc.*, **78** (18), 4514–4520.
97. Kofstad, P. (1962) *J. Electrochem. Soc.*, **109** (9), 776–781.
98. Stroud, E., Tripp, W.C., and Wimmer, J.M. (1974) *J. Am. Ceram. Soc.*, **57** (4), 172–175.
99. Balachandran, U. and Eror, N.G. (1982) *Mat. Res. Bull.*, **17**, 151–160.
100. Balachandran, U. and Eror, N.G. (1982) *J. Less-Common Met.*, **84**, 291–299.
101. Johannesen, O. and Kofstad, P. (1984) *J. Phys. Chem. Solids*, **45** (3), 239–250.
102. Johannesen, O. and Kofstad, P. (1984) *Solid State Ionics*, **12**, 235–242.
103. Roth, R.S. and Waring, J.L. (1970) *J. Res. Natl. Bur. Stand.*, **74A** (4), 485–493.
104. McHale, A.E. and Tuller, H.L. (1985) *J. Am. Ceram. Soc.*, **68** (12), 646–650.
105. Roth, R.S., Waring, J.L., and Brower, W.S. (1970) *J. Res. Natl. Bur. Stand.*, **74A** (4), 477–484.
106. McHale, A.E. and Tuller, H.L. (1983) *Radiat. Eff.*, **75**, 267–281.
107. Choi, G.M., Tuller, H.L., and Haggerty, J.S. (1989) *J. Electrochem. Soc.*, **136** (3), 835–838.
108. Dellis, J.L., Carpentier, J.L., Picot, J.C., Tellier, P., and Filal, M. (1992) *J. Alloys Compd.*, **189**, 157–161.
109. Choi, G.M. and Tuller, H.L. (1990) *J. Am. Ceram. Soc.*, **73** (6), 1700–1704.
110. Bredesen, R. and Kofstad, P. (1988) *Solid State Ionics*, **27**, 11–18.

5

Ion Transport in Metal Oxides

Roger A. De Souza

5.1

Introduction

The subject of ion transport in metal oxides is both broad and deep, and accordingly, there are many different ways of approaching the subject. This is due in part to the various driving forces that can cause ions to move. For example, ion motion may occur because of a concentration gradient, in which case one speaks of diffusion; or because of an electrical potential gradient, in which case one speaks of drift; or because of a temperature gradient, in which case one speaks of thermodiffusion. In this chapter, we use diffusion in the solid state as a means of examining ion transport in oxides. After the introductory sections covering the basics of diffusion, we consider the differences and the relationships between the most common diffusion coefficients. The possibilities of accelerated diffusion occurring along extended defects and diffusion across extended defects being hindered are also discussed. It is possible that, compared with diffusion in the bulk, diffusion along extended defects is retarded or that diffusion across extended defects is accelerated; these cases are of minor importance and, therefore, are not considered here. The final section provides a literature survey of anion and cation transport in several example systems.

There are two complementary approaches to treating diffusion: the macroscopic approach based on Fick's empirical equations and the microscopic approach based on atomic mechanisms and on random walk theory. The macroscopic definition provides the experimental basis for determining diffusion coefficients. At no point in the macroscopic treatment, however, does the atomic nature of matter enter into the treatment. Nonetheless, diffusion in solids results from many individual jumps of the diffusing particles, and it is mediated by point defects such as vacancies and interstitials. The benefit of the microscopic approach, then, is that it provides the theoretical basis for interpreting and expressing diffusion coefficients in terms of atomic quantities, such as point-defect concentrations, atomic jump distances, and activation barriers.

While writing this chapter, I referred to several books [1–6] and review articles [7–14]. My aim is to present the subject matter so that, in the end, beginners will

be able to understand, to apply, and perhaps even, to judge critically the literature data for ion transport in metal oxides.

5.2

Macroscopic Definition

Let us consider, at the level of a continuum, a single-phase system, in which the chemical component i is distributed inhomogeneously (i.e., its concentration c_i is not the same everywhere in the system). We now bring the system to a temperature at which i is mobile: as a result of diffusion, component i will move in such a manner that its concentration gradient decreases. The flux j_i (i.e., the amount of i passing through unit area of a reference plane per unit time) is given by Fick's first law. In one dimension (x), this law can be written as

$$j_i = -D_i \frac{\partial c_i}{\partial x}. \quad (5.1)$$

The factor of proportionality between the flux and the negative concentration gradient is D_i , the diffusion coefficient of i in the specific system. The minus sign in Eq. (5.1) indicates that the flux is directed toward lower concentrations, that is, "down the concentration gradient." As expected, Eq. (5.1) indicates that the flux tends to zero as $\partial c_i / \partial x$ tends to zero: the process of diffusion leads to the elimination of concentration gradients [1, 2].

In the majority of cases, the amount of diffusing component is conserved during the process of diffusion; this means that component i is neither created nor destroyed. If we consider a small volume element within the continuum phase, the flux of component i flowing into this small volume element minus the flux of i flowing out of it equals the rate of accumulation (or loss) of i within that volume element. This statement is captured mathematically by the continuity equation, which for diffusion in one dimension is

$$-\frac{\partial j_i}{\partial x} = \frac{\partial c_i}{\partial t}. \quad (5.2)$$

Substitution of Eq. (5.1) into Eq. (5.2) yields Fick's second law,

$$\frac{\partial c_i}{\partial t} = \frac{\partial}{\partial x} \left(D_i \frac{\partial c_i}{\partial x} \right), \quad (5.3)$$

which, if the diffusion coefficient is independent of position, reduces to

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial x^2}. \quad (5.4)$$

The second-order partial differential equation, Eq. (5.3) [or Eq. (5.4)], is sometimes called the "diffusion equation."

Although the one-dimensional treatment is often sufficient, Fick's first law is easily generalized to the three-dimensional case:

$$\mathbf{j}_i = -\mathbf{D}_i \nabla c_i \quad (5.5)$$

in which \mathbf{j}_i is the vector flux; the Nabla operator ∇ acts on the scalar concentration field $c_i(x, y, z)$; and \mathbf{D}_i is a symmetric second-rank tensor (in cubic crystals, it reduces to the single scalar quantity D_i). Analogous phenomena, in the sense of cause (driving force) and effect (flux), are Fourier's law of heat transport, $j_Q = -\kappa \nabla T$ (a flux of heat, j_Q , is driven by a gradient in temperature T , with the thermal conductivity κ as the proportionality constant), and Ohm's law, $I = -\sigma \nabla \phi$ (the current density I is driven by a gradient in electrical potential ϕ , with the electrical conductivity σ as the proportionality constant) [2].

It is important to note that, although Eq. (5.1) gives the macroscopic definition of a diffusion coefficient, the true driving force for diffusion is not the gradient in the concentration of i (∇c_i) but the gradient in its chemical potential ($\nabla \mu_i$); see Section 5.4. At chemical equilibrium, the gradient in the chemical potential of all mobile components is zero.

5.2.1

Two Solutions of the Diffusion Equation

From a practical point of view, Fick's first law [Eq. (5.1)] is not particularly useful for diffusion measurements in the solid state because it requires us to determine the diffusion flux. The accurate measurement of fluxes is seldom trivial and often virtually impossible; in addition, solid-state diffusivities are often low, so that the corresponding fluxes are also low. The optimal basis for determining a diffusion coefficient has proven to be the examination of concentration transients. One measures, for example, the concentration profile in a diffusion specimen as a function of position for a given time, and the diffusion coefficient D_i is obtained by describing the concentration profile with an appropriate solution to the diffusion equation [Eq. (4 or 5)] for the given initial and boundary conditions.

One point cannot be emphasized enough: If one can describe an entire concentration profile with the appropriate solution of the diffusion equation, one has incontrovertible evidence that diffusion in the solid state has taken place.

For the sake of illustration, we consider two experimental arrangements that are often used in diffusion studies. In both cases, the experiment is set up so that diffusion takes place in one dimension (x), in a medium that extends to infinity in the positive x direction ($x > 0$, semi-infinite medium) and that contains no diffusant at the start of the experiment, $c_i(x > 0, t = 0) = 0$; furthermore the diffusion coefficient D_i does not depend on position, so that Eq. (5.4) is the differential equation to be solved. The two cases we consider are referred to as the instantaneous source solution and the constant source solution (Figure 5.1). For a comprehensive treatment of the mathematics of diffusion, the reader is referred to the textbook by Crank [15].

In the first case, a small amount of diffusant is deposited at the plane $x = 0$. Its initial distribution is given by

$$c_i(x, 0) = M_i \delta(x), \quad (5.6)$$

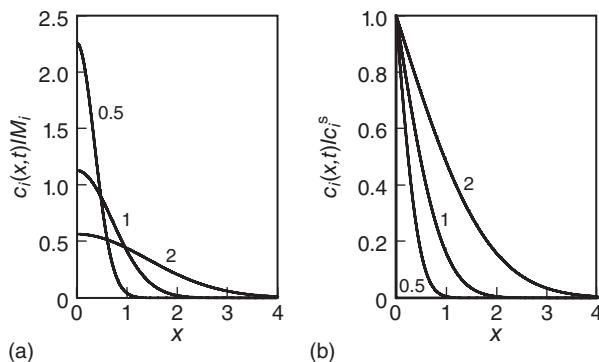


Figure 5.1 Solutions of the diffusion equation for the cases of an instantaneous source (a) and a constant source (b) plotted as normalized concentration against depth x

for various values of $\sqrt{4D_i t}$. The units of x and $\sqrt{4D_i t}$ are arbitrary but identical (if x is expressed in cm, D_i has the unit $\text{cm}^2 \text{s}^{-1}$ and t has the unit s).

M_i denotes the amount of diffusant per unit area and $\delta(x)$ the Dirac delta function. After diffusion for a time t , the concentration profile of i in the sample is described by

$$c_i(x, t) = \frac{M_i}{\sqrt{\pi D_i t}} \exp\left(-\frac{x^2}{4D_i t}\right). \quad (5.7)$$

The quantity $\sqrt{4D_i t}$ is a characteristic diffusion length.

In the second experimental arrangement, the surface of a semi-infinite medium is exposed continuously to a fixed concentration of diffusant (c_i^s), for example, by exposing the surface to an atmosphere of the diffusant. The corresponding solution of the diffusion equation is

$$c_i(x, t) = c_i^s \operatorname{erfc}\left(\frac{x}{2\sqrt{D_i t}}\right) \quad (5.8)$$

where $\operatorname{erfc}(z)$ is the complementary error function: $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$, with $\operatorname{erf}(z)$ being the Gaussian error function (a standard mathematical function), defined as $\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du$.

5.2.2

Dependence of the Diffusion Coefficient on Characteristic Thermodynamic Parameters

Phenomenologically, one often observes that, as a function of temperature T , the diffusion coefficient obeys Arrhenius-type behavior,

$$D_i(T) = D_{i,0} \exp\left(-\frac{\Delta H_{D_i}}{k_B T}\right), \quad (5.9)$$

that is, the behavior is characterized by just two quantities, the pre-exponential factor $D_{i,0}$ and the activation enthalpy of diffusion ΔH_{D_i} . Values of ΔH_{D_i} determined experimentally may be as low as tenths of an electron volt or as high as 10 eV but are generally of the order of 10⁰ eV.

For a binary metal oxide MO , the Gibbs phase rule stipulates that three variables are required to define the system thermodynamically. Two variables are temperature and hydrostatic pressure; the third variable is commonly chosen to be the oxygen partial pressure pO_2 (as it is easier experimentally to define and to vary pO_2 than pM , the partial pressure of the metallic component; furthermore, pO_2 can be varied experimentally over tens of orders of magnitude). Values of the diffusion coefficient measured as a function of pO_2 at constant temperature are found to obey a power law, with the power-law exponents m_{pO_2} ,

$$m_{pO_2} = \left(\frac{\partial \ln D_i}{\partial \ln pO_2} \right)_T. \quad (5.10)$$

Considered over a suitably large range of oxygen partial pressures, m_{pO_2} may take characteristic values of $\pm \frac{1}{6}$, $\pm \frac{1}{4}$, $\pm \frac{1}{2}$, and so on (see Section 4.3.3).

Similarly, for a doped binary oxide, at defined T and pO_2 , one often observes that D_i varies with dopant concentration [Dop] according to a power law, with exponent

$$m_{[\text{Dop}]} = \left(\frac{\partial \ln D_i}{\partial \ln [\text{Dop}]} \right)_{T,pO_2}. \quad (5.11)$$

In order to interpret and understand the activation energy of diffusion or one of the power-law exponents, one must consider diffusion from a microscopic standpoint.

5.3

Microscopic Definition

Fick's first law, as described in Eq. (5.1) or Eq. (5.5), provides the macroscopic definition of the diffusion coefficient D_i . The microscopic definition was derived, independently, by Einstein and by Smoluchowski. They considered that the diffusion arises as a result of the motion of atomic species (atoms, ions, molecules), and they examined the question of how far these species diffuse in a given time. They recognized that the quantity of interest is not the velocity of the diffusing species but their mean square displacement. Specifically, they related the mean square displacement $\langle x^2 \rangle$ of the diffusing species in the x -direction to the x -component of the diffusion coefficient, D_i^x , and to the time interval t during which diffusion takes place [1, 2, 5]:

$$\langle x_i^2 \rangle = 2D_i^x t. \quad (5.12)$$

This is now known as the Einstein relation or as the Einstein–Smoluchowski relation. For diffusion in an isotropic, three-dimensional medium, $\langle x_i^2 \rangle = \langle y_i^2 \rangle =$

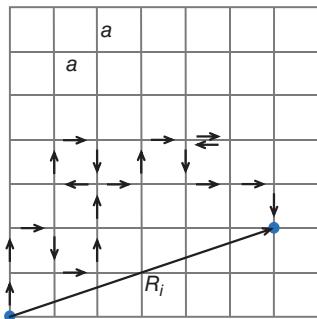


Figure 5.2 A single random walk of a single particle i on an empty two-dimensional lattice, with intersite jump distance of a . After $n = 20$ jumps, the particle has achieved a displacement R_i . Repeating the random walk of $n = 20$ jumps many times yields $\langle R_i \rangle = 0$ but $\langle R_i^2 \rangle = 20a^2$. Note: for each random walk, the particle covers a distance of $20a$; after many random walks, however, its root-mean-square displacement is $\sqrt{20}a$. For this two-dimensional case, the diffusion coefficient follows as $D_i = \frac{1}{4}a^2Z\Gamma_i$, with $Z = 4$. Adapted from Ref. [2].

$\langle z_i^2 \rangle = \langle R_i^2 \rangle / 3$, the mean square displacement is given by

$$\langle R_i^2 \rangle = 6D_i t. \quad (5.13)$$

At the microscopic scale, diffusion is considered to occur as the net result of many individual atomic jumps. Let us consider, then, a single particle (atom, ion, or molecule) diffusing on an otherwise empty three-dimensional lattice. The particle executes a series of consecutive jumps, each of distance a , from one site to a neighboring site (a in oxides is thus of the order of a few Å). Having made n jumps, the particle is characterized (Figure 5.2) by a displacement R from its starting position and a squared displacement R^2 . If each jump is independent of the previous jumps, that is, the particle has no memory of the past jumps made, the particle is said to make an uncorrelated random walk. Let us now repeat the walk of n jumps a large number of times and consider the average quantities. (As an alternative to averaging repeated trials of an isolated particle, one can average over a large set of dilute, noninteracting particles.) The average displacement $\langle R_i \rangle$ in this case is zero, because on average, the number of jumps in $+x$, $-x$, and so on will be the same. The mean square displacement $\langle R_i^2 \rangle$, however, is not zero. From the theory of random walks, one obtains for an uncorrelated random walker

$$\langle R_i^2 \rangle = na^2. \quad (5.14)$$

By combining Eqs (5.13) and (5.14) and introducing the jump rate of the particle to one of its Z neighbors, $\Gamma_i Z = n/t$, one obtains

$$D_i = \frac{1}{6}a^2Z\Gamma_i. \quad (5.15)$$

That is, we have expressed the diffusion coefficient in terms of certain atomic quantities: the particle's jump distance to a neighboring site a , the number of its neighboring sites Z , and its jump rate Γ_i [2].

5.3.1

Mechanisms of Diffusion

A variety of mechanisms have been proposed over the last century to explain atomic motion in crystalline solids. Nowadays, there is ample evidence that

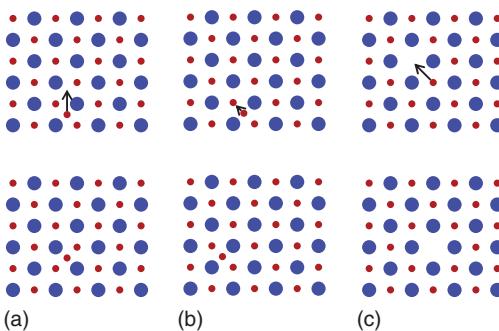


Figure 5.3 Common mechanisms of atomic migration in a binary compound with mobile species (red) and immobile species (blue): (a) interstitial, (b) interstitialcy, and (c) vacancy.

diffusion in crystalline solids takes place by defect-mediated mechanisms. Three common mechanisms are depicted in Figure 5.3.

In the interstitial mechanism (Figure 5.3a), sometimes called the direct interstitial mechanism, an interstitial ion executes a jump from one interstice to another. In the indirect interstitial mechanism (Figure 5.3b), known more commonly as the interstitialcy mechanism, an ion on an interstitial lattice site pushes an ion on a regular lattice site onto an interstitial site and occupies the regular lattice site. This process can occur with the three species moving in a direct line (collinear) or at an angle to one another (noncollinear). The third mechanism, and probably the most important, is the vacancy mechanism (Figure 5.3c): an ion occupying a regular lattice site jumps to a neighboring unoccupied site, that is, the migrating ion and the vacancy exchange sites.

For the sake of completeness, we also mention two mechanisms that were believed for a long time to be operative in crystalline solids but are now best regarded as hypothetical possibilities. The direct exchange of neighboring ions, in which two ions move simultaneously, requires large distortions of the lattice as the migrating species squeeze past each other; this makes this process energetically improbable for ionic solids. The ring mechanism, corresponding to the rotation of three (or more) ions as a group, requires less lattice distortion, but the collective nature of this complex mechanism makes it unlikely for most crystalline substances.

5.3.2

Diffusion Coefficients of Defects and Ions

In general, there will be a considerable difference between the diffusion coefficient of the defect, D_{def} , and the diffusion of the ion, D_{ion} . To see why this is so, we consider the concrete example of oxygen ions migrating in an oxide by a vacancy mechanism. The diffusion coefficient of the vacancies is [see Eq. (5.15)]

$$D_V = \frac{1}{6} a^2 Z \Gamma_V, \quad (5.16)$$

whereas the diffusion coefficient of the oxygen ions is

$$D_O = \frac{1}{6} a^2 Z \Gamma_O. \quad (5.17)$$

Each time a vacancy moves, an ion has to move, since the two species swap places. Thus, the total number of displacements of the ions and of the vacancies have to be equal

$$\Gamma_O[O] = \Gamma_V[V], \quad (5.18)$$

with $[i]$ denoting the concentration of i . For a dilute solution of vacancies, the site fraction of vacancies, n_V , is by definition many orders of magnitude less than unity, and it can thus be approximated by $n_V \approx [V]/[O]$. Hence, by combining Eqs (5.16)–(5.18), we obtain

$$D_O = D_V \frac{[V]}{[O]} \approx D_V n_V. \quad (5.19)$$

Thus, D_O and D_V will differ by the factor $n_V \ll 1$. One should note that n_V may vary considerably with temperature, oxygen partial pressure, and dopant concentration (see Chapter 4). D_V , at least for dilute solutions of vacancies, is independent of n_V and, as we will see later, only dependent on temperature. One consequence of Eq. (5.19) is that the mean square displacement of a vacancy, $\langle R_V^2 \rangle$, will be orders of magnitude larger than the mean square displacement of an oxygen ion, $\langle R_O^2 \rangle$.

An expression similar to Eq. (5.18) also holds for the diffusion of oxygen interstitials (I) in an oxide. Consequently, in an oxide with anti-Frenkel disorder, not only the oxygen vacancies but also the oxygen interstitials may contribute to the overall diffusion coefficient of the ions,

$$D_O[O] = D_V[V] + D_I[I]. \quad (5.20)$$

Equation (5.19) can also be derived by noting that the probability of an oxygen-ion jump is equal to the product of the vacancy hopping rate, Γ_V , and the probability of finding a vacancy adjacent to the ion, n_V . Hence, in the case of impurity diffusion by an interstitial mechanism (i.e., the impurity interstitials remain on the interstitial sublattice and do not undergo exchange with regular lattice species), $D_{\text{ion}} = D_{\text{def}}$ because the probability of finding an vacant interstitial site is essentially unity (for a dilute solution).

5.3.3

The Activation Barrier for Migration

In jumping from one site to another, an ion has to overcome an activation barrier, regardless of which of these three migration mechanisms is operative. Most of the time, the ion vibrates around its equilibrium position (which may be a regular lattice site or an interstitial position). Occasionally, it successfully executes a jump to a new site. Thereafter, it vibrates around its new position, waiting until it successfully makes the next jump. (As the jumps are random, the next successful

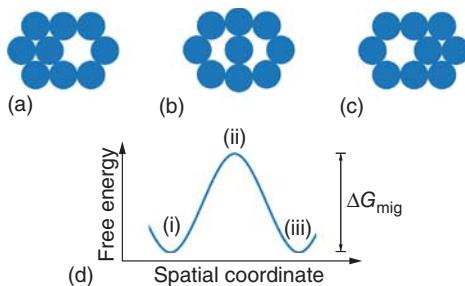


Figure 5.4 Simple schematic illustration of the ion movements involved in the jump of an ion to a neighboring vacant site (migration mediated by a vacancy mechanism). (a) Initial configuration; (b) Saddle-point

configuration; (c) Final configuration. (d) The change in the system's Gibbs energy associated with the ion movements in (a)–(c). ΔG_{mig} is the Gibbs activation energy of migration.

jump may bring the ion back to its original site.) In other words, the duration of a jump is short compared with the residence time of the ion on its site [2].

In Figure 5.4a–c, we see a simple schematic illustration of the migration process for the case of a vacancy mechanism. What is required for the ion to make a successful jump? In this simple picture, not only does the migrating ion have to move in the right direction, the surrounding lattice ions have to move out of the way. Only if both occur, will the ion complete its jump. Figure 5.4(d) shows the associated change in the Gibbs energy of the system as the migrating ion is moved reversibly from (i) to (iii) [1]; ΔG_{mig} , the Gibbs activation energy of migration, corresponds to the difference in the system's Gibbs energy between the initial and saddle-point configurations. The jump rate Γ is related to ΔG_{mig} through the Arrhenius law,

$$\Gamma = \nu_0 \exp\left(\frac{-\Delta G_{\text{mig}}}{k_B T}\right), \quad (5.21)$$

where the prefactor ν_0 denotes an attempt frequency of the order of the Debye frequency of the compound, k_B is Boltzmann's constant, and T is the absolute temperature. This is a simple description of a complicated many-body problem. For the detailed consideration of defect migration in solids including many-body effects, the interested reader is referred to Vineyard [16]. ΔG_{mig} , being a Gibbs energy, can be expressed in terms of an activation enthalpy of migration, ΔH_{mig} , and an activation entropy of migration, ΔS_{mig} ,

$$\Delta G_{\text{mig}} = \Delta H_{\text{mig}} - T \Delta S_{\text{mig}}. \quad (5.22)$$

By combining Eqs (5.19), (5.21), and (5.22), we thus find, for the case of oxygen-ion diffusion mediated by a vacancy mechanism in a cubic lattice, that the self-diffusion coefficient of oxygen is given by

$$D_O = n_V(T, pO_2) \frac{1}{6} a^2 Z \nu_0 \exp\left(\frac{\Delta S_{\text{mig},V}}{k_B}\right) \exp\left(\frac{-\Delta H_{\text{mig},V}}{k_B T}\right). \quad (5.23)$$

5.4

Types of Diffusion Experiments

There are a bewildering variety of ways to carry out diffusion experiments and, thus, a bewildering variety of diffusion coefficients. For metal oxides, the three most commonly determined coefficients are D^* , the tracer diffusion coefficient; D^δ , the chemical diffusion coefficient; and D^σ , the conductivity diffusion coefficient [6]. In the following we examine, first, what the basics of the three diffusion experiments are and, second, how the respective diffusion coefficients are related to the self-diffusion coefficients of the ions D_{ion} or of the defects D_{def} .

In order to obtain expressions relating the diffusion coefficients to one another, we require a theoretical framework to treat these processes, and the most convenient theoretical framework for treating these and other transport processes is linear irreversible thermodynamics [17]. For example, for a system that contains the mobile charge carriers (ions, electrons) i and k and that is exposed to a temperature gradient, one can express, according to linear irreversible thermodynamics, the vector fluxes of the two charge carriers and the vector flux of heat as

$$\mathbf{j}_i = L_{ii}\mathbf{X}_i + L_{ik}\mathbf{X}_k + L_{iQ}\mathbf{X}_Q \quad (5.24a)$$

$$\mathbf{j}_k = L_{ki}\mathbf{X}_i + L_{kk}\mathbf{X}_k + L_{kQ}\mathbf{X}_Q \quad (5.24b)$$

$$\mathbf{j}_Q = L_{Qi}\mathbf{X}_i + L_{Qk}\mathbf{X}_k + L_{QQ}\mathbf{X}_Q \quad (5.24c)$$

that is, in terms of phenomenological transport coefficients L and general thermodynamic driving forces, \mathbf{X} . Equation (5.24) states that a flux of i , say, can result directly from the driving force \mathbf{X}_i , but also indirectly from the driving forces \mathbf{X}_k and \mathbf{X}_Q . The thermodynamic driving force for heat transport is given by $\mathbf{X}_Q = -(1/T)\nabla T$; for the charge carrier i , the thermodynamic driving force is given, in the absence of external forces, by $\mathbf{X}_i = -T\nabla(\eta_i/T)$, where η_i denotes the electrochemical potential of i and is defined as (z_i is the charge number of i)

$$\eta_i = \mu_i + z_i e\phi. \quad (5.25)$$

The off-diagonal elements of the matrix of L coefficients lead to “cross” effects such as thermoelectricity (L_{eQ}) and thermodiffusion (L_{iQ}) and are thus also known as cross-coefficients. Onsager’s reciprocity theorem states that, in the absence of a magnetic field, the matrix of L coefficients is symmetric, that is, $L_{ik} = L_{ki}$, and so on. It is noted that an oxide was used as a model system to verify Onsager’s theorem experimentally [18].

Let us now, for the sake of simplicity, restrict the treatment to the one-dimensional case and to ideal solutions. By transforming Eq. (5.24) to the “reduced” heat formulation [7, 13, 14, 19]; assuming $L_{ik} = L_{ki}$ and so on; and using $L_{ii} = D_i[i]/k_B T$; one can write the flux of i as

$$j_i = -\frac{D_i[i]}{k_B T} \left[\frac{\partial \eta_i}{\partial x} + \left(\bar{S}_i + \frac{Q_i}{T} \right) \frac{\partial T}{\partial x} \right]. \quad (5.26)$$

where \bar{S}_i is the partial molar entropy and Q_i the reduced heat of transport of charge carrier i . Thus, we perceive that a flux of i can be driven by a gradient in chemical

potential (diffusion), by a gradient in electrical potential (drift), and by a gradient in temperature (thermodiffusion). The term $(\bar{S}_i + Q_i/T)$ may be positive or negative, and thus depending on the sign, the ion i may move down or up the temperature gradient. In the isothermal case, to which the three diffusion experiments refer, Eq. (5.26) reduces to

$$j_i = -\frac{D_i[i]}{k_B T} \left[\frac{\partial \mu_i}{\partial x} + z_i e \frac{\partial \phi}{\partial x} \right]. \quad (5.27)$$

The procedures for deriving the various diffusion coefficients from Eq. (5.24) or Eq. (5.27) are given in detail elsewhere [7, 8, 11]; due to limited space, only the end results are reproduced in the relevant sections. Furthermore, only the simplest cases are considered: many complications are possible [6, 11, 14, 20].

5.4.1

Chemical Diffusion

Consider an oxide of composition $MO_{1-\delta}$ in chemical equilibrium with the surrounding gas phase; that is, chemical potential of oxygen throughout the oxide is equal to that in the gas phase. A difference in the chemical potential of oxygen between the sample and the surrounding gas phase is now induced (either by changing the system's temperature or the oxygen activity of the gas phase), with the result that oxygen is either incorporated or removed from the oxide. The non-stoichiometry changes from its original value δ to a final value $\delta + \Delta\delta$. The process whereby this takes place ($MO_{1-\delta} + \Delta\delta \cdot (1/2)O_2 \rightarrow MO_{1-\delta+\Delta\delta}$) is called chemical diffusion (Figure 5.5). The chemical diffusion coefficient, according to Fick's first law [Eq. (5.1)], is denoted as D_i^δ , \tilde{D}_i , or $D_{i,chem}$: the true driving force, however, is the gradient in the chemical potential of oxygen [11].

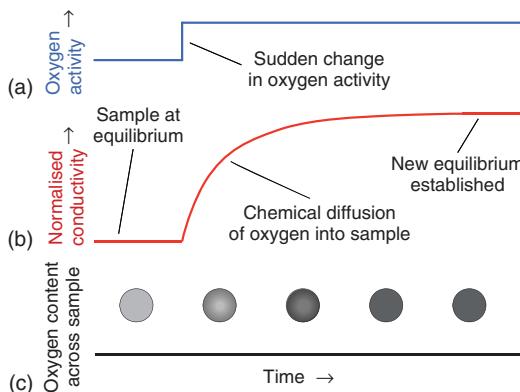


Figure 5.5 The chemical diffusion experiment consists of, for example, instantaneously increasing the activity of oxygen in the surrounding gas phase (a) and then monitoring the change in a characteristic

sample property (b) as a function of time, that is, one that depends on the oxygen content of the sample (c), such as the electrical conductivity, as the sample attains the new equilibrium with the gas phase.

Although, formally, oxygen is incorporated into the oxide as a neutral component (“O”), at the microscopic scale, there is coupled transport of charged species, otherwise known as ambipolar diffusion. Let us assume that the oxide $MO_{1-\delta}$ has oxygen vacancies and electrons as the dominant mobile defects. As oxygen is incorporated, there will be a flux of vacancies j_V and a flux of electrons j_e toward the surface; these two fluxes are given by [see Eq. (5.27)]

$$j_V = -\frac{D_V[V]}{k_B T} \left[\frac{\partial \mu_V}{\partial x} + 2e \frac{\partial \phi}{\partial x} \right] \quad (5.28a)$$

$$j_e = -\frac{D_e[e]}{k_B T} \left[\frac{\partial \mu_e}{\partial x} - e \frac{\partial \phi}{\partial x} \right] \quad (5.28b)$$

These two fluxes are not independent, however, as $2j_V - j_e = 0$. Furthermore, the internal gradient in the electrical potential, $-\partial\phi/\partial x$ (also known as the Nernst field), is common to both fluxes; this coupling has the effect of accelerating the slower moving moiety and slowing down the faster moving one. The detailed analysis yields

$$D_O^\delta = \frac{D_V[V] D_e[e]}{4D_V[V] + D_e[e]} \left(\frac{1}{2} \frac{\partial \ln pO_2}{\partial c_O} \right). \quad (5.29)$$

It is clear from Eq. (5.29) that the chemical diffusion coefficient of oxygen in $MO_{1-\delta}$, D_O^δ , may exhibit a rather complex behavior as a function of temperature and oxygen partial pressure, depending on how the individual parameters vary with T and pO_2 .

In some cases, Eq. (5.29) reduces to a simpler form: if, for instance, the two defects are dilute, noninteracting species, we find

$$D_O^\delta = \frac{D_V[V] D_e[e]}{4D_V[V] + D_e[e]} \left(\frac{1}{[V]} + \frac{4}{[e]} \right). \quad (5.30)$$

Furthermore, if the electrons are more mobile than the vacancies, $D_e \gg D_V$, and if electroneutrality is given by $[e] = 2[V]$, we obtain

$$D_O^\delta = 3D_V. \quad (5.31)$$

That is, D_O^δ is given by the diffusivity of the slowest moving defect (in this case, vacancies) multiplied by an “acceleration factor” (=3) because of the coupling through the Nernst field with the faster moving electrons. In this special case, the activation enthalpy of chemical diffusion is the activation enthalpy of vacancy migration: $\Delta H_{D_O^\delta} = \Delta H_{\text{mig}, V_O}$.

With the knowledge of D_O^δ , one can predict the time τ necessary for a sample’s nonstoichiometry to proceed, say, to 99% completion. A useful order-of-magnitude approximation for a slab sample of thickness $2l$ is $\tau_{D^\delta} \sim l^2/2D_O^\delta$. In certain cases, for instance, for thin samples, the kinetics of the stoichiometry change are governed by the surface reaction (see Section 5.5), and the time required is $\tau_{k^\delta} \sim 5l/2k_O^\delta$. For the intermediate regime, where both bulk diffusion and surface kinetics are important, $\tau_{D^\delta} \sim \tau_{k^\delta}$, a good approximation is $\tau \approx \tau_{D^\delta} + \tau_{k^\delta}$.

5.4.2

Tracer Diffusion

A tracer diffusion experiment refers to the use of radioactive or stable isotopes – chemically identical, labeled species i^* – to examine self-diffusion in condensed matter. Since one can distinguish between i^* and i , the motion of the indistinguishable i particles can be followed with the help of the tracers, i^* [11]. Tracer diffusion experiments are performed at constant sample composition. For the example oxide $MO_{1-\delta}$, this corresponds in the case of the cation tracer to $MO_{1-\delta} \rightarrow (M_{1-\alpha}M_\alpha^*)O_{1-\delta}$ and in the case of the anion tracer to $MO_{1-\delta} \rightarrow M(O_{1-\alpha}O_\alpha^*)_{1-\delta}$. The only driving force is the gradient in the chemical potential of the tracer (and is thus purely entropic); the chemical potentials of M and O are constant throughout the system.

For many metallic species, convenient radioisotopes are available; radioactive oxygen isotopes, however, are impracticable for diffusion measurements, as their half-lives are at most of the order of minutes. Consequently, it is the stable isotope ^{18}O that is used in tracer studies; it can be introduced into a sample either by diffusion annealing in a large volume of ^{18}O -enriched gas at an elevated temperature or by depositing a thin layer of $M^{18}\text{O}$ at room temperature and then diffusion annealing at an elevated temperature [12]. The former variant is illustrated in Figure 5.6.

The measured tracer diffusion coefficient of species i is related to the self-diffusion coefficient through the tracer correlation coefficient f^* ,

$$D_i^* = f^* D_i \quad (5.32)$$

f^* varies according to the migration mechanism (interstitial, vacancy, collinear/noncollinear interstitialcy) and the geometry of the sublattice on which diffusion takes place (see Table 5.1). f^* reflects the fact that tracer species do

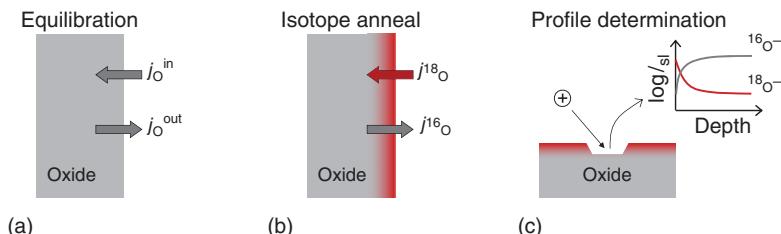


Figure 5.6 The tracer diffusion experiment [12]: (a) An oxide sample is subjected to pre-annealing in oxygen of normal isotopic abundance at given temperature and oxygen activity in order to equilibrate the sample with the surrounding atmosphere. (b) Subsequently, it is annealed, at the same temperature and oxygen activity, in an ^{18}O -enriched gas for a given time. At the sample surface, the dynamic equilibrium between gaseous oxygen and oxygen in the sample

leads to the incorporation of ^{18}O and the removal of ^{16}O (no net incorporation or removal of oxygen, only the exchange of one isotope into another). Subsequent diffusion of ^{18}O away from the interface and into the solid produces an oxygen isotope profile. (c) The isotope profile in the oxide is commonly determined by an ion beam analysis method, such as secondary ion mass spectrometry (SIMS).

Table 5.1 Tracer correlation coefficients for diffusion on various lattices by various mechanisms.

Lattice	Mechanism	*
Diamond	Vacancy	0.5
Simple cubic	Vacancy	0.6531
bcc cubic	Vacancy	0.7272
fcc cubic	Vacancy	0.7815
O in ABO_3 perovskite	Vacancy	0.69
Any lattice	Interstitial	1
Diamond	Interstitialcy (colinear)	0.727

not necessarily perform an uncorrelated random walk. Let us consider the case of vacancy migration on a simple cubic sublattice. The vacancies may move in all six migration directions with equal probability. This is not true, however, for the tracer species: if a vacancy and a tracer have just exchanged places, the most likely jump of the tracer is back to its original position. Its mean square displacement $\langle (R_i)^2 \rangle$ is hence less than that of a random walker, $\langle R_i^2 \rangle$; the tracer correlation coefficient is the ratio of the two, $f^* = \langle (R_i)^2 \rangle / \langle R_i^2 \rangle$, and thus takes values between zero and unity. The detailed analysis with linear irreversible thermodynamics begins with Eq. (5.24); considers three fluxes, for example, for a vacancy mechanism j_i , j_{i^*} , and j_V ; and yields f^* in terms of L_{ii} and L_{ii^*} [7, 8]. f^* deviates from unity if L_{ii^*} deviates from zero.

The activation enthalpy of tracer diffusion, $\Delta H_{D_i^*}$, is easily determined by performing measurements as a function of temperature (at constant oxygen partial pressure), but it is not so easy to interpret. For the case of vacancy transport, we find upon combining Eqs (5.19) and (5.32),

$$D_i^* = f^* D_V n_V. \quad (5.33)$$

Since the dependence of D_V and n_V on temperature can be expressed in exponential functions,

$$D_i^* = f^* D_{V,0} \exp\left(-\frac{\Delta H_{\text{mig},V}}{k_B T}\right) n_{V,0} \exp\left(-\frac{\Delta H_{\text{gen},V}}{k_B T}\right), \quad (5.34)$$

with $\Delta H_{\text{mig},V}$ being the activation enthalpy of vacancy migration and $\Delta H_{\text{gen},V}$, reflecting the change in vacancy concentration with temperature, we find

$$\Delta H_{D_i^*} = \Delta H_{\text{mig},V} + \Delta H_{\text{gen},V}. \quad (5.35)$$

Consequently, interpretation of the measured activation enthalpy of tracer diffusion requires quantitative knowledge of the defect chemistry (how exactly does the relevant defect concentration vary with temperature?). $\Delta H_{\text{gen},V}$, it should be noted, can be positive ([V] increasing with increasing temperature), negative ([V] decreasing with increasing temperature), or zero ([V] independent of temperature); it can take values up to several eV, and hence, depending on

the particular case, it may be much larger than, or much smaller than, or even comparable to, $\Delta H_{\text{mig},V}$.

5.4.3

Conductivity

In a sense, determining the electrical conductivity of a sample is the simplest of the three transport experiments. One applies an electrical field $E = -\nabla\phi$ to a sample and measures the resulting electrical current density I . The electrical conductivity σ is determined from Ohm's law

$$I = -\sigma \nabla\phi. \quad (5.36)$$

There are, however, several possible complications to this simple experiment. First, and most important, is that all mobile charged species contribute to the measured conductivity, all ionic and all electronic charge carriers, each conductivity contribution being the product of the concentration $[i]$, charge $z_i e$, and mobility u_i

$$\sigma_{\text{tot}} = \sum_i \sigma_i = \sum_i [i] z_i e u_i. \quad (5.37)$$

Generally, the mobilities of electronic charge carriers are orders of magnitude larger than those of ionic charge carriers; hence, a small concentration of electrons or holes (minority defects) may predominantly contribute to the electrical conductivity. To isolate the ionic contribution, one may have to use the appropriate electron-blocking but ion-conducting electrodes, or else determine, in addition to the total conductivity, the ionic transference number t_{ion} , that is, the proportion of the conductivity carried by the ions, $t_{\text{ion}} = \sigma_{\text{ion}}/\sigma_{\text{tot}}$ (see Chapter 9.5). The second complication is that the sample's electrical response may be dominated by that of the electrodes, for instance, and obtaining the bulk contribution requires either the use of four-point measurement geometries in DC mode or frequency-dependent impedance spectroscopy studies. Third, it is only for small driving forces (see below) that Eq. (5.36) constitutes a linear law, that is, the current density I is proportional to the driving force $-\nabla\phi$, with the constant of proportionality, the measured conductivity, being independent of driving force.

Having determined the ionic conductivity of the bulk phase in the linear regime, one can now calculate the conductivity diffusion coefficient with the aid of the Nernst–Einstein equation,

$$D_i^\sigma = \sigma_i \frac{k_B T}{[i](z_i e)^2}. \quad (5.38)$$

There are various routes to derive Eq. (5.38), but all of them, it is emphasized, assume the charge carriers under consideration to be noninteracting and dilute [3, 9]. From linear irreversible thermodynamics, for example, one considers a sample of uniform composition ($\nabla\mu_i = 0$) and temperature ($\nabla T = 0$), to which a gradient in the electrical potential $\nabla\phi$ (as the sole thermodynamic driving force) is applied. Comparison of Eq. (5.27), which is only valid for dilute, noninteracting charge

carriers, with Eq. (5.36) yields Eq. (5.38). It is to be noted that, although there are various diffusion coefficients, with $D_{\text{ion}} \neq D_{\text{def}}$ in general, there is only one measured conductivity, that is, $\sigma_{\text{ion}} \equiv \sigma_{\text{def}}$. Thus, from the measured conductivity, one can calculate D_{ion} if [ion] is known, and D_{def} , if [def] is known. It is mentioned that the ratio of D_i^* and D_i^* (two diffusion coefficients that can be independently measured) is called the Haven ratio [9]

$$\frac{D_i^*}{D_i^{\sigma}} = H_R, \quad (5.39)$$

and that for dilute, noninteracting defects, $H_R = f^*$.

Finally, we return to the topic of ion migration under high fields. Let us consider the migration of a positively charged interstitial ion in one dimension. In the absence of an applied field, see Figure 5.7a, the ion will jump to vacant sites in the forward-to-backward directions with equal probability, $v \exp[-\Delta G_{\text{mig}}/k_B T]$. The field alters these probabilities by altering the barriers, as indicated in Figure 5.7b: it increases the probability of motion in the direction of the field to $v \exp[-(\Delta G_{\text{mig}} - (1/2)a z_i E)/k_B T]$ and decreases the probability of motion in the opposite direction in an analogous manner. The net motion in the direction of the field is proportional to the difference in the forward and backward rates. The detailed treatment yields the current density I [21, 22]

$$I = z_i e[i]av \exp\left(-\frac{\Delta G_{\text{mig}}}{k_B T}\right) 2 \sinh\left(\frac{z_i e a E}{2 k_B T}\right). \quad (5.40)$$

For small fields, specifically for $|az_i e E| \ll 2k_B T$, Eq. (5.40) reduces to the linear law of Eq. (5.36) with a field-independent conductivity. For large fields, on the other hand, the jumps of increased probability dominate (e.g., for positive ions in the direction of the field), and the current density then displays an exponential variation with the field. In other words, the measured conductivity depends strongly on the field, $I/E = \sigma(E)$, and the field-induced enhancement of the conductivity over the low-field value can be orders of magnitude. The

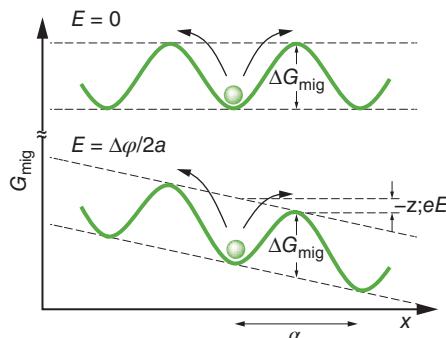


Figure 5.7 Schematic illustration of a positive interstitial ion overcoming an activation barrier of migration ΔG_{mig} : (a) Gibbs energy profile at zero applied field. (b) Gibbs energy profile at nonzero applied field.

fields required to achieve significant effects, however, are large: to increase the oxygen-ion conductivity of an oxide (with a typical oxygen-ion jump distance of $a = 0.3 \text{ nm}$) at room temperature by 25%, one requires a field of $E \approx 1 \text{ MV/cm}$. (A field of $E \approx 2 \text{ MV/cm}$ will yield an increase in conductivity of 220%.) Such fields are close to the fields at which dielectric breakdown occurs in macroscopic samples; at nanoscale distances, though, this limit may be shifted to even higher values. Nevertheless, it is worth noting that ionic transport can be accelerated exponentially by temperature and/or electric field.

5.5

Mass Transport along and across Extended Defects

Real oxide samples, one should recognize, are not single crystals containing only point defects: extended defects, such as dislocations and grain boundaries, will, in general, also be present. In addition, real samples are finite in extent and, thus, are bounded by surfaces or interfaces with other phases. A variety of paths may therefore be available for diffusing species (see Figure 5.8); one differentiates between: (i) bulk diffusion (also termed volume or lattice diffusion), which refers to mass transport within a single grain; (ii) grain-boundary diffusion, which refers to mass transport along the region of crystallographic misorientation between two grains; (iii) dislocation or “pipe” diffusion, which refers to mass transport along dislocations; and (iv) surface diffusion, which refers to mass transport along a crystal surface.

For a given material, the diffusion coefficient of component i along a grain boundary, dislocation, or surface (D_i^{gb} , D_i^{dis} , or D_i^{s} , respectively) will vary according to the structural characteristics of the extended defect. In the case of planar defects (grain boundaries and surfaces), the diffusion coefficient will be a function of the interface (mis)orientation; D_i^{gb} , for example, will vary with the parameters that characterize the grain boundary: the tilt and twist axes, the tilt and twist angles, and the interface plane. Measurements on polycrystals will thus provide

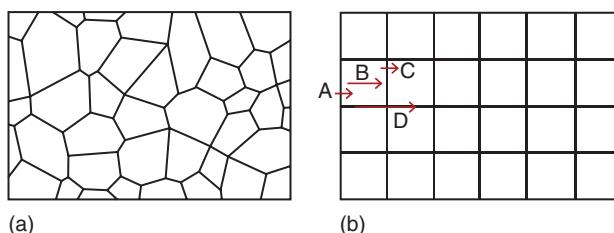


Figure 5.8 Mass-transport processes in a polycrystalline solid. (a) Cross section of a polycrystalline solid. (b) The brick-layer model, an idealized representation of the microstructure shown in (a). The arrows indicate possible transport processes, with the length of

the arrows being inversely proportional to the resistance of the associated process. A – hindered transport across a surface; B – transport in the grain bulk; C – hindered transport across a grain boundary; D – enhanced transport along a grain boundary.

an average over all the grain boundaries contained within the investigated volume. In the case of line defects, D_i^{dis} will vary according to the dislocations' character (edge/screw) and its Burgers vector. These alternative paths offered by extended defects become important, if diffusion in the bulk is slow. In such cases, this fast-path diffusion, or short-circuit diffusion, may contribute significantly to mass transport or may even govern the overall behavior.

Why should diffusion along these extended defects occur faster than in the bulk? The atomic arrangements within grain boundaries and within dislocation cores are considered to be more open than in the bulk phase, suggesting less hindrance for the migrating species and thus accelerated rates of mass transport. Although there is much experimental data that confirms this picture for metallic systems [23–25], it is far from certain that this aspect is also universally applicable to oxides. In an (ionic) oxide, an ion diffusing along an extended defect may have to pass, in a migration jump, ions of the same polarity – a process that is avoided in the bulk phase. Furthermore, the intrinsic structure of the extended defects may offer preferential sites for point defects: as a result, there may be a high concentration of defects within the extended defects, but they are locked into the structure and thus essentially immobile [26]. A more open arrangement within the extended defect does not, therefore, guarantee accelerated rates of ion transport. In addition, the extended defects may be electrostatically charged, with global charge neutrality being satisfied by attendant, enveloping tubes of space charge in which the concentrations of mobile, charged point-defects are modified drastically from their values in the electroneutral bulk [27].

In addition to fast-path diffusion processes that take place in parallel to diffusion in the bulk phase, there may be slower processes in series with bulk diffusion (see Figure 5.8). Hindered mass transport across interfaces (surfaces, grain boundaries) constitute such slower serial processes, and they may also influence (or even govern) the overall diffusion behavior. Here, because the concentration is not a continuous function across the interface, one cannot consider a gradient in concentration, and thus, one cannot define a diffusion coefficient, as in Eq. (5.1). Instead, one describes the flux across the interface in terms of a transfer coefficient k_i and the drop in concentration across the interface, Δc_i , in the direction of the flux:

$$j_i = k_i \Delta c_i. \quad (5.41)$$

The case of transport being limited by grain boundaries is illustrated schematically in Figure 5.9.

Why may transport across an interface be hindered? What does k_i refer to, on a microscopic level? There are several possible causes for k_i taking a finite value. At a grain boundary, for example, the crystallographic mismatch between the two grains may conceivably result in the matter flux being diminished, either because of the considerable perturbations of the bulk structure at the interface or because of differences in the orientations of the grains in layered structures. The magnitudes of such effects are probably small, though. In contrast, huge effects are observed, where space-charge layers, depleted of mobile charge

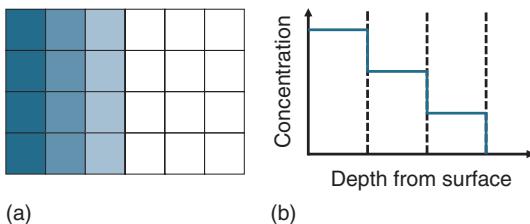


Figure 5.9 Diffusion in a polycrystalline sample with grain size w wherein transport across grain boundaries is slow ($k_i^{\text{gb}} \ll D_i/w$) and diffusion along grain boundaries is negligible. (a) Cross section of a polycrystal: diffusant enters the sample from

the left. (b) Concentration profiles across the polycrystal, showing the decrease in concentration (Δc_i) at the grain boundaries (whose positions are indicated by the dashed, vertical lines).

carriers, are present at the grain boundaries [6]. Huge effects are also observed in polycrystalline samples, in which a second phase covers each grain; such second phases are often SiO_2 -based compounds that exhibit much lower rates of oxygen transport [28].

In the case of oxygen transport across a surface, that is, across a gas|solid interface, the transfer coefficient k_{O}^s characterizes the exchange flux of the dynamic equilibrium between oxygen in the gas phase and oxygen in the solid. The forward reaction, for example, requires, in addition to several charge transfer steps, the adsorption and dissociation of oxygen molecules on the surface and the incorporation of the resulting oxygen moiety into the crystal lattice. The most likely rate-determining step is either dissociation or charge transfer, leading to dissociation [29].

Thus, in a polycrystal, there is a network of serial and parallel mass-transport processes that may be operative. In the following sections, mathematical models for describing (a) fast-path diffusion along extended defects, such as grain boundaries and dislocations, and (b) hindered mass transport across grain boundaries and surfaces are presented.

5.5.1

Accelerated Transport

Extended Defects

The standard model for describing fast grain-boundary diffusion in a polycrystalline sample treats the system as thin grain-boundary slabs of width w^{gb} in which the diffusion coefficient D_i^{gb} is greater than the diffusion coefficient D_i in the bulk grains of width w [30]. The overall behavior that one observes depends on a number of parameters, such as the diffusion coefficients D_i^{gb} and D_i , the diffusion time, and the grain size. Equivalently, the case of fast diffusion along dislocations is treated in terms of tubes of diameter r^{dis} , present at a dislocation density of d , in which diffusion occurs at an enhanced rate D_i^{dis} . Based on the distance over which the bulk structure is perturbed, w^{gb} and r^{dis} are generally assumed to be ca. 1 nm. If, however, space charge zones in which the relevant defects are accumulated are

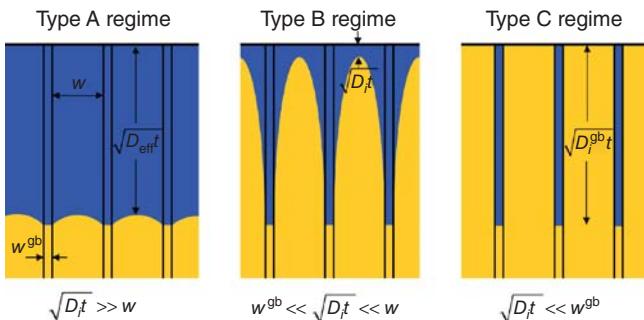


Figure 5.10 Illustration of three regimes of diffusion kinetics in a polycrystal identified by Harrison [30]. The polycrystal consists of grains of width w and diffusion coefficient D_i and grain boundaries of width w_{gb} and

diffusion coefficient D_i^{gb} . The three regimes are defined by the inequalities shown in the bottom line of the figure (t is the diffusion time). Adapted from Ref. [10].

present at such extended defects, the effective values for w_{gb} and r^{dis} will be much larger. Following the example of Atkinson [10], the three characteristic cases identified by Harrison [31] (see Figure 5.10) in the order of increasing diffusion time, that is, in reverse order, are discussed.

Type-C kinetics are observed for short diffusion times, for which the diffusant has had insufficient time to penetrate a significant distance into the bulk phase. Diffusion, therefore, takes place solely along the grain boundaries, without any flux leakage into adjacent grains. The concentration profile that is obtained, for a constant diffusion source, say, follows a complementary error function [Eq. (5.8)], with the measured diffusion coefficient corresponding to D_i^{gb} .

At longer times, diffusion in the lattice cannot be ignored and, provided the boundaries act independently, type-B diffusion kinetics are observed. “Independently” means that diffusant that moves down one short-circuit path and enters the grains is unlikely to reach another short-circuit path. In concentration profiles, fast short-circuit diffusion makes itself apparent as a “tail,” following the usual bulk diffusion profile. For independent grain boundaries, the tail is linear in a plot of $\ln c$ versus $x^{6/5}$ [32], and analysis of the tail’s slope yields the product $D_i^{\text{gb}} w_{\text{gb}}$; for independent dislocations, the tail is linear in a plot of $\ln c$ versus x [33], and analysis of the tail’s slope yields the product $D_i^{\text{dis}} r^{\text{dis}}$ (Figure 5.11).

At further longer times, the diffusion fringes around neighboring grain boundaries overlap extensively and a diffusing moiety may visit many grains and grain boundaries during the diffusion time. The grain boundaries do not act independently. This is type-A diffusion, and the concentration profile follows, for a constant diffusion source, a complementary error function; that is, it is indistinguishable from pure bulk diffusion. In this case, however, the effective diffusion coefficient D_{eff} is a function of D_i and D_i^{gb} .

The whole picture becomes far more complicated if space charge zones that are depleted of the mobile defects are present. There may be fast diffusion along

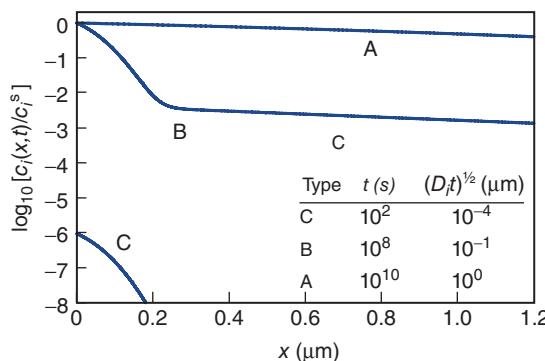


Figure 5.11 Simulated concentration profiles obtained at various times for a single crystal with dislocation density $d = 3 \times 10^7 \text{ cm}^{-2}$, dislocation radius $r_i^{\text{dis}} = 1 \text{ nm}$,

bulk diffusion coefficient $D_i = 10^{-18} \text{ cm}^2 \text{ s}^{-1}$, and dislocation diffusion coefficient $D_i^{\text{dis}} = 10^{-12} \text{ cm}^2 \text{ s}^{-1}$. Adapted from Ref. [10].

the interface, but the flux is prevented from leaving the interface by the depletion space-charge zones. In this case, the diffusion kinetics will not correspond to Harrison types A, B, or C.

5.5.2

Hindered Transport

Extended Defects

Owing to its importance in ionic oxides, we consider, specifically in this section, transport across interfaces hindered by depletion space-charge layers. This can be investigated experimentally by means of all three transport experiments (chemical, tracer, conductivity), but the tracer approach currently offers one major advantage: it is capable of resolving the profile within the space-charge zone. This is shown in Figure 5.12 for the case of a depletion space-charge zone at a surface.

When observed with moderate depth resolution (Figure 5.12c), the profile shows a drop in isotope fraction, $\Delta n_{\text{eff}}^* = j^*/k_{\text{eff}}$, that is, the combined effect of the limited surface-reaction kinetics and the depletion space-charge layer. At ultrahigh depth resolution (Figure 5.12b), the profile within the space-charge layer becomes apparent, as well as the drop in isotope fraction that arises from the surface-reaction kinetics, $\Delta n^* = j^*/k^s$.

5.6

Case Studies

The aim of this section is not to provide an overview of diffusion in oxides, for there are far too many oxide compounds in which diffusion has been examined (see Refs [36–38] for compilations of diffusion data). The aim is not even to provide a comprehensive review of all diffusion data published for selected systems. Studies of diffusion in many oxides date back to the 1960s, and correspondingly, there is a

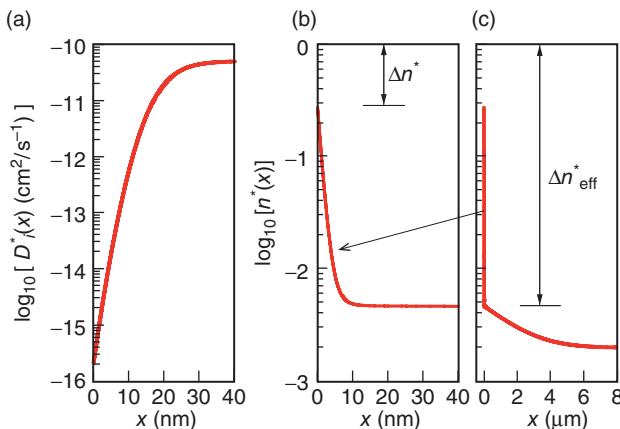


Figure 5.12 Isotope transport through an equilibrium surface space-charge layer depleted of oxygen vacancies [34, 35]. (a) Local variation of the oxygen tracer diffusion coefficient, $D^*(x) = f^* D_V n_V(x) \approx D^*(\infty) \exp[-2e\phi(x)/k_B T]$, which arises from

oxygen vacancy depletion near the surface (see Section 4.4). Solving Eq. (5.3) with this spatially variant $D^*(x)$ yields the isotope profile shown in (b) [the first 40 nm] and in (c) [the entire profile].

huge amount of data available. Furthermore, the experimental investigation of diffusion is, in general, far from simple,¹⁾ and as a consequence, some published data is inevitably of questionable quality. The aim of this section, rather, is to present diffusion data for four selected systems, in order to illustrate the current level of understanding. To this end, the author has taken a critical look at the available data, and decided in favor of those results for which (a) there are no obvious errors in the experimental procedure or in the data analysis; and (b) there is good agreement between the data obtained by diverse groups (wherever possible, using various techniques).

For three selected systems, diffusion data for both cations and anions are available, and transport across and along extended defects has also been examined. The three systems are: strontium titanate, SrTiO_3 (an oxide with small defect concentrations); yttria-stabilized zirconia (YSZ), $\text{Zr}_{1-x}\text{Y}_x\text{O}_{2-x/2}$ (an oxide with large defect concentrations); and alumina, Al_2O_3 (an oxide with small diffusion coefficients for both cations and anions). The fourth system is tantalum pentoxide, Ta_2O_5 , and at present, only the data for oxygen-ion transport in the bulk phase is available. Ta_2O_5 is discussed to draw attention to the complexity of ion transport in this “simple” oxide. Three of these four systems are unusual in the sense that they exhibit $D_{\text{anion}} \gg D_{\text{cation}}$, whereas generally for oxides, it is the other way round, that is, the cations are more mobile than the anions, MgO [39], NiO [40], CoO [11, 40], ZnO [38], TiO_2 [39], MgAl_2O_4 [39]. It is worth noting that, if the ions of one sublattice show high diffusion coefficients, the ions of opposite polarity

1) To quote W. Oswald (1891): “To make accurate experiments on diffusion is one of the most difficult problems in practical physics.”

will generally exhibit low diffusion coefficients: one sublattice has to be essentially immobile for an oxide to be classed as a solid.

Comparison of diffusion data is performed best at the level of defect diffusion coefficients, D_{def} . The benefit for systems in which the interactions between point defects are negligible is, as noted previously, that the defect diffusion coefficients are independent of defect concentration; at high defect concentrations, D_{def} is generally expected to vary with defect concentration. Extraction of D_{def} from measured D^* , D^δ , or D^σ does require, however, detailed knowledge of the defect chemistry; as we will see, this is not always available.

Reliable data for fast diffusion along grain boundaries or dislocations is seldom available. This is not solely due to the difficulties in carrying out such experiments. Extended defects often act as sinks for impurities or sites at which second phases are to be found. In other words, grain-boundary diffusion data are far more sensitive to impurities than bulk diffusion data [10]. The best approach appears to be the combination of diffusion studies with crystallographic and chemical characterization at the atomic level (with High-Resolution Transmission Electron Microscopy (HRTEM)); needless to say, studies of this sort are rare.

5.6.1

Strontium Titanate

The perovskite oxide SrTiO_3 has long served as a model system for studying point-defect behavior in perovskite oxides [41].

Given the prohibitively high energies needed to form interstitial anions or cations in the close-packed $AB\text{O}_3$ perovskite structure [42–45], the migration of ions in $AB\text{O}_3$ perovskite lattices is expected to involve vacancies as opposed to interstitials. Indeed, experimental studies of the defect chemistry of diverse $AB\text{O}_3$ perovskites are all consistent with the predominance of vacancies over interstitials. The dominant intrinsic disorder is SrO partial Schottky in SrTiO_3 (but TiO_2 partial Schottky in BaTiO_3). For oxygen-ion migration, there is, in fact, direct evidence from both atomistic simulations [43, 46–48] and experiment (neutron scattering [49, 50]) that it occurs by oxygen ions jumping into vacant sites along the $\langle 110 \rangle$ directions of the BO_6 octahedra.

In Figure 5.13 we summarize the defect diffusivities obtained for V_O^- and V_{Sr}'' in SrTiO_3 and for V_O^- and V_{Ti}'''' in BaTiO_3 . The oxygen-vacancy diffusion coefficients were obtained from acceptor-doped samples under oxidizing ($[\text{Acc}'] = 2[\text{V}_\text{O}^-]$) or highly reducing ($[\text{e}'] = 2[\text{V}_\text{O}^-]$) conditions, that is, under conditions where oxygen vacancies dominate. The cation-vacancy diffusion coefficients were obtained from donor-doped samples in which cation vacancies are, or are becoming, the dominant defects ($[\text{Don}'] = 2[\text{V}_{\text{Sr}}'']$; $[\text{Don}'] = 4[\text{V}_{\text{Ti}}'''']$). By comparing defect diffusion coefficients, we can compare results from samples with different dopant concentrations but also with different dopant types (acceptor and donor). We emphasize that the data were measured for weakly doped systems rather than solid solutions (cf. YSZ).

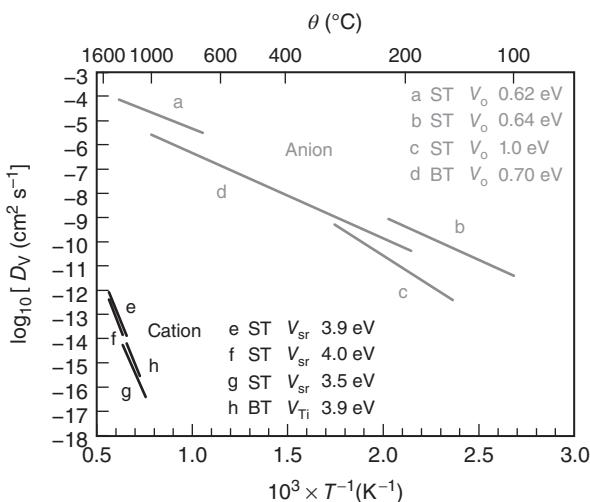


Figure 5.13 Comparison of vacancy diffusion coefficients obtained experimentally for SrTiO_3 and BaTiO_3 . Oxygen-vacancy diffusion in SrTiO_3 : (a) [35], (b) [51], (c) [52].

Oxygen-vacancy diffusion in BaTiO_3 , (d) [53]. Strontium-vacancy diffusion in SrTiO_3 : (e) [54], (f) [55], (g) [55]. Titanium-vacancy diffusion in BaTiO_3 , (h) [56].

There is extensive literature on oxygen-ion migration in SrTiO_3 (see Ref. [35] and references therein). From a critical analysis of the literature, in terms of both the absolute value of the vacancy diffusivity and the activation enthalpy of migration, one finds good agreement between macroscopic (chemical [57], tracer [35]) diffusion experiments; microscopic (nuclear spin relaxation [58] and anelastic relaxation [59]) hopping experiments; and atomistic simulations [60, 61]. This analysis yields an activation enthalpy of migration of $\Delta H_{\text{mig},V} = 0.62 \text{ eV}$ for the range $948 \leq T/\text{K} \leq 1623$ (indicated by a in Figure 5.13).

At lower temperatures, defect interactions involving oxygen vacancies may become important. There is experimental evidence that acceptor-dopant cations, such as Ni'_{Ti} [52], Mn'_{Ti} [62, 63], and Fe'_{Ti} [62, 63] trap the oxygen vacancies and increase the local activation energies for vacancy motion [64]. The consequence of the long-range migration of oxygen vacancies is an increase in the measured activation enthalpy of diffusion or conduction. In the case of Ni-doped SrTiO_3 (c in Figure 5.13), for example, the measured activation enthalpy, from conductivity investigations, increases to ca. 1 eV [52]. Not all acceptor-dopant cations cause a change in activation enthalpy; however, the interactions are evidently negligible in the case of Al'_{Ti} (b in Figure 5.13), and the migration enthalpy remains at $\Delta H_{\text{mig},V} \approx 0.6 \text{ eV}$ down to $T \approx 370 \text{ K}$ [51, 65, 66].

On moving from SrTiO_3 (a and b in Figure 5.13) to BaTiO_3 (d in Figure 5.13), one notices a small decrease in the absolute value of the oxygen-vacancy diffusivity, arising from a small increase in activation enthalpy and a decrease in the pre-exponential factor. In general, perovskites of the composition $A^{\text{II}}B^{\text{IV}}\text{O}_3$

or $A^{III}B^{III}O_3$ display activation enthalpies of vacancy migration in the range of 0.5–0.8 eV [67]; $\Delta H_{\text{gen},V}$, in contrast, may vary from zero to several electron volts.

Let us now consider the data on the diffusion of cation vacancies in titanate perovskites (e–h in Figure 5.13). Surprisingly, the data agree closely with one another, in terms of both absolute magnitude and activation enthalpy. This is particularly surprising: one would expect a considerable difference between the diffusion of V''_{Sr} in SrTiO_3 and that of V''''_{Ti} in BaTiO_3 , for Sr^{2+} is a large, divalent cation residing on the *A*-site sublattice, whereas Ti^{4+} is a small, tetravalent cation residing on the *B*-site sublattice. This behavior is not specific to the titanate perovskites. Remarkably similar diffusion coefficients D_{ion} and activation enthalpies $\Delta H_{D_{\text{ion}}}$ have been reported for *A*-site and *B*-site cations in BaTiO_3 [68, 69], $(\text{La},\text{Sr})(\text{Ga},\text{Mg})\text{O}_3$ [70], $(\text{Ba},\text{Sr})(\text{Co},\text{Fe})\text{O}_3$ [71], and MgSiO_3 [72, 73]. In those cases for which atomistic simulations have been performed, the predicted values for ΔH_{D_A} are close to the measured values, whereas the predicted values for ΔH_{D_B} are two to three times higher than those found experimentally. For example, the predicted activation energy of Sr vacancy migration in SrTiO_3 [47, 74, 75] is $\Delta H_{D_{V_{\text{Sr}}}} \sim 4$ eV, while $\Delta H_{D_{V_{\text{Ti}}}} > 8$ eV. Furthermore, atomistic simulations suggest that *B*-site migration may only take place with a relatively low activation enthalpy if an *A*-site vacancy is directly adjacent to the *B*-site vacancy [44, 45, 75].

Walsh *et al.* [47] also predict that the activation enthalpy for Sr vacancy migration decreases from 3.7 to 2.9 eV when there is a neighboring oxygen vacancy, while that for oxygen-vacancy migration increases from 0.53 to 0.89 eV with a neighboring strontium vacancy. This may explain why activation enthalpies for cation transport in titanate perovskites are either around 3.7 eV or around 2.8 eV. The former are donor-doped samples containing relatively few oxygen vacancies; the latter are acceptor-doped samples with many oxygen vacancies. Furthermore, it suggests that with the data shown in Figure 5.13, one can only predict oxygen diffusion coefficients in acceptor-doped samples and strontium/barium (and titanium?) diffusion coefficients in donor-doped material.

Grain boundaries in acceptor-doped SrTiO_3 and BaTiO_3 , which are free of second phases, are universally found to be highly resistive. Their behavior can be quantitatively accounted for in terms of positively charged grain boundaries and negatively charged space-charge layers [26, 77]. There is currently no evidence of fast transport of oxygen along grain boundaries or dislocations. Schaffrin, in a chemical diffusion experiment [77], and Watanabe *et al.*, in a tracer diffusion experiment [76], found polycrystalline samples to behave as indicated in Figure 5.9a and thus observed decrease in oxygen (concentration/isotope) profiles at grain boundaries, as indicated in Figure 5.9b. No preferential penetration of oxygen along grain boundaries was observed in either case. The results obtained by Watanabe *et al.* [78] are reproduced in Figure 5.14. (See also Ref. [79] for chemical diffusion experiments on a bicrystal.)

A recent study found no evidence for the fast diffusion of oxygen along edge dislocations comprising a low-angle tilt grain boundary [27]. The oxygen isotope profiles at the grain boundary were identical to those for the bulk phase, under both oxidizing and reducing conditions. In addition, static lattice calculations indicated

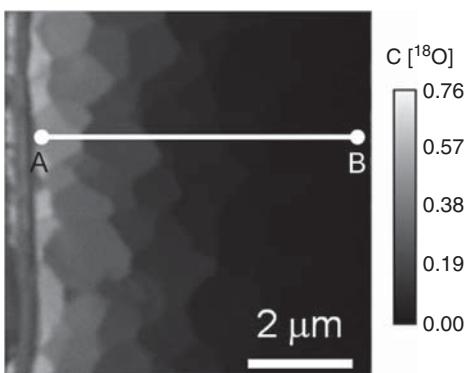


Figure 5.14 SIMS analysis of a cross section of a BaTiO_3 polycrystalline sample exposed to an $^{18}\text{O} / ^{16}\text{O}$ anneal at $T = 1123\text{ K}$ for 1 h. The grain boundaries are seen to hinder

oxygen transport. Fast transport of oxygen along grain boundaries is not observed [76]. Reproduced with kind permission of Trans Tech Publications.

a thermodynamic driving energy for space-charge formation at edge dislocations and suggested that the activation barriers for oxygen-ion migration were higher at dislocations than in the bulk. It was also argued that there are major doubts concerning the four literature reports on fast-path diffusion of oxygen along dislocations in SrTiO_3 [80–83], not only in the experimental procedure but also in the analysis and the interpretation of the results [27]. There is, therefore, no experimental evidence of fast oxygen diffusion along dislocations in SrTiO_3 . It appears that in oxide systems, in which one ion is highly mobile in the bulk phase (e.g., oxygen ions in perovskite-structured oxides and in fluorite-structured oxides), ion transport along extended defects is hindered [27, 84, 85].

5.6.2

Yttria-Stabilized Zirconia (YSZ)

ZrO_2 adopts monoclinic symmetry at room temperature. Above $T = 1480\text{ K}$, it transforms to a tetragonal structure, and finally, at $T = 2650\text{ K}$, it transforms to the cubic fluorite structure [86]. Upon substituting oxides of lower valent cations (Y_2O_3 , CaO), the cubic structure can be stabilized to lower temperatures. Many other AO_2 compounds also crystallize in the fluorite structure: HfO_2 is similar to ZrO_2 , monoclinic at room temperature. CeO_2 , ThO_2 , UO_2 , and PuO_2 are cubic at room temperature. In cubic fluorite-structured oxides, migration of anions and cations appears to occur by vacancy mechanism [38, 87, 88].

In Figure 5.15, we summarize selected tracer diffusion data obtained for YSZ with ca. 9.5 mol% Y_2O_3 . Oxygen diffusion data [89–91] show astonishingly good agreement over the range $423 \leq T/\text{K} \leq 1273$, and when combined together, they yield an activation enthalpy for oxygen diffusion of $\Delta H_{D_O^*} = (1.02 \pm 0.01)\text{ eV}$ (a in Figure 5.15). (See also Weller *et al.* [92].) It is worth emphasizing how

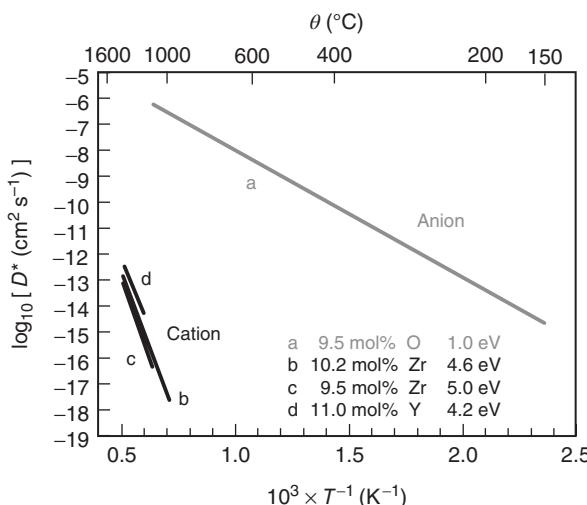


Figure 5.15 Comparison of oxygen and cation diffusion coefficients obtained experimentally for yttria-stabilized zirconia with about 9.5 mol% Y_2O_3 . (a) Oxygen tracer

diffusion [91] based on data from Refs [89, 90, 91]; (b) zirconium tracer diffusion [97]; (c) - zirconium tracer diffusion [98]; (d) yttrium tracer diffusion [99].

extraordinarily fast oxygen diffuses in this material: within 60 s at 1000 °C, an oxygen ion has moved $\sqrt{\langle R_O^2 \rangle} \approx 100 \mu\text{m}$.

The migration of oxygen ions in YSZ takes place by a vacancy mechanism. In an $\text{AO}_2 - M_2\text{O}_3$ solid solution, such as YSZ, long-range transport of oxygen is not a random walk of vacancies. It is complicated because of defect interactions [96–99]: vacancies interact with the aliovalent substituent cations; oxygen vacancies interact with one another (i.e., there is a repulsive vacancy–vacancy interaction); and the migration barriers for oxygen vacancies are modified by aliovalent substituent cations. The measured activation enthalpy of about 1 eV is therefore an effective value.

There is naturally much oxygen transport data obtained from conductivity experiments. Since the vacancies, in this case, can hardly be regarded as dilute, noninteracting defects, the use of the Nernst–Einstein equation to calculate D_O^* as given in Eq. (5.38) is incorrect. In other words, the Haven ratio H_R is not equal to $f^* = 0.65$ but takes values between 0.3 and 1.4 as a function of temperature [89, 92].

In comparison with diffusion of O, the diffusion of Y and that of Zr (b–d in Figure 5.15) are seen both to be much slower and to be characterized by much high activation enthalpies, $\Delta H_{D_{\text{cat}}^*} = 4–5$ eV. Studies from Solmon *et al.* [98] and Kilo *et al.* [93, 95] agree that Y diffusion is slightly faster than Zr diffusion. There is a problem, however, in interpreting these data. D_{Zr}^* is found experimentally to decrease with increasing $[\text{Y}'_{\text{Zr}}]$, with a power-law exponent $m_{[\text{Dop}]} \approx -2$; this suggests that cation vacancies are responsible for cation diffusion [97]:

$$D_{\text{Zr}}^* \propto [\text{V}_{\text{Zr}}'''] \propto \frac{K_{\text{Sch}}}{[\text{V}_\text{O}^\cdot]} 2 \propto \frac{K_{\text{Sch}}}{[\text{Y}'_{\text{Zr}}]^2}. \quad (5.42)$$

Atomistic simulations support this conclusion, predicting the Schottky disorder energy to be lower than Frenkel disorder energy [100, 101]. The problem is that the activation energies of cation-vacancy migration $\Delta H_{\text{mig,cat}}$ from atomistic simulations (both static [101] and molecular dynamic [102] simulations) are also around 5 eV. Since the measured activation enthalpy of diffusion is the sum of migration and generation terms [see Eq. (5.35)], $\Delta H_{D^*} = \Delta H_{\text{mig,V}_{\text{Zr}}} + \Delta H_{\text{gen,V}_{\text{Zr}}}$, one comes to the surprising conclusion that the concentration of cation vacancies hardly varies with temperature: $\Delta H_{\text{gen}} \approx 0$. The defect-chemical analysis [see Eq. (5.42)], however, suggests $\Delta H_{\text{gen}} = \Delta H_{\text{Sch}} \approx 6 \text{ eV}$ [100, 101], that is, a strong dependence on temperature. The reasons for this inconsistency are, at present, unclear but may be related to the high concentration of point defects ($\text{Y}'_{\text{Zr}}, \text{V}_\text{O}^\cdot$) in experimentally examined systems.

Oxygen-ion transport across grain boundaries is found to be hindered, and two effects have been reported extensively in the literature. In samples with high impurity levels of SiO_2 , a highly resistive, siliceous, second phase forms at the grain boundaries; in samples with negligible levels of SiO_2 , for which, according to microscopic TEM studies, the grain boundaries are free of such phases, the diminished transport rate is attributed to the presence of space-charge layers [28, 103].

Enhanced oxygen-ion diffusion along grain boundaries in nanocrystalline YSZ ceramics has been reported [104]. This study, however, employed ceramics that contained some porosity. A later study of oxygen diffusion in dense, nanocrystalline YSZ ceramics found no enhanced oxygen diffusion along grain boundaries [84]; the previous report of high grain-boundary diffusivity could be attributed to diffusion along pores and microcracks. Given that oxygen diffusion takes place so rapidly in the bulk phase, it is perhaps unsurprising to find no enhancement along grain boundaries. From TEM investigations [105, 106], grain boundaries in YSZ have been observed to be oxygen-deficient, that is, they contain higher concentrations of oxygen vacancies than the bulk phase, but these vacancies are not necessarily mobile; they may be locked in place. In any case, molecular dynamics (MD) simulations of selected grain-boundary orientations indicate that oxygen transport along the boundary is less rapid than in the bulk phase [85, 107]. Plastically deformed YSZ single crystals (i.e., samples with higher dislocation densities) were found experimentally to exhibit higher conductivities than undeformed samples, but the degree of enhancement was small (of the order of several percent) [108]. Tracer diffusion studies would be helpful in confirming that this is a real effect and not a measurement artifact.

Enhanced cation diffusion along extended defects has been reported [38]. Kilo *et al.* [95], for example, observed Harrison type-B kinetics for diffusion of Y along dislocation networks. With an effective width of 3 Å, D_Y^{gb} was found to be 10^5 higher than D_Y . The activation enthalpy obtained was $\Delta H_{D_Y^*,\text{dis}} = 3.5 \text{ eV}$ and, thus, about 70% of the bulk value. This is consistent with the simple picture of a more open structure of the dislocation cores. It is not clear, however, whether

the enhanced diffusion occurs within the dislocation cores and/or within the attendant space-charge layer (in which $V_O^{\cdot\cdot\cdot}$ are depleted, and thus, $V_{Zr}^{\prime\prime\prime\prime}$ are strongly enhanced). This example emphasizes that fast-path diffusion is generally observed only when bulk diffusion is slow.

5.6.3

Alumina

Al_2O_3 crystallizes in a number of different polymorphs. Here, we are concerned with $\alpha\text{-Al}_2\text{O}_3$, the mineral corundum. (Note that $\beta\text{-Al}_2\text{O}_3$ and $\beta''\text{-Al}_2\text{O}_3$ are mixed oxides of aluminum and sodium and excellent Na^+ -conductors.)

Mass transport in $\alpha\text{-Al}_2\text{O}_3$ is well characterized but poorly understood. Two points are undisputed (see Figure 5.16): Al diffuses faster than O in the bulk and diffusion in the bulk is slow. Even at temperatures within 10% of the melting point ($\theta_{mp} \approx 2500\text{ }^\circ\text{C}$), the diffusion coefficient of the fastest moving species (Al) reaches only $D_{\text{Al}}^* \sim 10^{-10}\text{ cm}^2/\text{s}$. Interpreting the experimental data plotted in Figure 5.16, however, is rather difficult: in the literature, mass transport in $\alpha\text{-Al}_2\text{O}_3$ is referred to as the “corundum conundrum” [109–111].

There are four sets of data for aluminum tracer diffusion in Al_2O_3 [112–114, 122]. The three datasets shown in Figure 5.16 (a [112], b [113], c [114]) agree with each other remarkably well. They can be described by a single Arrhenius slope (d in Figure 5.16), with an activation enthalpy of $\Delta H_{D_{\text{Al}}^*} = (4.0 \pm 0.1)\text{ eV}$. This is close to the value obtained from atomistic simulations for the migration

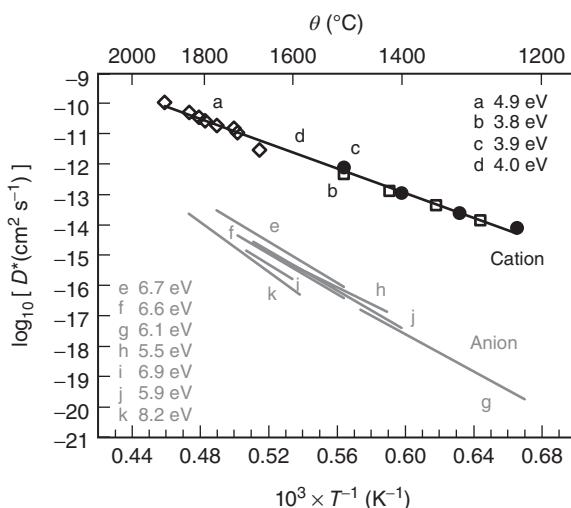


Figure 5.16 Tracer diffusion data for Al and O in $\alpha\text{-Al}_2\text{O}_3$. Al diffusion: (a) [112], (b) [114], (c) [115]. d is a fit to the three experimental datasets (a–c). O diffusion: (e) [118], (f) [119], (g) [120], (h) [121], (i) [122], (j) [123], (k) [124].

of aluminum vacancies $\Delta H_{\text{mig},V_{\text{Al}}} = 3.8 \text{ eV}$ [123] and to the value obtained from a defect-chemical analysis of electrical conductivity behavior of Ti-doped samples, assuming vacancy migration, of 3.7 eV [124].

The agreement is remarkable because the samples investigated in the three studies [112–114] were nominally undoped in two cases (with the samples in a [112] presumably having different amount and type of impurities as compared to those in b [113]) and Ti-doped in the third case (c [114]). One would perhaps expect small differences between the two undoped data sets and a considerable difference between these two and the Ti-doped sample. In any case, the insensitivity of D_{Al}^* toward Ti-doping is inconsistent with standard defect chemistry: donor dopants Ti_{Al} increase V''_{Al} , and thus, D_{Al}^* should increase substantially.

The conundrum is just as puzzling for oxygen diffusion [115–121]. Isothermal diffusion coefficients exhibit surprisingly little scatter, despite referring to different samples with varying amounts of impurities. The activation enthalpies of diffusion do exhibit some scatter, varying from 5.5 to 8.2 eV, but a value of $(6.5 \pm 0.5) \text{ eV}$ would describe most data. Whether oxygen diffusion is due to oxygen vacancies or oxygen interstitials is unclear. Simulations have, to date, only considered vacancy migration [123, 125], with the most detailed study [125] predicting $\Delta H_{\text{mig},V_{\text{O}}} = 5.3 \text{ eV}$. This is at the lower limit of the experimental values, if one assumes $\Delta H_{\text{gen},V_{\text{O}}} = 0 \text{ eV}$, which is only the case if all samples examined were effectively acceptor-doped from impurities, and electroneutrality is thus $[\text{Acc}'_{\text{Al}}] = 2[V'_{\text{O}}]$. If some samples contained a net concentration of donors, with $[\text{Don}'_{\text{Al}}] = 3[V'''_{\text{Al}}]$ as the electroneutrality condition, one finds $\Delta H_{\text{gen},V_{\text{O}}} = 7–8 \text{ eV}$ (as predicted energies for the Schottky quintet are 21–25 eV). This would effectively require $\Delta H_{\text{mig},V_{\text{O}}}$ to be zero. Hence, it appears that all samples contained a net concentration of acceptors, and oxygen migrates by a vacancy mechanism. It remains to be seen why the predicted value is lower than most experimental values, if all samples were acceptor-doped and if, perhaps, oxygen diffusion by an interstitial(cy) mechanism provides closer agreement with the experiment.

$\alpha\text{-Al}_2\text{O}_3$ represents one of the few oxide systems for which there are data for fast diffusion of anions and cations along dislocations. There is no reliable data on fast diffusion of Al along dislocations, but there is data on fast diffusion of the impurities Cr and Ti [126]. By assuming a dislocation radius of 1 nm, Nakagawa *et al.* found that D_{dis} was 4–7 orders of magnitude faster than that in the bulk, and the activation enthalpies were comparable or smaller than the values for diffusion in the bulk. Indirect evidence of accelerated oxygen diffusion along dislocations was first obtained by Reed and Wuensch [121]. Later studies [116, 118, 127] reported diffusion coefficients along dislocations 4–6 orders of magnitude higher. Particularly noteworthy is the good agreement, in terms of not only the absolute magnitude but also the activation enthalpy of diffusion, between the data obtained by Tang *et al.* [127], who examined the annihilation of dislocation dipoles [$\Delta H_{D_{\text{O}}^{\text{dis}}} = (4.5 \pm 1.3) \text{ eV}$], and the data obtained by Nakagawa *et al.* [118], who used the tracer method and observed profiles consistent with Harrison type-B

kinetics [$\Delta H_{D_O^{*,\text{dis}}} = 4.8 \text{ eV}$]. Both values of $\Delta H_{D_O^{\text{dis}}}$ are significantly lower than the values of ΔH_{D_O} obtained for oxygen diffusion in the bulk.

Fast diffusion of Al along grain boundaries has not been observed directly, but data has been extracted from creep experiments by Cannon *et al.* [128]. The absolute values are comparable to those obtained for the impurity diffusion of Cr along grain boundaries [129]. Again, the activation enthalpies for accelerated grain-boundary diffusion are lower than the values for the bulk. Oxygen apparently diffuses faster along grain boundaries [130, 131], but there are issues concerning the data [111].

In summary, fast diffusion along extended defects is observed in $\alpha\text{-Al}_2\text{O}_3$ because diffusion in the bulk is slow.

5.6.4

Tantalum Pentoxide

The crystal chemistry of tantalum pentoxide is a subject of much debate [132]. It is agreed that there are high-temperature and low-temperature modifications, termed H- Ta_2O_5 and L- Ta_2O_5 , respectively, and that the transition temperature is around 1350 °C [132]. For L- Ta_2O_5 – the phase we are interested in – an orthorhombic β phase was regarded for many years to be the most stable. Recently, extensive DFT calculations [133–135] have indicated that the β phase is unstable with respect to an orthorhombic λ phase. In addition, only the λ phase has a calculated electronic band gap that agrees well with the experimental value of about 4 eV.

The conductivity of nominally pure L- Ta_2O_5 has been examined in detail [136–142]. Figure 5.17 summarizes the diffusion coefficients that have been calculated from these conductivities (on the assumption that the measured conductivity is entirely due to oxygen ions) with the Nernst–Einstein equation [Eq. (5.38)].

At higher temperatures, data from various studies show good agreement, with an activation enthalpy of oxygen diffusion of $\Delta H_{D_O} = (1.8 \pm 0.1) \text{ eV}$. Most of the data also show surprisingly little variation in the absolute magnitude, despite the differing amount and nature of impurities that were undoubtedly present in the various samples investigated. At lower temperatures, there is a change in the activation enthalpy, to $\Delta H_{D_O} \sim 0.7 \text{ eV}$, as first noted by Kofstad [136], who also remarked that this change is accompanied by the equilibration kinetics becoming rather sluggish. In later studies [141, 142], the temperature at which the change occurs was revised from about 750 to about 850 °C. In this low-temperature regime, the calculated diffusivities are significantly different from one another.

Most authors [138, 139, 141, 142] explained the observed behavior in terms of mobile oxygen vacancies formed to compensate acceptor impurities. The change in activation enthalpy was ascribed by McHale and Tuller [141] to limited impurity solubility. Johannessen and Kofstad [140], however, found that the intentional addition of acceptor dopants ($\text{Mg}_{\text{Ta}}''', \text{Ca}_{\text{Ta}}''', \text{Y}_{\text{Ta}}'',$ or Hf_{Ta}') resulted in a decrease

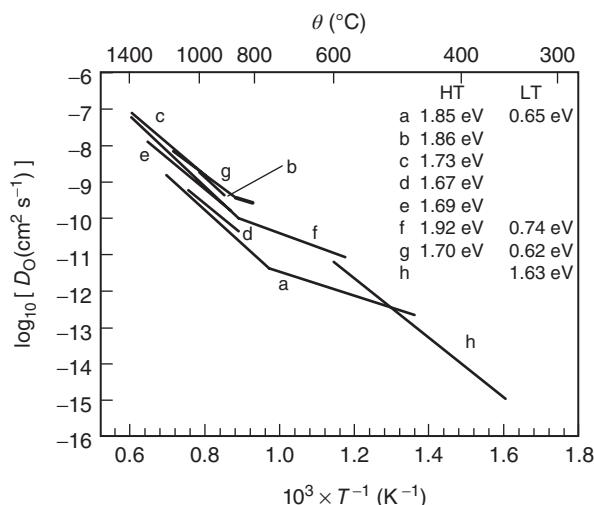


Figure 5.17 Diffusion data for oxygen in L-Ta₂O₅ obtained from electrical conductivity measurements at high oxygen partial pressures ((a) [136], (b) [137], (c) [138], (d) [139], (e) [140], (f) [141], and (g) [142]) and

from tracer diffusion measurements ((h) [143]). Note: The tracer data [143] are not corrected for tracer correlation, as f^* for L-Ta₂O₅ is not known.

in the high-temperature conductivity. Bredesen and Kostad [144] subsequently discussed various possible defect models, including anti-Frenkel disorder.

Recently, oxygen transport in L-Ta₂O₅ was examined directly by means of ¹⁸O tracer experiments [143]. The activation enthalpy of oxygen tracer diffusion in this low-temperature regime was found to be $\Delta H_{D_O} = (1.63 \pm 0.17)$ eV. By extrapolating these tracer data to higher temperatures, one finds excellent agreement with the converted high-temperature conductivity data. This suggests that the low-temperature conductivity, with its much lower activation enthalpy, has a different origin, for example, it is due to electron holes. Division of the activation enthalpy of oxygen diffusion, $\Delta H_{D_O} \sim 1.8$ eV, into the migration and generation contributions requires knowledge of the migration enthalpy of oxygen interstitials and the enthalpy of anti-Frenkel disorder. No data is currently available for either of these enthalpies.

Acknowledgments

I thank S. Beschnitt, H. Schraknepper, M. Schie, R. Waser, and M. Martin for careful and critical reading of the manuscript.

References

1. Shewmon, P. (1989) *Diffusion in Solids*, TMS, Warrendale, PA.
2. Mehrer, H. (2007) *Diffusion in Solids*, Springer-Verlag, Berlin, Heidelberg, New York.
3. Schmalzried, H. (1971) *Festkörperreaktion*, Verlag Chemie, Weinheim.
4. Balluffi, R.W., Allen, S.M., and Carter, W.C. (2005) *Kinetics of Materials*, John Wiley & Sons, Inc., Hoboken, NJ.
5. Borg, R.J. and Dienes, G.J. (1988) *An Introduction to Solid State Diffusion*, Academic Press, San Diego, CA.
6. Maier, J. (2004) *Physical Chemistry of Ionic Materials: Ions and Electrons in Solids*, John Wiley & Sons, Ltd, Chichester.
7. Howard, R.E. and Lidiard, A.B. (1964) Matter transport in solids. *Rep. Prog. Phys.*, **27** (1), 161, doi: <http://stacks.iop.org/0034-4885/27/i=1/a=305>.
8. Wagner, C. (1975) Equations for transport in solid oxides and sulfides of transition metals. *Prog. Solid State Chem.*, **10** Part 1, 3–16, doi: 10.1016/0079-6786(75)90002-3.
9. Murch, G.E. (2001) Diffusion kinetics in solids, in *Phase Transformations in Materials* (ed. G. Kostorz), Wiley-VCH Verlag GmbH, Weinheim, pp. 173–238.
10. Atkinson, A. (1984) Diffusion along grain boundaries and dislocations in oxides, alkali halides and carbides. *Solid State Ionics*, **12**, 309–320, doi: 10.1016/0167-2738(84)90160-7.
11. Martin, M. (2005) Diffusion in oxides, in *Diffusion in Condensed Matter* (eds P. Heijmans and J. Kaerger), Springer-Verlag, Berlin, Heidelberg, New York, pp. 211–249.
12. De Souza, R.A. and Martin, M. (2009) Probing diffusion kinetics with secondary ion mass spectrometry. *MRS Bull.*, **34** (12), 907–914.
13. Janek, J., Sann, J., Mogwitz, B., Rohnke, M., and Kleine-Boymann, M. (2012) Degradation of functional materials in temperature gradients – thermodiffusion and the soot effect. *J. Korean Ceram. Soc.*, **49** (1), 56–65.
14. Lee, T., Kim, H.S., and Yoo, H.I. (2014) From onsager to mixed ionic electronic conductors. *Solid State Ionics*, **262**, 2–8, doi: 10.1016/j.ssi.2013.10.048.
15. Crank, J. (1975) *The Mathematics of Diffusion*, 2nd ed. Oxford University Press, Oxford.
16. Vineyard, G.H. (1957) Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids*, **3** (1-2), 121–127, doi: 10.1016/0022-3697(57)90059-8.
17. de Groot, S.R. and Mazur, P. (1962) *Non-Equilibrium Thermodynamics*, North Holland, Amsterdam.
18. Lee, D.K. and Yoo, H.I. (2006) Electron-ion interference and Onsager reciprocity in mixed ionic-electronic transport in TiO₂. *Phys. Rev. Lett.*, **97**, 255901, doi: 10.1103/PhysRevLett.97.255901.
19. Wagner, C. (1972) The thermoelectric power of cells with ionic compounds involving ionic and electronic conduction. *Prog. Solid State Chem.*, **7**, 1–37, doi: 10.1016/0079-6786(72)90003-9.
20. Kim, H.S. and Yoo, H.I. (2011) Compilation of all the isothermal mass/charge transport properties of the mixed conducting La₂NiO₄+[small delta] at elevated temperatures. *Phys. Chem. Chem. Phys.*, **13**, 4651–4658, doi: 10.1039/C0CP02120B.
21. Verwey, E. (1935) Electrolytic conduction of a solid insulator at high fields the formation of the anodic oxide film on aluminium. *Physica*, **2** (1-12), 1059–1063, doi: 10.1016/S0031-8914(35)90193-8.
22. Mott, N.F. and Gurney, R.W. (1950) *Electronic Processes in Ionic Crystal*, 2nd ed. Oxford University Press, Oxford.
23. Kaur, I., Mishin, Y., and Gust, W. (1995) *Fundamentals of Grain and*

- Interphase Boundary Diffusion*, John Wiley & Sons, Ltd, Chichester.
24. Peterson, N. (1983) Grain-boundary diffusion in metals. *Int. Mater. Rev.*, **28** (1), 65–91, doi: 10.1179/imr.1983.28.1.65.
 25. Balluffi, R. and Mehl, R. (1982) Grain boundary diffusion mechanisms in metals. *Metall. Trans. A*, **13** (12), 2069–2095, doi: 10.1007/BF02648378.
 26. De Souza, R.A. (2009) The formation of equilibrium space-charge zones at grain boundaries in the perovskite oxide SrTiO₃. *Phys. Chem. Chem. Phys.*, **11**, 9939–9969, doi: 10.1039/B904100A.
 27. Metlenko, V., Ramadan, A., Gunkel, F., Du, H., Schraknepper, H., Hoffmann-Eifert, S., Dittmann, R., Waser, R., and De Souza, R. (2014) Do dislocations act as atomic autobahns for oxygen in the perovskite oxide SrTiO₃? *Nanoscale*, **6** (21), 12–864–12–876. doi: 10.1039/C4NR04083J.
 28. Lee, J.H. (2009) Highly resistive intergranular phases in solid electrolytes: an overview. *Monatsh. Chem. Chem. Mon.*, **140** (9), 1081–1094, doi: 10.1007/s00706-009-0111-0.
 29. De Souza, R.A. (2006) A universal empirical expression for the isotope surface exchange coefficients (k^*) of acceptor-doped perovskite and fluorite oxides. *Phys. Chem. Chem. Phys.*, **8**, 890–897, doi: 10.1039/B511702J.
 30. Fisher, J.C. (1951) Calculation of diffusion penetration curves for surface and grain boundary diffusion. *J. Appl. Phys.*, **22** (1), 74–77, doi: 10.1063/1.1699825.
 31. Harrison, L.G. (1961) Influence of dislocations on diffusion kinetics in solids with particular reference to the alkali halides. *Trans. Faraday Soc.*, **57**, 1191–1199, doi: 10.1039/TF9615701191.
 32. Le Claire, A.D. (1963) The analysis of grain boundary diffusion measurements. *Br. J. Appl. Phys.*, **14** (6), 351, doi: 10.1088/0508-3443/14/i=6/a=317.
 33. Le Claire, A.D. and Rabinovitch, A. (1981) A mathematical analysis of diffusion in dislocations. I. Application to concentration 'tails'. *J. Phys. C*: *Solid State Phys.*, **14** (27), 3863, doi: 10.1088/0022-3719/14/i=27/a=011.
 34. De Souza, R.A. and Martin, M. (2008) Using ¹⁸O/¹⁶O exchange to probe an equilibrium space-charge layer at the surface of a crystalline oxide: method and application. *Phys. Chem. Chem. Phys.*, **10**, 2356–2367, doi: 10.1039/B719618K.
 35. De Souza, R.A., Metlenko, V., Park, D., and Weirich, T.E. (2012) Behavior of oxygen vacancies in single-crystal SrTiO_3 : equilibrium distribution and diffusion kinetics. *Phys. Rev. B*, **85**, 174109, doi: 10.1103/PhysRevB.85.174109.
 36. Harrop, P.J. (1968) Self-diffusion in simple oxides. *J. Mater. Sci.*, **3**, 206–222.
 37. Freer, R. (1980) Self-diffusion and impurity diffusion in oxides. *J. Mater. Sci.*, **15**, 803–824.
 38. Borchardt, G., Gömann, K., Kilo, M., and Schmidt, H. (2008) Diffusion in ceramics, in *Ceramics Science and Technology*, Vol. 1, Chapter 4 (eds R. Riedel and I. Chen), Wiley-VCH Verlag GmbH, Weinheim, pp. 105–180.
 39. Van Orman, J.A. and Crispin, K.L. (2010) Diffusion in oxides. *Rev. Mineral. Geochem.*, **72** (1), 757–825, doi: 10.2138/rmg.2010.72.17.
 40. Peterson, N. (1984) Impurity diffusion in transition-metal oxides. *Solid State Ionics*, **12**, 201–215, doi: 10.1016/0167-2738(84)90149-8.
 41. De Souza, R.A., Fleig, J., Merkle, R., and Maier, J. (2003) SrTiO₃: a model electroceramic. *Z. Metallkunde*, **94**, 218–225, doi: 10.3139/146.030218.
 42. Akhtar, M.J., Akhtar, Z.U.N., Jackson, R.A., and Catlow, C.R.A. (1995) Computer simulation studies of strontium titanate. *J. Am. Ceram. Soc.*, **78** (2), 421–428, doi: 10.1111/j.1151-2916.1995.tb08818.x.
 43. Islam, M.S. (2000) Ionic transport in ABO₃ perovskite oxides: a computer modelling tour. *J. Mater. Chem.*, **10**, 1027–1038, doi: 10.1039/A908425H.
 44. De Souza, R.A., Islam, M.S., and Ivers-Tiffée, E. (1999) Formation and migration of cation defects in the perovskite oxide LaMnO₃. *J. Mater. Chem.*, **9**, 1621–1627, doi: 10.1039/A901512D.

45. De Souza, R.A. and Maier, J. (2003) A computational study of cation defects in LaGaO₃. *Phys. Chem. Chem. Phys.*, **5**, 740–748, doi: 10.1039/B209062G.
46. Cherry, M., Islam, M., and Catlow, C. (1995) Oxygen ion migration in perovskite-type oxides. *J. Solid State Chem.*, **118** (1), 125–132, doi: 10.1006/jssc.1995.1320.
47. Walsh, A., Catlow, C.R.A., Smith, A.G.H., Sokol, A.A., and Woodley, S.M. (2011) Strontium migration assisted by oxygen vacancies in SrTiO₃ from classical and quantum mechanical simulations. *Phys. Rev. B*, **83**, 220301–. doi: 10.1103/PhysRevB.83.220301.
48. Schie, M., Marchewka, A., Müller, T., De Souza, R.A., and Waser, R. (2012) Molecular dynamics simulations of oxygen vacancy diffusion in SrTiO₃. *J. Phys. Condens. Matter*, **24** (48), 485002, doi: 10.1088/0953-8984/24/i=48/a=485002.
49. Lerch, M., Boysen, H., and Hansen, T. (2001) High-temperature neutron scattering investigation of pure and doped lanthanum gallate. *J. Phys. Chem. Solids*, **62** (3), 445–455, doi: 10.1016/S0022-3697(00)00078-0.
50. Günter, M.M., Boysen, H., Corte, K., Lerch, M., and Suard, E. (2005) In-situ investigation of oxygen diffusion in Sr-, Mg-doped LaGaO₃ superionic conductors with a simultaneously applied electric field. *Z. Kristallogr.*, **220**, 218–224, doi: 10.1524/zkri.220.2.218.59125.
51. Guo, X. (2009) Comment on “colossal ionic conductivity at interfaces of epitaxial ZrO₂:Y₂O₃/SrTiO₃ heterostructures”. *Science*, **324** (5926), 465, doi: 10.1126/science.1168940.
52. Waser, R. (1991) Bulk conductivity and defect chemistry of acceptor-doped strontium titanate in the quenched state. *J. Am. Ceram. Soc.*, **74** (8), 1934–1940, doi: 10.1111/j.1151-2916.1991.tb07812.x.
53. Kessel, M. (2012) *Anionen- und Kationendiffusion in Barium- und Strontiumtitannat*. PhD thesis. RWTH Aachen University.
54. Meyer, R., Waser, R., Helmbold, J., and Borchardt, G. (2003) Observation of vacancy defect migration in the cation sublattice of complex oxides by ¹⁸O tracer experiments. *Phys. Rev. Lett.*, **90**, 105901, doi: 10.1103/PhysRevLett.90.105901.
55. Gomann, K., Borchardt, G., Schulz, M., Gomann, A., Maus-Friedrichs, W., Lesage, B., Kaitasov, O., Hoffmann-Eifert, S., and Schneller, T. (2005) Sr diffusion in undoped and La-doped SrTiO₃ single crystals under oxidizing conditions. *Phys. Chem. Chem. Phys.*, **7**, 2053–2060, doi: 10.1039/B418824A.
56. Preis, W. and Sitte, W. (2006) Electronic conductivity and chemical diffusion in n-conducting barium titanate ceramics at high temperatures. *Solid State Ionics*, **177** (35–36), 3093–3098, doi: 10.1016/j.ssi.2006.07.053.
57. Schwarz, D.B. and Anderson, H.U. (1975) Determination of oxygen chemical diffusion coefficients in single crystal SrTiO₃ by capacitance manometry. *J. Electrochem. Soc.*, **122** (5), 707–710, doi: 10.1149/1.2134297.
58. Hackmann, A. and Kanert, O. (1991) NMR investigation of defect properties in single crystal SrTiO₃. *Radiat. Eff. Defects Solids*, **119-121** (2), 651–656, doi: 10.1080/10420159108220797.
59. Cordero, F. (2007) Hopping and clustering of oxygen vacancies in SrTiO₃ by anelastic relaxation. *Phys. Rev. B*, **76**, 172106, doi: 10.1103/PhysRevB.76.172106.
60. Cuong, D.D., Lee, B., Choi, K.M., Ahn, H.S., Han, S., and Lee, J. (2007) Oxygen vacancy clustering and electron localization in oxygen-deficient SrTiO₃: LDA+U study. *Phys. Rev. Lett.*, **98**, 115503, doi: 10.1103/PhysRevLett.98.115503.
61. Lontsi-Fomena, M., Villesuzanne, A., Doumerc, J.P., Frayret, C., and Pouchard, M. (2008) A density functional theory study of oxygen diffusion in LaAlO₃ and SrTiO₃. *Comput. Mater. Sci.*, **44** (1), 53–60, doi: 10.1016/j.commatsci.2008.01.046.
62. Müller, K.A. (1981) Paramagnetic point and pair defects in oxide perovskites. *J. Phys. France*, **42** (4), 551–557, doi: 10.1051/jphys:01981004204055100.

63. Merkle, R. and Maier, J. (2003) Defect association in acceptor-doped SrTiO₃: case study for Fe_{Ti}V_O and Mn_{Ti}V_O. *Phys. Chem. Chem. Phys.*, **5**, 2297–2303, doi: 10.1039/B300205P.
64. Schie, M., Waser, R., and De Souza, R.A. (2014) Oxygen vacancy behavior in strontium titanate: more than a nearest neighbor phenomenon. *J. Phys. Chem. C*, **118**, 15185–15192, doi: 10.1021/jp504436t.
65. Wang, C.C., Lei, C.M., Wang, G.J., Sun, X.H., Li, T., Huang, S.G., Wang, H., and Li, Y.D. (2013) Oxygen-vacancy-related dielectric relaxations in SrTiO₃ at high temperatures. *J. Appl. Phys.*, **113** (9), 094103, doi: 10.1063/1.4794349.
66. De Souza, R.A. and Ramadan, A.H.H. (2013) Ionic conduction in the SrTiO₃|YSZ|SrTiO₃ heterostructure. *Phys. Chem. Chem. Phys.*, **15**, 4505–4509, doi: 10.1039/C3CP44399J.
67. Ishigaki, T., Yamuchi, S., Kishio, K., Mizusaki, J., and Fueki, K. (1988) Diffusion of oxide ion vacancies in perovskite-type oxides. *J. Solid State Chem.*, **73** (1), 179–187, doi: 10.1016/0022-4596(88)90067-9.
68. Körfer, S., De Souza, R.A., Yoo, H.I., and Martin, M. (2008) Diffusion of Sr and Zr in BaTiO₃ single crystals. *Solid State Sci.*, **10** (6), 725–734, doi: 10.1016/j.solidstatesciences.2007.06.017.
69. Yoo, H.I., Lee, C.E., De Souza, R.A., and Martin, M. (2008) Equal mobility of constituent cations in BaTiO₃. *Appl. Phys. Lett.*, **92** (25), 252103, doi: 10.1063/1.2951606.
70. Schulz, O., Martin, M., Argirusis, C., and Borchardt, G. (2003) Cation tracer diffusion of ¹³⁸La, ⁸⁴Sr and ²⁵Mg in polycrystalline La_{0.9}Sr_{0.1}Ga_{0.9}Mg_{0.1}O_{2.9}. *Phys. Chem. Chem. Phys.*, **5**, 2308–2313, doi: 10.1039/B301882M.
71. Harvey, S.P., De Souza, R.A., and Martin, M. (2012) Diffusion of La and Mn in Ba_{0.5}Sr_{0.5}Co_{0.8}Fe_{0.2}O_{3-δ} [small delta] polycrystalline ceramics. *Energy Environ. Sci.*, **5**, 5803–5813, doi: 10.1039/C1EE02740A.
72. Yamazaki, D., Kato, T., Yurimoto, H., Ohtani, E., and Toriumi, M. (2000) Silicon self-diffusion in MgSiO₃ perovskite at 25 GPa. *Phys. Earth Planet. Inter.*, **119** (3-4), 299–309, doi: 10.1016/S0031-9201(00)00135-7.
73. Holzapfel, C., Rubie, D.C., Frost, D.J., and Langenhorst, F. (2005) Fe-Mg interdiffusion in (Mg,Fe)SiO₃ perovskite and lower mantle reequilibration. *Science*, **309** (5741), 1707–1710, doi: 10.1126/science.1111895.
74. Thomas, B., Marks, N., and Begg, B. (2007) Defects and threshold displacement energies in SrTiO₃ perovskite using atomistic computer simulations. *Nucl. Instrum. Methods Phys. Res., Sect. B*, **254** (2), 211–218, doi: 10.1016/j.nimb.2006.11.069.
75. Mizoguchi, T., Takahashi, N., and Lee, H.S. (2011) First-principles study on migration mechanism in SrTiO₃. *Appl. Phys. Lett.*, **98** (9), 091909, doi: 10.1063/1.3560464.
76. Waser, R. and Hagenbeck, R. (2000) Grain boundaries in dielectric and mixed-conducting ceramics. *Acta Mater.*, **48** (4), 797–825, doi: 10.1016/S1359-6454(99)00367-5.
77. Schaffrin, C. (1976) Oxygen diffusion in BaTiO₃ ceramic. *Phys. Status Solidi A*, **35** (1), 79–88, doi: 10.1002/pssa.2210350109.
78. Watanabe, K., Sakaguchi, I., Hishita, S., Haneda, H., and Ohashi, N. (2013) Oxygen diffusion in BaTiO₃ ceramic. *Key Eng. Mater.*, **566**, 262–265, doi: 10.4028/www.scientific.net/KEM. 566.262.
79. Leonhardt, M., Jamnik, J., and Maier, J. (1999) In situ monitoring and quantitative analysis of oxygen diffusion through Schottky-barriers in SrTiO₃ bicrystals. *Electrochim. Solid-State Lett.*, **2** (7), 333–335, doi: 10.1149/1.1390827.
80. Waugh, J.S., Paladino, A.E., DiBenedetto, B., and Wantman, R. (1963) Effect of dislocations on oxidation and reduction of single-crystal SrTiO₃. *J. Am. Ceram. Soc.*, **46** (1), 60, doi: 10.1111/j.1151-2916.1963.tb13774.x.
81. Paladino, A.E., Rubin, L.G., and Waugh, J.S. (1965) Oxygen ion diffusion in single crystal SrTiO₃. *J. Chem. Phys. Solids*, **26**, 391–397,

- <http://www.sciencedirect.com/science/article/pii/002236976590168X>.
82. Amante, J.C., Crawley, J.D., Kim, J., and Lemberger, T.R. (1991) Oxygen tracer diffusion in SrTiO₃ and thin films of YBa₂Cu₃O_{7-x} on SrTiO₃, in *Point Defects and Related Properties of Ceramics, Ceramic Trans* (eds T. Mason and J.L. Routbort), American Ceramics Society, pp. 303–312.
 83. Szot, K., Speier, W., Carius, R., Zastrow, U., and Beyer, W. (2002) Localized metallic conductivity and self-healing during thermal reduction of SrTiO₃. *Phys. Rev. Lett.*, **88**, 075 508, doi: 10.1103/PhysRevLett.88.075508.
 84. De Souza, R.A., Pietrowski, M.J., Anselmi-Tamburini, U., Kim, S., Munir, Z.A., and Martin, M. (2008) Oxygen diffusion in nanocrystalline yttria-stabilized zirconia: the effect of grain boundaries. *Phys. Chem. Chem. Phys.*, **10**, 2067–2072, doi: 10.1039/B719363G.
 85. Fisher, C. and Matsubara, H. (1998) Oxide ion diffusion along grain boundaries in zirconia: a molecular dynamics study. *Solid State Ionics*, **113-115**, 311–318, doi: 10.1016/S0167-2738(98)00380-4.
 86. Kisi, E.H. and Howard, C.J. (1998) Crystal structures of zirconia phases and their inter-relation. *Key Eng. Mater.*, **153-154**, 1–35, doi: 10.4028/www.scientific.net/KEM.153-154.1.
 87. Ando, K. and Oishi, Y. (1983) Diffusion characteristics of actinide oxides. *J. Nucl. Sci. Technol.*, **20** (12), 973–982, doi: 10.1080/18811248.1983.9733499.
 88. Matzke, H. (1987) Atomic transport properties in UO₂ and mixed oxides (U, Pu)O₂. *J. Chem. Soc., Faraday Trans. 2*, **83**, 1121–1142, doi: 10.1039/F29878301121.
 89. Manning, P., Sirman, J., De Souza, R., and Kilner, J. (1997) The kinetics of oxygen transport in 9.5 mol single crystal yttria stabilised zirconia. *Solid State Ionics*, **100** (1-2), 1–10, doi: 10.1016/S0167-2738(97)00345-7.
 90. Röwer, R., Knöner, G., Reimann, K., Schaefer, H.E., and Södervall, U. (2003) Oxygen diffusion in YSZ single crystals at relatively low temperatures. *Phys. Status Solidi B*, **239** (2), R1–R3, doi: 10.1002/pssb.200309011.
 91. Pietrowski, M.J., De Souza, R.A., Fartmann, M., ter Veen, R., and Martin, M. (2013) Oxygen isotope transport properties of yttria-stabilized zirconia (YSZ) in O₂- and H₂O-containing atmospheres. *Fuel Cells*, **13** (5), 673–681, doi: 10.1002/fuce.201300087.
 92. Weller, M., Herzog, R., Kilo, M., Borchardt, G., Weber, S., and Scherrer, S. (2004) Oxygen mobility in yttria-doped zirconia studied by internal friction, electrical conductivity and tracer diffusion experiments. *Solid State Ionics*, **175** (1-4), 409–413, doi: 10.1016/j.ssi.2003.12.044.
 93. Kilo, M., Borchardt, G., Lesage, B., Kaitasov, O., Weber, S., and Scherrer, S. (2000) Cation transport in yttria stabilized cubic zirconia: 96Zr tracer diffusion in (ZrxY1-x)O2-x/2 single crystals with 0.15>x>0.48. *J. Eur. Ceram. Soc.*, **20** (12), 2069–2077, doi: [http://dx.doi.org/10.1016/S0955-2219\(00\)00072-8](http://dx.doi.org/10.1016/S0955-2219(00)00072-8).
 94. Solmon, H., Chaumont, J., Dolin, C., and Monty, C. (2003) Y, Zr, and O self diffusion in Zr_{1-x}Y_xO_{2-x/2} (x=0.17). *J. Appl. Phys.*, **94** (12), 7547–7552, doi: 10.1063/1.1628379.
 95. Kilo, M., Taylor, M.A., Argirusis, C., Borchardt, G., Lesage, B., Weber, S., Scherrer, S., Scherrer, H., Schroeder, M., and Martin, M. (2003) Cation self-diffusion of 44Ca, 88Y, and 96Zr in single-crystalline calcia- and yttria-doped zirconia. *J. Appl. Phys.*, **94** (12), 7547–7552, doi: 10.1063/1.1628379.
 96. Murray, A., Murch, G., and Catlow, C. (1986) A new hybrid scheme of computer simulation based on hades and monte carlo: application to ionic conductivity in Y³⁺ doped CeO₂. *Solid State Ionics*, **18-19**, Part 1, 196–202, doi: 10.1016/0167-2738(86)90111-6.
 97. Meyer, M., Nicoloso, N., and Jaenisch, V. (1997) Percolation model for the anomalous conductivity of fluorite-related oxides. *Phys. Rev. B*, **56**, 5961–5966, doi: 10.1103/PhysRevB.56.5961.
 98. Krishnamurthy, R., Yoon, Y.G., Srolovitz, D.J., and Car, R. (2004)

- Oxygen diffusion in yttria-stabilized zirconia: a new simulation model. *J. Am. Ceram. Soc.*, **87** (10), 1821–1830, doi: 10.1111/j.1151-2916.2004.tb06325.x.
99. Martin, M. (2006) On the ionic conductivity of strongly acceptor doped, fluorite-type oxygen ion conductors. *J. Electroceram.*, **17** (2-4), 765–773, doi: 10.1007/s10832-006-6007-z.
100. Dwivedi, A. and Cormack, A.N. (1990) A computer simulation study of the defect structure of calcia-stabilized zirconia. *Philos. Mag. A*, **61** (1), 1–22, doi: 10.1080/01418619008235554.
101. Kilo, M., Jackson, R.A., and Borchardt, G. (2003) Computer modelling of ion migration in zirconia. *Philos. Mag.*, **83** (29), 3309–3325, doi: 10.1080/14786430310001605001.
102. González-Romero, R., Meléndez, J., Gómez-García, D., Cumbre, F., Domínguez-Rodríguez, A., and Wakai, F. (2011) Cation diffusion in yttria-zirconia by molecular dynamics. *Solid State Ionics*, **204–205**, 1–6, doi: <http://dx.doi.org/10.1016/j.ssi.2011.10.006>.
103. Guo, X. and Waser, R. (2006) Electrical properties of the grain boundaries of oxygen ion conductors: acceptor-doped zirconia and ceria. *Prog. Mater Sci.*, **51** (2), 151–210, doi: 10.1016/j.pmatsci.2005.07.001.
104. Knöner, G., Reimann, K., Röwer, R., Södervall, U., and Schaefer, H.E. (2003) Enhanced oxygen diffusivity in interfaces of nanocrystalline ZrO₂-Y₂O₃. *Proc. Natl. Acad. Sci. U.S.A.*, **100** (7), 3870–3873, doi: 10.1073/pnas.0730783100.
105. Backhaus-Ricoult, M.B., Badding, M., and Thibault, Y. (2006) Grain boundary segregation and conductivity in yttria-stabilized zirconia. *Ceram. Trans.*, **179**, 173–191.
106. Lei, Y., Ito, Y., Browning, N.D., and Mazanec, T.J. (2002) Segregation effects at grain boundaries in fluorite-structured ceramics. *J. Am. Ceram. Soc.*, **85** (9), 2359–2363, doi: 10.1111/j.1151-2916.2002.tb00460.x.
107. González-Romero, R.L., Meléndez, J.J., Gómez-García, D., Cumbre, F.L., and Domínguez-Rodríguez, A. (2012) A molecular dynamics study of grain boundaries in YSZ: structure, energetics and diffusion of oxygen. *Solid State Ionics*, **219**, 1–10, doi: 10.1016/j.ssi.2012.05.004.
108. Otsuka, K., Kuwabara, A., Nakamura, A., Yamamoto, T., Matsunaga, K., and Ikuhara, Y. (2003) Dislocation-enhanced ionic conductivity of yttria-stabilized zirconia. *Appl. Phys. Lett.*, **82** (6), 877–879, doi: 10.1063/1.1544440.
109. Heuer, A. and Lagerlof, K. (1999) Oxygen self-diffusion in corundum (α -Al₂O₃): a conundrum. *Philos. Mag. Lett.*, **79** (8), 619–627, doi: 10.1080/095008399177002.
110. Heuer, A. (2008) Oxygen and aluminum diffusion in α -Al₂O₃: how much do we really understand? *J. Eur. Ceram. Soc.*, **28** (7), 1495–1507, doi: 10.1016/j.eurceramsoc.2007.12.020.
111. Harding, J.H., Atkinson, K.J.W., and Grimes, R.W. (2003) Experiment and theory of diffusion in alumina. *J. Am. Ceram. Soc.*, **86** (4), 554–559, doi: 10.1111/j.1151-2916.2003.tb03340.x.
112. Paladino, A.E. and Kingery, W.D. (1962) Aluminum ion diffusion in aluminum oxide. *J. Chem. Phys.*, **37** (5), 957–962, doi: 10.1063/1.1733252.
113. Fielitz, P., Borchardt, G., Ganschow, S., and Bertram, R. (2012) 26Al tracer diffusion in nominally undoped single crystalline α -Al₂O₃. *Defect Diffus. Forum*, **323–325**, 75–79.
114. Fielitz, P., Borchardt, G., Ganschow, S., Bertram, R., and Markwitz, A. (2008) 26Al tracer diffusion in titanium doped single crystalline α -Al₂O₃. *Solid State Ionics*, **179** (11-12), 373–379, doi: 10.1016/j.ssi.2008.03.007.
115. Oishi, Y., Ando, K., and Kubota, Y. (1980) Self-diffusion of oxygen in single crystal alumina. *J. Chem. Phys.*, **73** (3), 1410–1412, doi: 10.1063/1.440201.
116. Prot, D. and Monty, C. (1996) Self-diffusion in α -Al₂O₃. II. Oxygen diffusion in ‘undoped single’ crystals. *Philos. Mag. A*, **73** (4), 899–917, doi: 10.1080/01418619608243695.
117. Lagerlof, K.P.D., Mitchell, T.E., and Heuer, A.H. (1989) Lattice diffusion

- kinetics in undoped and impurity-doped sapphire (α -Al₂O₃): a dislocation loop annealing study. *J. Am. Ceram. Soc.*, **72** (11), 2159–2171, doi: 10.1111/j.1151-2916.1989.tb06049.x.
118. Nakagawa, T., Nakamura, A., Sakaguchi, I., Shibata, N., Lagerlöf, K.P.D., Yamamoto, T., Haneda, H., and Ikuhara, Y. (2006) Oxygen pipe diffusion in sapphire basal dislocation. *J. Ceram. Soc. Jpn.*, **114** (1335), 1013–1017.
119. Oishi, Y., Ando, K., Suga, N., and Kingery, W.D. (1983) Effect of surface condition on oxygen self-diffusion coefficients for single-crystal Al₂O₃. *J. Am. Ceram. Soc.*, **66** (8), C-130–C-131, doi: 10.1111/j.1151-2916.1983.tb10106.x.
120. Cawley, J.D., Halloran, J.W., and Cooper, A.R. (1991) Oxygen tracer diffusion in single-crystal alumina. *J. Am. Ceram. Soc.*, **74** (9), 2086–2092. doi: 10.1111/j.1151-2916.1991.tb08264.x.
121. Reed, D.J. and Wuensch, B.J. (1980) Ion-probe measurement of oxygen self-diffusion in single-crystal Al₂O₃. *J. Am. Ceram. Soc.*, **63** (1-2), 88–92, doi: 10.1111/j.1151-2916.1980.tb10655.x.
122. Gall, M.L., Lesage, B., and Bernardini, J. (1994) Self-diffusion in α -Al₂O₃ I. Aluminium diffusion in single crystals. *Philos. Mag. A*, **70** (5), 761–773, doi: 10.1080/01418619408242929.
123. Jacobs, P.W.M. and Kotomin, E.A. (1993) Defect energies for pure corundum and for corundum doped with transition metal ions. *Philos. Mag. A*, **68** (4), 695–709, doi: 10.1080/01418619308213992.
124. Mohapatra, S.K. and Kröger, F.A. (1977) Defect structure of α -Al₂O₃ doped with titanium. *J. Am. Ceram. Soc.*, **60** (9-10), 381–387, doi: 10.1111/j.1151-2916.1977.tb15517.x.
125. Aschauer, U., Bowen, P., and Parker, S. (2009) Oxygen vacancy diffusion in alumina: new atomistic simulation methods applied to an old problem. *Acta Mater.*, **57** (16), 4765–4772, doi: 10.1016/j.actamat.2009.06.061.
126. Nakagawa, T., Nakamura, A., Sakaguchi, I., Shibata, N., Mizoguchi, T., Yamamoto, T., Haneda, H., Ohashi, N., and Ikuhara, Y. (2011) Cation diffusion along basal dislocations in sapphire. *Acta Mater.*, **59** (3), 1105–1111, doi: 10.1016/j.actamat.2010.10.041.
127. Tang, X., Lagerlöf, K.P.D., and Heuer, A.H. (2003) Determination of pipe diffusion coefficients in undoped and magnesia-doped sapphire (α -Al₂O₃): a study based on annihilation of dislocation dipoles. *J. Am. Ceram. Soc.*, **86** (4), 560–565, doi: 10.1111/j.1151-2916.2003.tb03341.x.
128. Cannon, R.M., Rhodes, W.H., and Heuer, A.H. (1980) Plastic deformation of fine-grained alumina (Al₂O₃): I, interface-controlled diffusional creep. *J. Am. Ceram. Soc.*, **63** (1-2), 46–53, doi: 10.1111/j.1151-2916.1980.tb10648.x.
129. Bedu-Amissah, K., Rickman, J.M., Chan, H.M., and Harmer, M.P. (2007) Grain-boundary diffusion of Cr in pure and Y-doped alumina. *J. Am. Ceram. Soc.*, **90** (5), 1551–1555, doi: 10.1111/j.1551-2916.2007.01584.x.
130. Oishi, Y. and Kingery, W.D. (1960) Self-diffusion of oxygen in single crystal and polycrystalline aluminum oxide. *J. Chem. Phys.*, **33** (2), 480–486, doi: 10.1063/1.1731170.
131. Prot, D., Le Gall, M., Lesage, B., Huntz, A.M., and Monty, C. (1996) Self-diffusion in α -Al₂O₃. IV. Oxygen grain-boundary self-diffusion in undoped and yttria-doped alumina polycrystals. *Philos. Mag. A*, **73** (4), 935–949, doi: 10.1080/01418619608243697.
132. Garg, S., Krishnamurthy, N., Awasthi, A., and Venkatraman, M. (1996) The O-Ta (oxygen-tantalum) system. *J. Phase Equilib.*, **17** (1), 63–77, doi: 10.1007/BF02648373.
133. Andreoni, W. and Pignedoli, C.A. (2010) Ta₂O₅ polymorphs: structural motifs and dielectric constant from first principles. *Appl. Phys. Lett.*, **96** (6), 062901, doi: 10.1063/1.3308475.
134. Lee, S.H., Kim, J., Kim, S.J., Kim, S., and Park, G.S. (2013) Hidden structural order in orthorhombic Ta₂O₅. *Phys. Rev. Lett.*, **110**, 235 502, doi: 10.1103/PhysRevLett.110.235502.
135. Guo, Y. and Robertson, J. (2014) Oxygen vacancy defects in Ta₂O₅ showing

- long-range atomic re-arrangements. *Appl. Phys. Lett.*, **104** (11), 112906, doi: 10.1063/1.4869553.
- 136.** Kofstad, P. (1962) On the defect structure of Ta₂O₅. *J. Electrochem. Soc.*, **109** (9), 776–781, doi: 10.1149/1.2425554.
- 137.** Elyutin, V., Lenskaya, T., Pavlov, Y.A., and Polyakov, V.P. (1971) *Dokl. Akad. Nauk. SSSR (Tech Phys.)*, **199**, 62. Cited by Freer [35].
- 138.** Stroud, J.E., Tripp, W.C., and Wimmer, J.M. (1974) Defect structure of Ta₂O₅. *J. Am. Ceram. Soc.*, **57** (4), 172–175, doi: 10.1111/j.1151-2916.1974.tb10848.x.
- 139.** Balachandran, U. and Eror, N. (1982) Electrical conductivity in Ta₂O₅. *Mater. Res. Bull.*, **17** (2), 151–160, doi: 10.1016/0025-5408(82)90141-6.
- 140.** Johannessen, O. and Kofstad, P. (1984) The effect of oxide additions on the electrical conductivity of the low-temperature form of Ta₂O₅. *J. Phys. Chem. Solids*, **45** (3), 239–250, doi: 10.1016/0022-3697(84)90027-1.
- 141.** McHale, A. and Tuller, H. (1985) Defects and charge transport in β-Ta₂O₅: I, analysis of the conductivity observed in nominally pure β-Ta₂O₅. *J. Am. Ceram. Soc.*, **68** (12), 646–650, doi: 10.1111/j.1151-2916.1985.tb10118.x.
- 142.** Dellis, J., Carpentier, J., Picot, J., Tellier, P., and Filal, M. (1992) Défauts de structure et transport de charges dans β-Ta₂O₅ en équilibre thermodynamique à haute température. *J. Alloys Compd.*, **189** (2), 157–161, doi: 10.1016/0925-8388(92)90700-J.
- 143.** Lederer, U. (2014) *Sauerstoffdiffusion in Tantaloxid*. Master's thesis. RWTH Aachen University.
- 144.** Bredesen, R. and Kofstad, P. (1988) A reinterpretation of the defect structure of L-Ta₂O₅. *Solid State Ionics*, **27** (1-2), 11–18, doi: 10.1016/0167-2738(88)90484-5.

6

Electrical Transport in Transition Metal Oxides

Franklin J. Wong and Shriram Ramanathan

6.1 Overview

Functional oxides crystallize into a multitude of structures and chemistries. The purpose of this chapter is to develop structural and chemical intuition for understanding the electronic properties, particularly electrical transport, of 3d transition metal oxides (TMOs). The relationship between crystal and electronic structures is discussed conceptually in Section 6.2 and is used throughout the chapter. Three distinct electrical transport mechanisms commonly observed in crystalline oxides are considered in Section 6.3: band transport, small polaron transport, and hopping via defect states. Many of the electronic properties of TMOs are directly related to the electron filling of the transition-metal 3d orbitals. Materials systems with partially occupied transition-metal 3d-like states exhibit strong correlations, and they must be treated differently from uncorrelated materials. We consider three categories of oxides: (i) $3d^0$ and $3d^{10}$ band insulators in Section 6.4, (ii) half-filled Mott insulators in Section 6.5, and (iii) correlated systems with orbital degrees of freedom and exhibit metal–insulator transitions (MITs) in Section 6.6. The transport mechanisms introduced in Section 6.3 are applied to analyze the electrical properties of doped band insulators and half-filled Mott insulators in Sections 6.4 and 6.5, respectively. Section 6.6 exclusively focuses on the general features of MITs observed in some correlated oxides. We do not, however, attempt to explore in great depth the theoretical formalism of the physical models, nor do we intend for this chapter to be an exhaustive experimental review of research on particular materials. The readers are referred to more specialized monographs and articles in the appropriate sections of the chapter.

In this chapter, we do not discuss perovskite oxides, which can occur in a wide range of crystal chemistry combinations and hence warrant a separate work; readers are referred to a comprehensive review by Goodenough and Zhou [1]. We also note another excellent reference monograph, which examines a wide range of oxides as well as theoretical foundations of electronic structure and transport by Tsuda *et al.* [2].

While there is a general understanding of the electronic properties of TMOs, it is often the case that the details of any specific material remain unresolved issues of ongoing research. Covering a large territory, our overview presents a few overarching concepts to describe the vast phenomenology exhibited by TMOs and at the same time highlights some gaps in understanding. Whenever possible, recent work on thin-film oxides will be presented. The growing interest in oxides for electronic and energy conversion applications warrants revisiting some of the outstanding problems in the field.

6.2

Structure of Transition Metal Oxides

This section provides a brief overview of the crystal structures that are covered in this chapter, qualitatively introduces an electronic structure from a bonding approach, and illustrates the relationship between electronic and crystal structures. See Chapter 3 for more detailed discussions.

6.2.1

Crystal Structures of Oxides

It is often instructive to *see* the crystal structures of TMOs in terms of polyhedral cages formed by the oxygen anions with the metal cations residing within the cages. We are concerned with only octahedral oxygen cages and refer to them as $\text{MO}_{6/\beta}$ octahedra, where M is the transition metal, β is the coordination number of oxygen, and $\beta : 6$ is the metal : oxygen stoichiometric ratio. We briefly review the $\text{MO}_{6/\beta}$ octahedra connectivity of different crystal structures and refer the readers to Chapter 3 for images. A complete discussion of structural chemistry and more elaborate illustrations can be found in Refs. [3–5].

The cubic **rock-salt** structure (Figure 3.3a) is common for MO oxides with $\text{MO}_{6/6}$ octahedra, whose corners are connected to one another along the $\langle 100 \rangle$ axes and edges are connected along the $\langle 110 \rangle$ axes. MnO (Section 6.5.2) and NiO (Section 6.5.3) crystallize in the rock-salt structure.

M_2O_3 oxides can crystallize in the rhombohedral **corundum** structure, which we reference with the four-index Miller index notation. $\text{MO}_{6/4}$ octahedra are arranged in the (0001) plane with two opposite faces along the [0001] direction; within the (0001) plane, the M cations form a honeycomb lattice (Figure 3.3c). The stacking is such that, along the *c* axis of corundum, each $\text{MO}_{6/4}$ octahedron shares a face with *just* one other octahedron, forming cation pairs along the *c* axis. Fe_2O_3 (Section 6.5.4), Ti_2O_3 (Section 6.6.3), and V_2O_3 (Section 6.6.4) crystallize in the corundum structure.

MO_2 oxides can crystallize in the tetragonal **rutile** structure (Figure 3.3b). $\text{MO}_{6/3}$ octahedra along the *c* axis share edges, and they share corners on the *ab* plane. SnO_2 (Section 6.4.1), TiO_2 (Section 6.4.2), and VO_2 (Section 6.6.2) crystallize in the rutile structure.

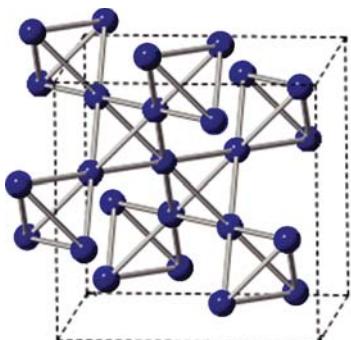


Figure 6.1 The B-site positions of spinel AB_2O_4 . The unit cell is shown by dotted lines.

Spinel oxides have the chemical formula AB_2O_4 , wherein the A cations occupy 1/8 of the tetrahedral interstices and the B cations occupy 1/2 of the interstices of the face-centered cubic oxygen sublattice (Figure 3.3f). An interesting aspect of the AB_2O_4 structure is that if only the B cations are considered, they form vertices of corner-connected tetrahedra (Figure 6.1) – a feature that is relevant in the discussion of Fe_3O_4 (Section 6.6.5).

6.2.2

Bonding and Electronic Structure

The $\text{MO}_{6/\beta}$ octahedron is the building block of not only the crystal structures but also the electronic structures of metal oxides. In 3d TMOs, the relevant electronic bands are formed by the bonding between the 3d orbitals of the transition metals and the 2p orbitals of the oxygen; in the tight-binding approach and by considering the ionicity of metal–oxygen bonds, the *occupied* bonding orbitals have mainly oxygen 2p character, and the antibonding orbitals have mainly transition-metal 3d character. Depending on the cation and its oxidation state, these 3d-like antibonding states can be unfilled, partially filled, or completely filled, and we denote the filling of these states simply by the cation electron configuration; the filling of these 3d-like states largely determines the electronic properties of TMOs.

Octahedral coordination causes the lifting of the fivefold degeneracy of the transition-metal 3d orbitals of a free cation into three degenerate lower energy t_{2g} orbitals ($|xy\rangle$, $|yz\rangle$, and $|xz\rangle$) and two degenerate higher energy e_g orbitals ($|x^2-y^2\rangle$ and $|z^2\rangle$), as shown in Figure 3.13 (see Ref. [6] for details). In a solid, these orbitals are dispersed in energy and form bands. The x , y , z coordinates are chosen such that each direction is pointing at a corner of an octahedron (Figure 6.2a); thus, they should be considered as *local* coordinates of each octahedron and may be distinct from the lattice vectors of the crystal structure. The t_{2g} orbitals engage in π bonding with the oxygen 2p orbitals, and the e_g orbitals take part in σ bonding; examples are shown in Figure 6.2b,c. When two octahedra share an edge, there is a 90° metal–oxygen–metal bond, and the metal cations are brought close enough for one of their 3d orbitals to engage in fairly strong direct bonding that circumvents the oxygen (Figure 6.2d). This direct bonding has

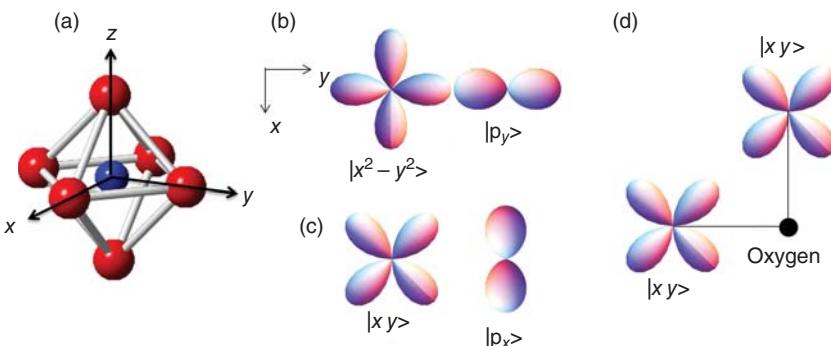


Figure 6.2 (a) Metal–oxygen octahedron with the local x , y , z coordinates defined. Schematics of (b) σ and (c) π bonding between the metal 3d orbitals

and oxygen 2p orbitals, creating e_g and t_{2g} states, respectively, as well as (d) direct metal–metal bonding in a 90° metal–oxygen–metal bond.

huge ramifications for rutile-type VO_2 (Section 6.6.2). Similarly, direct bonding is relevant for face-sharing octahedra, and we leave the discussion to Section 6.6.3 on corundum Ti_2O_3 . Except for rock salt, the other crystal structures do not possess the full symmetry of their metal octahedra building blocks. Crystal structures of lower symmetries result in further lifting of electronic degeneracy; they are second-order effects that will be treated as perturbations.

6.3

Models of Electrical Transport

6.3.1

Band Transport of Carriers

In the band model, the spectrum of electronic eigenstates (electronic structure) can be calculated to an initial approximation by assuming a static lattice of ion cores, that is, the Born–Oppenheimer approximation [7]. An independent-particle one-electron picture is applicable when the valence electrons are sufficiently mobile. The existence of a periodic lattice and electron screening leads to renormalization of the bare electron mass (m_0) to the band electron effective mass (m^*). This renormalized particle that still carries the same unit electric charge is called a quasiparticle; we can treat the dynamics of these quasiparticles in an electric field just as we would for bare electrons, with the only difference being their m^* .

Carrier electrons in a material drift upon the application of an electric field E , resulting in a current density J , related to the electrical conductivity σ of the material: $J = \sigma E$. For carriers in an isotropic band,

$$\sigma = \frac{ne^2\tau}{m^*}, \quad (6.1)$$

where n is the concentration of band electrons contributing to the current, e is the unit electric charge, and τ is the mean free scattering time. Equation 6.1 describes electrons that essentially move freely except for partaking in occasional scattering events on average every time interval τ . Carrier electrons are often described by their drift mobility

$$\mu = \frac{e\tau}{m^*}.$$

If mobile band electrons are responsible for current flow in a material, the carrier μ can be deduced from Hall effect measurements by:

$$\mu_H = \sigma |R_H|,$$

where μ_H is the Hall mobility and R_H is the Hall coefficient, which is negative for electrons and positive for holes for band conduction. We specify the term Hall mobility μ_H , because it does not necessarily have to be equal to μ , as shown in the next section. But for band transport, the two are approximately equal to within a factor close to unity.

6.3.2

Electronic Bandwidth

Because of the prevalence of m^* and electronic bandwidth B in the discussion of transport, an approximate relation between the two is useful. In the nearest neighbor tight-binding model, the dispersion of electronic bands would have the form:

$$E_{TB}(k) = t \sum_{ij} (1 - \cos(k \cdot R_{ij})), \quad (6.2)$$

where $E_{TB}(k)$ are the tight-binding eigenenergies corresponding to the states of momentum k , t is the overlap integral of orbitals between the nearest neighbors, and R_{ij} are the nearest neighbor i,j vectors. For such a band, $B = 2zt$, where z is the number of nearest neighbors of a site. By assuming that the intersite spacing is a , near the band edge ($k \approx 0$), m^* can be approximated as:

$$m^* \approx \frac{\hbar^2}{2ta^2} = \frac{z\hbar^2}{Ba^2}$$

For a typical lattice constant of 3 Å and coordination number of 6, an m^* of m_0 corresponds to a bandwidth of 5.1 eV. An extremely narrow bandwidth of 0.1 eV, comparable to a typical optical phonon energy, corresponds to an m^* of $51m_0$.

6.3.3

Small Polaron Formation

Based on the Born–Oppenheimer approximation, the electronic structure of a crystal can be calculated by assuming a static lattice; however, a lattice is dynamic and can exhibit collective excitations – phonons. If the lattice is polar, one can

envision that the motion of ions may strongly affect the motion of carriers, and phonons can dress electrons, thereby forming polaron states with altered energies. Starting from a continuum model, the change in electron energy ΔE upon it being confined within of lattice polarization cloud of radius r is given by [8]:

$$\Delta E(r) = \frac{\hbar^2}{2m^*r^2} - \frac{e^2}{8\pi\epsilon_0 r} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_{dc}} \right), \quad (6.3)$$

where ϵ_∞ and ϵ_{dc} are the high (optical) frequency and DC dielectric constants, respectively. The first positive term is the reduced kinetic energy of an electron confined within a radius of r compared to being delocalized. The negative second term is the difference in Coulomb energy of an electron being screened by other electrons only as compared to by the lattice as well. The equilibrium polaron state has a radius r_p given by

$$\frac{1}{r_p} = \frac{m^*c^2}{8\pi\epsilon_0\hbar^2} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_{dc}} \right) \quad (6.4)$$

and the corresponding polaron energy $E_p = \Delta E(r_p)$ can be determined similarly. Figure 6.3a shows a carrier electron locally trapped within a potential well formed by lattice distortions; this trapping occurs even if the lattice is perfect, that is, free of defects. The energy well moves with the carrier and is not an inherent part of the lattice. If the trapping energy is high and the polaron size r_p approaches the lattice constant, the polaron is considered *small*. Since the transport of large polarons is similar to that of pure band conduction, in the next section, we focus only on small polaron conduction.

It is transparent from Eq. 6.4 that heavy mass electrons residing in a lattice with large lattice polarizability but small electronic polarizability are susceptible to formation of small polarons (small r_p). It is not uncommon that all of these characteristics can be found concurrently in a single ionic material. Though strictly speaking, Eq. 6.3 breaks down when r_p approaches a lattice constant because a continuum model would not be justified, this simple analysis gives the correct intuition about the features leading to small polaron formation. However, one should not rely on Eqs. 6.3 and 6.4 for accurate E_p and r_p values of small polarons. Analytical expressions to estimate E_p and r_p can be found in Ref. [9]. As a small polaron depends strongly on the energy required to induce local lattice distortions that trap the carrier, today it is common to use *ab initio* calculations to determine

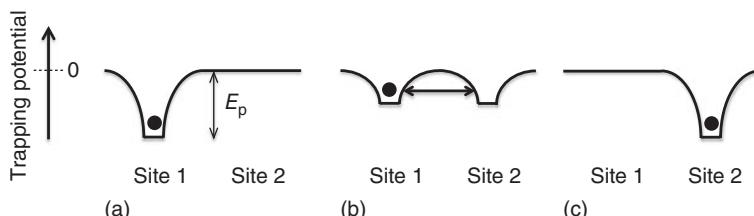


Figure 6.3 Schematic of small polaron hopping. The carrier hops from (a) one site to (c) another via (b) an intermediate state with a thermally activated excited lattice configuration.

E_p rather than rely on extensions of a continuum-based model [10–12]. A small polaron is trapped within a radius r_p that is the smallest characteristic length of the crystal lattice such as a bond length [13].

6.3.4

Small Polaron Transport

The motion of a small polaron, trapped by local lattice distortions, needs to be activated by thermally excited phonons and is considered as hopping transport. For the small polaron to hop from one site to another, the system must be excited to an intermediate lattice state shown in Figure 6.3b, wherein the potential of the site with the electron is coincident with that of a neighboring empty site. In this intermediate state, the carrier can move between both sites. Net hopping occurs when the carrier ends up being trapped in a new site (Figure 6.3c). The trapping potential follows the carrier. By assuming that one optical phonon branch with energy $\hbar\omega_0 = k_B\theta_E$ is responsible for the local lattice distortions, then at $T > 1/2\theta_E$, the carrier μ of small polaron hopping (sph) takes the following functional form:

$$\mu^{\text{sph}}(T) = \mu^{\text{sph}0}(T) \exp\left(-\frac{E_a}{k_B T}\right) \quad (6.5)$$

In the nonadiabatic limit, wherein carrier hopping is much slower than local lattice vibrations, $\mu^{\text{sph}0}(T) \propto 1/T^{3/2}$ and $E_a \approx 1/2E_p$. In the adiabatic limit, wherein carriers hop immediately in response to local lattice vibrations, $\mu^{\text{sph}0}(T) \propto 1/T$ and $E_a = 1/2E_p - t$. Holstein and coworkers performed Hall effect calculations on carriers moving in a two-dimensional hexagonal lattice and concluded that the μ_H exhibits weaker activation behavior with an activation energy of $1/6E_p$ in the non-adiabatic case and $1/6E_p - t (> 0)$ in the adiabatic case [14, 15]. In addition, within this lattice hopping geometry, it was shown that $\mu_H \gg \mu$ if $E_a \gg k_B T$ in the non-adiabatic case, but there are no definite restrictions for the adiabatic case. Furthermore, because of its extremely weak activated behavior, the μ_H of adiabatic hopping can even decrease with increasing temperature if t approaches $1/6E_p$, due to prefactors in the μ_H [15]. Here we only survey the features of small polaron hopping conductivity that will be used to interpret the data presented in Sections 6.4 and 6.5. Detailed discussions of the theory of small polaron hopping can be found in Refs. [9, 16, 17].

Holstein emphasized the possibility of small polaron band conduction at low temperatures [18]. In the low temperature regime, when hopping transport is negligible, the primary effect of temperature is the reduction of small polaron electronic bandwidth; the small polaron bandwidth (B_p) is reduced from the electronic bandwidth in a static lattice (B) [2]:

$$B_p = B \exp\left[-\gamma \coth\left(\frac{\hbar\omega_0}{2k_B T}\right)\right]$$

γ is the electron–phonon coupling constant for small polarons approximated as:

$$\gamma = \frac{1.25}{\pi} \frac{e^2}{\hbar\omega_0 a} \frac{1}{4\pi\varepsilon_0} \left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_{dc}} \right)$$

and a is a characteristic intersite spacing of the crystal. It is argued that in this regime, the μ of small polarons would have similar temperature dependence as large polarons, but would possess a much larger m^* . Rather than increasing with increasing temperature as in the hopping regime, the μ decreases with increasing temperature in the small polaron band transport regime. Because the bandwidth is expected to be very small, small polaron band transport can be thought of as a tunneling process. In principle, the crossover between small polaron band to hopping transport would be manifested as a minimum in temperature-dependent μ . It is shown in Sections 6.4 and 6.5 that signatures of transition from hopping to band transport in real materials systems are often not conclusive and have been a subject of debate in the study of electrical transport in oxides.

6.3.5

Thermopower (Seebeck Coefficient)

Thermopower measurements provide a means to decouple n and μ from the measured conductivity. The Seebeck coefficient for a nondegenerate n-type semiconductor is given by

$$S = -\frac{k_B}{e} \left(A - \ln \frac{n}{N} \right) = -\frac{k_B}{e} \left(A + \frac{E_c - E_f}{k_B T} \right), \quad (6.6)$$

where A is a constant that depends on the carrier scattering mechanism, E_f is the Fermi level, and E_c is the conduction band edge energy. In principle, dual measurements of conductivity and thermopower allow for the extraction of n and hence the temperature dependence of μ without relying on the Hall effect; in practice, quantitative results still depend on assumptions, because the thermopower enables a direct measurement of $E_c - E_f$ and not n . N is the density of states (DOS) parameter that appears in the formula for n :

$$n = N \exp \left(-\frac{E_c - E_f}{k_B T} \right) \quad (6.7)$$

For conventional semiconductors with wide, spin-degenerate bands ($B \gg k_B T$), $N = 2(2\pi m^* k_B T / h^2)^{3/2}$; however, in the literature of oxides, the narrow bandwidth limit ($B_p < k_B T$) is often assumed and hence N is considered as the number density of cations times the band degeneracy. This ambiguity is illustrated in Section 6.5.3.

6.3.6

Hopping Transport via Defect States

Since defects from both intrinsic nonstoichiometry and extrinsic impurities are prevalent in TMOs, we must consider electrical conduction via localized defect states within the band gap of an insulator. Here we assume that the binding energy

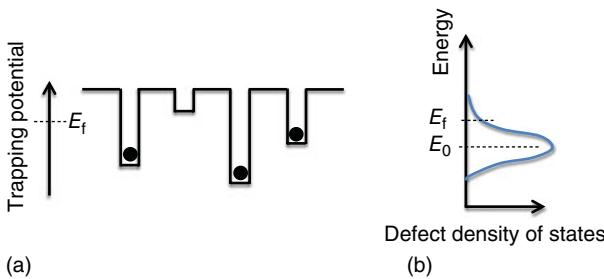


Figure 6.4 (a) Schematic of a spatially random trapping potential landscape created by defects. (b) Hypothetical defect density of states with respect to energy.

of the carrier to the defect states is large enough as compared to $k_B T$ such that band conduction is ignored; this condition is valid (i) at low temperatures, or (ii) when the defects are deep-level, or (iii) if both conditions are satisfied. This section follows the discussion in Ref. [19]. Unlike in the case of small polarons, the trapping potential landscape is static and does not follow the carrier (Figure 6.4a). The problem of hopping transport via localized states in a system with disorder is treated by an Abraham-Miller network of resistances between each pair of localized states in the system. The physical intuition is that the probability of hopping from one site to another site is low (resistance value high) when their spatial separation and/or their energy differences are large; conduction occurring via defect sites can be calculated by summing over all the hopping paths that percolate the sample (see Refs. [17, 19] for details).

We will investigate a particular case of a doped semiconductor, in which some analytical transport formulas can be attained. Suppose that the defects have a DOS and electron filling level within the gap of the semiconductor shown in Figure 6.4b. Let us first consider the case of nearest neighbor hopping (nnh), in which the carriers hop close to the shortest distances possible. nnh will be dominated by the most abundant defect states, that is, states near the maximum DOS (Figure 6.4b). The corresponding energy is labeled as E_0 , and we assume that it is lower than E_f . Since carrier hopping must occur from a filled site to an empty site, and $E_0 < E_f$, nnh is limited primarily by the probability that a final state is unoccupied. Therefore, the temperature-dependent conductivity is given by

$$\sigma^{\text{nnh}}(T) = \sigma^{\text{nnh}0} \exp\left(-\frac{E_f - E_0}{k_B T}\right) \quad (6.8)$$

for $E_f - E_0 \gtrsim 3k_B T$ and the prefactor $\sigma^{\text{nnh}0}$ depends on the density of defects and the localization length of the defect states. Thermally activated nnh with this functional form would be most pronounced if there is a sharp maximum in the defect DOS.

On the other hand, if the temperature is reduced such at $k_B T \ll E_f - E_0$, there would not be many unfilled states near E_0 for nnh. Instead, conduction would be

dominated by the states near E_f , as illustrated in Figure 6.4b to have a considerably lower DOS than at E_0 . The average hopping distance depends on temperature – hence the term variable range hopping (vrh) – and it can be substantially larger than the average distance between a pair of neighboring defect sites. Mott showed that it has the following functional form [20]:

$$\sigma^{\text{vrh}}(T) = \sigma^{\text{vrh}0} \exp\left[-\left(\frac{T_0}{T}\right)^{1/(d+1)}\right], \quad (6.9)$$

where d is the spatial dimensionality of the system with the assumption of a constant DOS in the vicinity of E_f . For three-dimensional systems, it yields the famous Mott-1/4 law.

Simple analytical expressions of hopping conduction via localized defect sites can be attained for the cases of nnh among the states near the maximum DOS (Eq. 6.8) or variable-range hopping among the localized states near E_f (Eq. 6.9). Because of the assumptions implied in Figure 6.4b, these two forms of hopping are distinct, and they dominate at different temperature regimes. However, details depend heavily on the shape of the defect DOS relative to position of E_f – information rarely known for oxides.

Hopping via defect sites is distinct from small polaron hopping, which in principle can occur in a perfect crystal. In practice, carriers and defects are introduced in an insulator (or semiconductor) concurrently. Because they have the same functional forms for conductivity, it is difficult to distinguish the difference between small polaron hopping and neighbor hopping via defect sites solely from electrical transport measurements, but the dramatically weaker temperature dependence of variable-range hopping makes it identifiable. Thermally activated μ is often used as the telltale signature of small polaron hopping.

6.3.7

Bad Metallic Behavior

For band transport in a metallic system, one expects that inelastic carrier scattering increases with temperature, thereby reducing carrier mean free path. However, what if the mean free path becomes so small that it approaches the lattice spacing [21]? This condition is known as the Mott–Ioffe–Regel (MIR) limit. It has been argued that, for metals, the resistivity should increase up to the MIR limit and then saturate at higher temperatures. Resistivity saturation is observed in a variety of heavy fermionic systems [22]. However, there are a variety of materials that appear to violate the MIR criterion, and their resistivity values continue to increase despite the corresponding mean free paths being unphysically small, that is, less than a lattice constant. Because the resistivity increases with temperature, the carriers are not trapped and the transport mechanism is not thermally activated hopping. This unique phenomenology is often coined as bad metallic behavior, and its microscopic origins remain an unresolved subject of active research. It was pointed out that the bad metallic behavior is often found in metallic systems near Mott transitions, suggesting the relevance of electron correlations [22, 23].

6.4

Band Insulators

In this section, we compare two oxide semiconductors that crystallize in rutile structures, SnO_2 and TiO_2 – uncorrelated band insulators. SnO_2 is the only material covered in this chapter that is not a TMO, but we include it for comparative purposes, illustrating how its electronic features contrast with those of TMOs.

6.4.1

SnO_2 : 3d^{10} System

Sn is a sp-valent Group IV element. The lowest conduction band of SnO_2 is formed by antibonding Sn 5s-like states. SnO_2 can be doped n-type, but not p-type, possibly due to the spontaneous creation of compensating nonstoichiometric defects upon moving the Fermi level close to the valence band edge [24]. A recurring theme for many oxide semiconductors, as well as other wide-gap semiconductors in general, is that it is difficult, if even possible, to achieve electrically active doping of both types in the thermodynamic limit [25].

Unlike that previously assumed, Janotti and Van de Walle showed that oxygen vacancies are unlikely to be responsible for room-temperature n-type conductivity in unintentionally doped SnO_2 , and H doping is a possible culprit [24]. Figure 6.5a shows that as the Fermi level reaches roughly above the midgap level, the formation energy of neutral oxygen vacancies becomes lower than that of ionized oxygen vacancies; therefore, oxygen vacancies should be a midgap donor.

N-type impurity doping can be achieved by the substitution of Sn^{4+} with Ta^{5+} , Nb^{5+} , or Sb^{5+} as well as substitution of O^{2-} with F^- . Recent computational studies show that since the charge densities of the conduction band states are dispersed more around the Sn-sites, F-doping of the O-site results in better charge

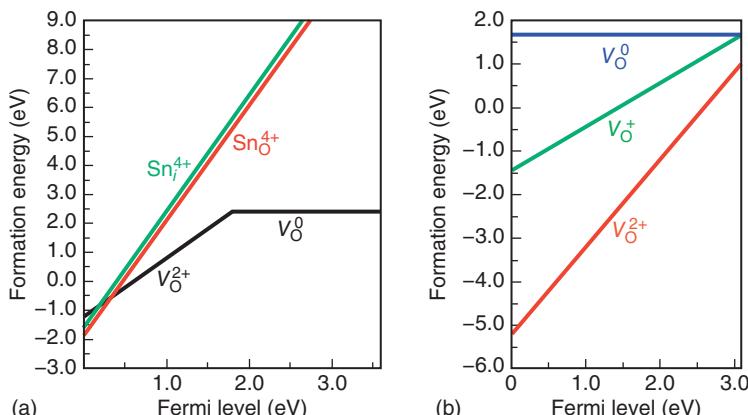


Figure 6.5 Formation energies of native donor defects in (a) SnO_2 and (b) TiO_2 . Fermi level of 0 represents the valence band edge, and the highest value represents the conduction band edge [24].

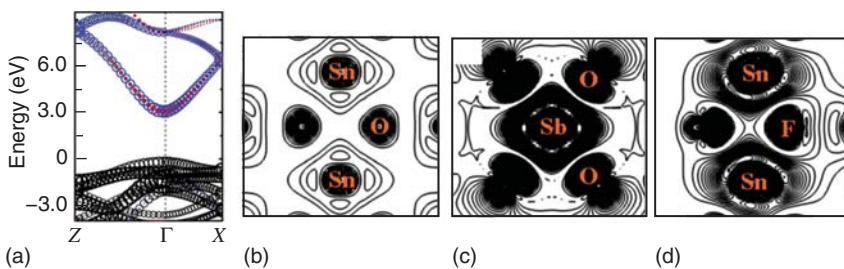


Figure 6.6 (a) Calculated band structure of SnO₂. Charge density contours (in steps of 0.4 electrons per cubic angstrom) near the conduction band minimum of (b) pure, (c) Sb-doped, and (d) F-doped SnO₂ [26].

delocalization than Sb-doping of the Sn-site (Figure 6.6) [26]. From cyclotron resonance studies, the m^* values of SnO₂ were measured to be $0.234m_0$ and $0.299m_0$ parallel and orthogonal to the rutile *c* axis, respectively [27]. The anisotropy arises from the tetragonal symmetry of the rutile structure. The relatively low m^* values tell us that we can safely use band transport to describe the electrical conductivity of n-type SnO₂, and a judicious choice of impurity dopants will hence result in high μ values. A typical electron μ value in SnO₂ at room temperature is $\sim 200\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ [27].

6.4.2

TiO₂: 3d⁰ System

In contrast to SnO₂, the lowest conduction bands of TiO₂ are formed by Ti 3d-like (t_{2g}) states (Section 6.2 and Section 4.5.2.3).¹⁾ Janotti and Van de Walle showed that nominally undoped TiO₂ can be *n*-type with oxygen vacancies (Figure 6.5b) [24]. Because the 3d orbitals of Ti are less dispersed than the 5s orbitals of Sn, the conduction bands of TiO₂ are narrower in energy and the m^* values are greater as compared to those of SnO₂. Yagi *et al.* estimates the conduction band m^* to be $2\text{--}4m_0$ and $10\text{--}16m_0$ parallel and orthogonal to the *c* axis, respectively [28]. However, another report estimates it to be $\sim 25\text{--}125m_0$ [29]. If the larger values are correct, small polarons may be relevant; they would correspond to bandwidths on the order of $k_B T$ at room temperature and optical phonon energies. However, both ranges were deduced by μ_H measurements, which are not the same as μ values for small polaron transport. At room temperature, the highest μ value measured for TiO₂ single crystals is roughly $1\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ [30], which is substantially lower than that of SnO₂. Generally, μ values between 0.1 and $1.0\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ are considered to be the lower limits of the applicability of band transport. It has been argued by Austin and Mott that the true electron m^* in TiO₂ is $\sim 100m_0$, suggesting a likely band narrowing effect due to the formation of small polarons [9], but Hendry *et al.*

1) Here, we focus only on the TiO_{6/3} octahedra building blocks in the discussion of symmetry. The overall tetragonal structure of rutile slightly lifts the degeneracy of the t_{2g} states, which will be discussed in Section 6.6.

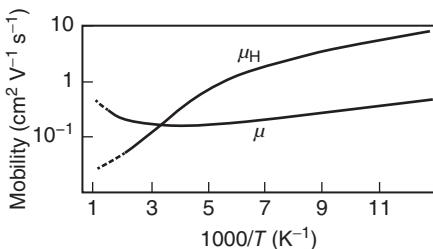


Figure 6.7 Hall and drift electron mobility of electron-doped TiO_2 [9].

claims intermediate-sized polarons as deduced from terahertz spectroscopy measurements from 10 to 300 K [30].

With independent measurements of μ_H and μ , Bogomolov *et al.* found that in general $\mu_H \neq \mu$, and μ increases with temperature above ~ 300 K, with a small activation energy of 0.07 eV along the rutile *c* axis and 0.13 eV in an orthogonal axis (Figure 6.7) [31]. It can be interpreted that the crossover from small polaron band to hopping transport occurs as the temperature is raised above ~ 300 K, and $\mu_H \neq \mu$ is a signature of small polarons. However, $\mu_H < \mu$ in the hopping regime, a characteristic that does not fit the nonadiabatic hopping picture. It may not be easily reconciled even with the adiabatic model, though adiabatic hopping can be consistent with $\mu_H < \mu$ under some conditions [15]. Furthermore, since the time of the measurements in 1968, there have not been any other definitive data on thermally activated μ , despite a lot of work on thermally activated conductivity – for example, Ref. [32]. Small polaron transport in TiO_2 is an issue that needs to be re-examined experimentally. Recent theoretical studies have shown small polaron stability in TiO_2 and anisotropic activation energies [12]. However, the high experimental μ_H values of $\sim 1000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ [28] and μ values of $\sim 50 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ [9] at low temperatures cannot be readily accounted for by small polaron (band) transport. Actually, more convincing signatures of small polarons in TiO_2 are obtained from optical absorption measurements. In an attempt to resolve the debate and seemingly experimental inconsistencies, Janotti *et al.* proposed the coexistence of band electrons and small polarons in TiO_2 , with the former dominating transport measurements and the latter optical measurements [10]. Emin notes the possibility, theoretically, of free carrier and small polaron coexistence in three-dimensional systems [13].

6.5

Half-Filled Mott Insulators

6.5.1

Correlations and the Hubbard

It was pointed out by Mott that the applicability of band theory, which relies on electrons in solids to behave as independent quasiparticles, fails to explain the



Figure 6.8 Schematics of (a) the ground state and (b) an excited state of a system with one electron and one orbital per lattice site described by Hubbard Hamiltonian for $U \gg t > 0$.

electrical properties of many TMOs. As shown in Section 6.2.2, the relevant bands in TMOs have 3d-like character; they are not as dispersed in energy and in space as bonding states of sp-valent materials, but at the same time they are delocalized enough to contribute to bonding. Electron–electron correlation effects play a significant role in more localized 3d-like states, rendering the one-electron picture inadequate. The most elegant approach to introduce correlation is using the Hubbard Hamiltonian [33, 34]:

$$H = - \sum_{i,j,\sigma} t_{ij} (c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma}) + \sum_i U_i n_{i\uparrow} n_{i\downarrow} \quad (6.10)$$

assuming one relevant electron orbital per site, where t_{ij} is the transfer integral between an orbital on site i and that on site j , c_i^\dagger and c_i are electron creation and annihilation operators, respectively, of an orbital on site i , $n_{i\sigma}$ is the electron number operator of site i and spin σ , and U_i is the on-site potential. The first term is the tight-binding Hamiltonian in Eq. 6.2 written in second quantization. The second term encapsulates the physics of correlation effects; there is an energy cost of U for both spin-up and spin-down electrons in the same orbital. For uncorrelated materials such as Si ($U=0$), the band spin degeneracy is 2. By assuming $t_{ij}=t$ for nearest neighbors i,j and 0 otherwise, on average each site contains one electron, and $U \gg t > 0$, it can be shown that the ground state is insulating and has antiparallel nearest neighbor spins (Figure 6.8a). An excited electronic state would be described by the pairing up of two electrons of opposite spins on one site, leaving a “hole” in another site (Figure 6.8b), costing a finite energy $\sim U$, and hence making the system an insulator.

It turns out that this result holds for cases of two orbitals and two electrons per site, three orbitals and three electrons per site, and so on. We categorize systems that meet this requirement as (perfectly) “half-filled” systems, and they often turn out to be *stable* Mott insulators with substantial band gaps. Formally, extra indices will have to be introduced to treat multiorbital systems, and details can be found in Ref. [33]. Though the origin of band gaps is different, half-filled Mott insulators can be treated approximately as semiconductors in which carriers can be introduced through impurities and nonstoichiometry.

The first electron transfer term in Eq. 6.10 is spin-conserving; therefore, charge transport in magnetic oxides is influenced by their magnetic *structures*, as shown in Section 6.5.4. We will not, however, discuss the magnetic *interactions* involved that produce the ordered magnetic *structures* of different materials; the interested reader is referred to the classic treatise by Goodenough [35].

6.5.2

MnO: 3d⁵ System

Mn²⁺ in MnO has five 3d electrons, so MnO is a half-filled system. If U is sufficiently larger than t , there would be a gap separating a filled band of one spin and an empty band of another spin, resulting in a Mott insulator (Section 3.5.3.2). The electron filling is hence more specifically $t_{2g}^3 e_g^2$ with maximized spin angular momentum on each site. The experimental optical gap of MnO is 2.8–3.6 eV [36], and the band gap deduced by electrical transport is 2.6 eV [37]. MnO crystallizes in the rock-salt structure, and below its Néel temperature of 118 K [38], all of the spins of Mn²⁺ cations within a (111) plane are in parallel alignment, while spins of adjacent (111) planes are in antiparallel alignment. Most electrical transport measurements in MnO have been performed above its Néel temperature and hence are not affected by the long-range magnetic order.

Based on the lower activation energy calculated from thermopower [$e\partial S/\partial(1/T)$] as compared to that from resistivity measurements [$2.3k_B\partial \log \rho / \partial(1/T)$] in intrinsically p-type single crystalline MnO, Pai and Honig deduced that the hole μ in MnO is thermally activated with an activation energy of ~ 0.44 eV from 293 to 520 K [39]. Figure 6.9 shows the temperature dependence of hole μ . Crevecoeur and de Wit showed that hole transport is also thermally activated in Li-doped MnO with an activation energy of ~ 0.3 eV over a wider temperature range of 200–1000 K [40]. The estimated hole μ of MnO at room temperature is $\sim 10^{-4}$ cm² V⁻¹ s⁻¹ [39, 40], while μ_H is $\sim 10^{-3}$ cm² V⁻¹ s⁻¹ at about room temperature [40] and has weaker temperature dependence [37]. All of these features are consistent with nonadiabatic small polaron hopping of holes in MnO. The propensity for small polaron formation of holes in MnO has been captured by a recent *ab initio* calculation by Peng and Lany; interestingly, they also predict that on the other hand, electrons can be *free* and exhibit band transport [11]. Experimentally, it appears to be difficult to achieve n-type MnO; nevertheless, there is some evidence of high electron μ values in samples with intrinsically induced carrier electrons at high temperatures, though these values were deduced by Hall data [41].

6.5.3

NiO: 3d⁸ System

NiO is a 3d⁸ ($t_{2g}^6 e_g^2$) system that crystallizes in the rock-salt structure. The filled t_{2g} states do not affect its electrical properties, so they need not be considered. Hence, NiO can be also considered as a half-filled e_g^2 system, and it is a Mott insulator with a sizeable optical gap of ~ 3.5 eV [42] (Section 4.5.3.1). It exhibits antiferromagnetic ordering below its Néel temperature of 525 K, with moments of Ni²⁺ cations within the same (111) plane in parallel and antiparallel alignment between adjacent (111) planes [43]. NiO is a p-type semiconductor, and it can be doped extrinsically by substitutional Li or intrinsically by cation vacancies.

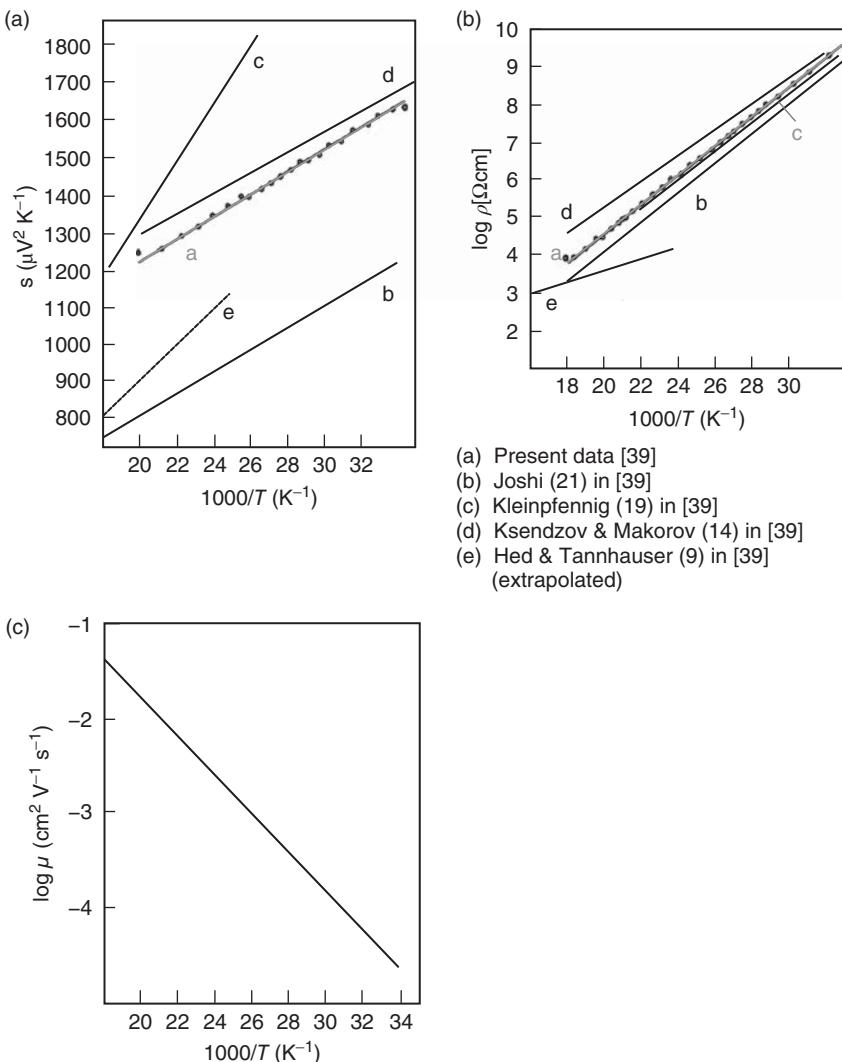


Figure 6.9 The combination of (a) thermopower S and (b) resistivity ρ measurements is used to calculate (c) the temperature dependence of μ in intrinsically p-type MnO. The different letters

corresponding to each curve denote different samples. The mobility curve in (c) was extracted from the data points shown as symbols in (a) and (b). [39].

With dual measurements of conductivity and thermopower, Morin first reported thermally activated μ data in NiO ceramic samples; since the activation energies depend on Li dopant concentration, the effects of defects were uncertain [44]. However, also by electrical and thermopower measurements, Bosman and Crevecoeur found that in slightly Li-doped ($\sim 0.1\%$) NiO ceramics, the hole μ does

not exhibit thermally activated behavior [45]. Figure 6.10 shows the temperature dependence of both μ and μ_H . The two μ curves are calculated based on different assumptions: (i) the DOS parameter N of Eq. 6.7 is a constant equal to two times the Ni cation concentration and (ii) N is proportional to $T^{3/2}$ (see Section 6.3.5). With the second assumption, it can be seen that the hole μ at room temperature is $\sim 4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, and its temperature dependence appears to be described by band transport; the authors argue that this scenario is more plausible [16, 45]. The striking differences between the μ and μ_H values, though do not appear to be consistent with band transport, could be related to the magnetic ordering of NiO, but electron spin is typically not included in general models of the Hall effect. Furthermore, Bosman and van Daal pointed out that although $\mu \neq \mu_H$, mere inequality does not readily suggest small polarons since in most small polaron hopping models with realistic materials parameters, $\mu < \mu_H$, contrary to the observed data [16]. However, based on the same data, Austin and Mott favor the first scenario that assumes the narrowband limit (constant N); they argue that the upturn in μ is indicative of the onset of small polaron hopping being the dominant conduction mechanism, and at lower temperatures, small polaron band transport may be relevant [9]. They claim adiabatic small polaron hopping

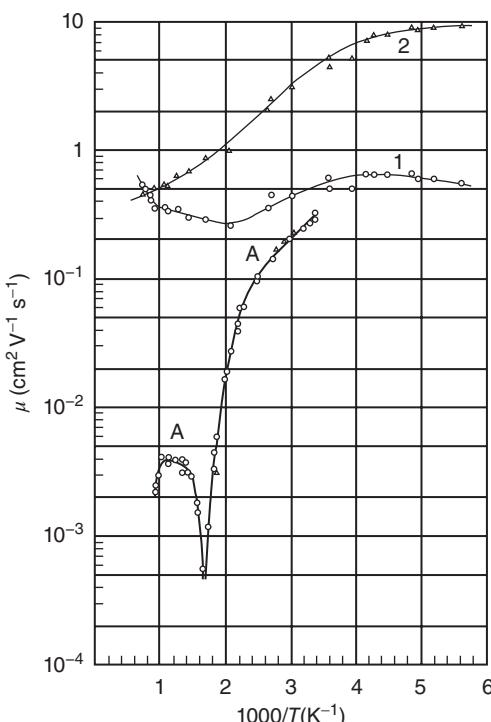


Figure 6.10 Hall (A) and drift (1 and 2) mobilities of 0.09% Li-doped NiO. The drift mobilities were calculated from the same data with two different assumptions: (1) $N = \text{constant}$ and (2) $N \propto T^{3/2}$ (see explanation in the main text) [45].

with an activation energy of $\sim 0.1\text{--}0.2\text{ eV}$ to describe the high-temperature hole transport in NiO [9], and $\mu > \mu_H$ can be possible in the adiabatic limit [13, 15]. It would imply that the e_g bands are wider in NiO than in MnO. Magnetic moments present yet another complication to the analysis of Hall data. The Hall coefficient of p-type NiO undergoes a sign change at $\sim 600\text{ K}$, $\sim 75\text{ K}$ above the Néel temperature, which Bosman and van Daal attribute to the anomalous component of the Hall effect [16]. If small polaron hopping is operative, the sign of the Hall coefficient should depend on the hopping geometry [17], which may change near the onset of magnetic ordering; this possibility has not been addressed in the literature.

Still there are reports of thermally activated μ ($\sim 10^{-5}\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ at room temperature) with sizeable activation energies of $\sim 0.3\text{ eV}$ ($\sim 0.1\text{ eV}$) below (above) the Néel temperature in NiO single crystals intrinsically doped with small amounts of Ni vacancies [46], while others claim rather high hole μ values ($\sim 35\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ at room temperature), which should be fully explained by band transport [47]. The mechanism of hole transport in NiO, perhaps surprisingly, remains an unresolved issue that has not received much recent attention and requires further investigation.

6.5.4

$\alpha\text{-Fe}_2\text{O}_3$: 3d⁵ System

$\alpha\text{-Fe}_2\text{O}_3$ crystallizes in the corundum structure, and it is a 3d⁵ system similar to MnO. It is antiferromagnetic below its Néel temperature of 948 K,²⁾ and its antiferromagnetic state is characterized by parallel spin alignment of Fe³⁺ cations within the (0001) basal planes and alternate alignment between adjacent basal planes. N-type doping can be achieved by substitution of Fe³⁺ with 4+ cations, including Ti, Si, Ge, and Zr [48, 49]. Experimental studies [48] and later computational studies [49] showed that, while Ti impurities have the tendency to trap carriers, the other dopants do not. A carrier electron is of minority spin, and it can travel from one Fe to another Fe cation if the cation moments are in parallel alignment, but not when they are in antiparallel alignment (see Figure 6.11). Therefore, the magnetic structure of Fe₂O₃ is thought to strongly affect electron transport. In fact, computational studies have claimed up to 4 orders of magnitude of difference in conductivity within and orthogonal to the basal plane [50]. While there is some experimental evidence of transport anisotropy in Fe₂O₃ single crystals and textured samples, the results are less drastic and convincing [51, 52]. Regardless, one would intuitively anticipate transport anisotropy, even though it has to be better quantified experimentally.

The combination of resistivity, thermopower, and Hall effect measurements has led to claims and debate over the relevance of small polaron transport in Fe₂O₃

2) There is a small degree of spin canting away from a perfect antiferromagnetic structure, but we will disregard this detail.

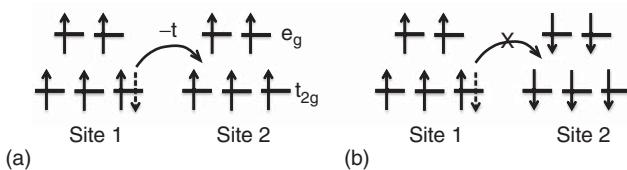


Figure 6.11 A $t_{2g}^3e_g^2$ system doped with electrons, and the extra carrier is shown as a dashed arrow. (a) The carrier can transfer from one site to an adjacent site with parallel spin alignment (b) but not to one with antiparallel spin alignment.

similarly to that in NiO, and Refs. [16, 53] provide a good analysis of the experimental data.

Fe_2O_3 thin films can be grown epitaxially on sapphire, which has the same crystal structure. In a thin-film study of Ti-doped Fe_2O_3 grown on (0001) Al_2O_3 , Zhao *et al.* found that above $\sim 150\text{ K}$, the resistivity ρ appears to fit the functional form $T \exp(E_a/k_B T)$ (see Eq. 6.5), consistent with adiabatic small polaron hopping with $E_a \sim 0.1\text{ eV}$; however, it is not clear whether μ is in fact a thermally activated quantity (Figure 6.12). At lower temperatures, $\rho \sim \exp[(T_0/T)^{1/3}]$ (see Eq. 6.9), consistent with variable-range hopping in a two-dimensional material [54]. Because in principle, spin-preserved transfer of minority spin carrier electrons ought to be confined to the basal plane with parallel cation moment alignment, two-dimensional transport can be rationalized. However, it is unclear when nnh via defects is ever the dominant conduction mechanism due to the limited temperature range measured and its functional form similar to that of the conductivity of small polaron hopping. Furthermore, if the defect DOS does not have a strong peak, a clear regime of nnh is not expected.

More recently, Fe_2O_3 has also received a lot of attention as a candidate oxygen evolution photoelectrode for water splitting [55], an area that has revived interest in the electrical transport behavior of Fe_2O_3 and will likely be an area of significant research growth.

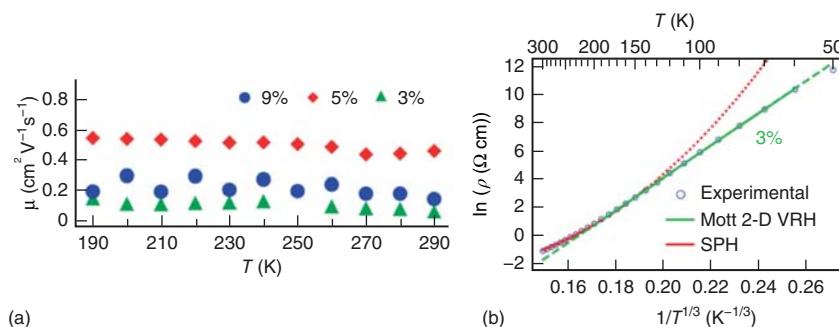


Figure 6.12 (a) Hall mobility values of epitaxial n-type $(\text{Fe}_{1-x}\text{Ti}_x)_2\text{O}_3$ films at different dopant percentages ($x \times 100\%$). (b) Resistivity curve of a $x = 0.03$ doped film fitted

to the functional forms for two-dimensional (2-D) Mott variable-range hopping (vrh) and small polaron hopping (sph) [54].

6.5.5

Summary

In this section as well as in Section 6.4.2, we see that it is often extremely difficult to achieve unambiguous fits of the well-defined transport models presented in Section 6.3 to actual data in TMOs. The materials presented in this section are Mott insulators, meaning that their 3d-like bands are narrow enough such that U is sufficient to stabilize the insulating state. However, the overlap integral may or may not be small enough as compared to E_p , $\hbar\omega_o$, and $k_B T$ to strictly apply some of the narrowband-limit small polaron transport formulas commonly used. More recent theoretical and computational studies have treated the small polaron problem without assuming limiting cases [56, 57], but the possible effects of electron correlations and magnetic interactions have not been fully accounted for. Furthermore, in real materials, point defects introduce carriers, which in turn can be bound to the defect sites; therefore, thermally activated resistivity observed in doped oxides is not a telltale sign of small polaron hopping, and carrier μ measurements that do not depend on the popular Hall effect are required. At low temperatures, hopping via localized defects can dominate carrier transport, making it difficult, if not impossible, to observe true low-temperature small polaron conduction. Our intention is to highlight how transport models have been applied to electrical data, some complicating factors in interpretation, as well as open questions yet to be addressed on the electrical properties of TMOs.

6.6

Temperature-Induced Metal–Insulator Transitions in Oxides

6.6.1

Orbitals and Metal–Insulator Transitions

In 1959, Morin reported the observation of temperature-induced MITs in titanium and vanadium dioxides (Figure 6.13) [58]. For a compound to exhibit a MIT that can be traversed by temperature, both an insulating and a metallic state must be accessible within an energy scale on the order of its transition temperature. Consequently, it is quite rare for a compound to exhibit a temperature-induced MIT, and some common features of those that do are discussed in this section.

We will cover VO_2 , Ti_2O_3 , and V_2O_3 . Octahedrally coordinated V^{4+} in VO_2 and Ti^{3+} in Ti_2O_3 have $3d^1$ (t_{2g}^1) electron configurations. Therefore, in first order, the single electron has a choice of three (or six if assuming spin degeneracy) degenerate t_{2g} orbitals to occupy – hence, VO_2 and Ti_2O_3 have orbital degrees of freedom. In the previous section, we showed examples of stable Mott insulators with sizeable band gaps (>2 eV), which do not have orbital degrees of freedom. However, when the single $3d^1$ electron has a choice of occupying degenerate orbitals with

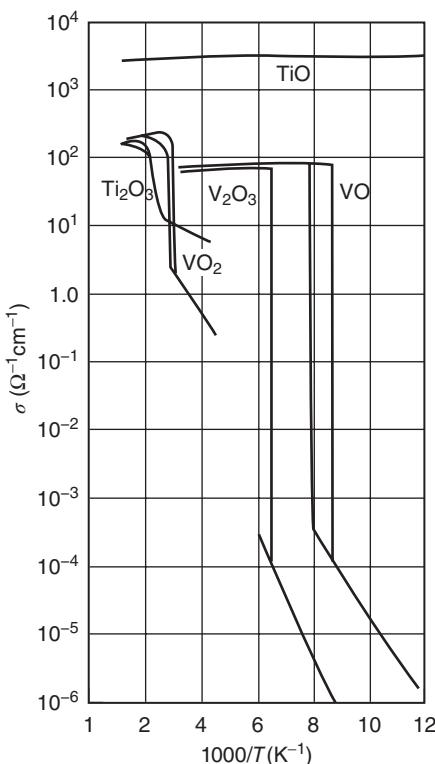


Figure 6.13 Temperature-dependent resistivity of several titanium and vanadium oxides [58].

equal probability, it is extremely difficult to attain an insulating state with Mott physics alone for reasonable values of U . V^{3+} in V_2O_3 has an electron configuration of $3d^2$, and therefore V_2O_3 has orbital degrees of freedom as well. In this section, we provide a brief introduction to the orbital physics involved in MITs; in particular, we will focus on the difference in orbital occupation between the insulating and metallic states of these compounds. We only briefly mention some phenomena observed in VO in Section 6.6.6, but since it typically occurs as a nonstoichiometric compound [59], since 1959, there have been few to no other reports on a temperature-induced MIT in VO. This section ends with an examination of the Verwey transition in Fe_3O_4 . Unlike the previous materials discussed, Fe_3O_4 is a mixed-valent compound with an average Fe valence state of 2.67, and the insulating state entails charge ordering. The insulating states of all of these materials have small band gaps of 0.1–0.7 eV, and there are uncertainties in the measured values. The focus of the section is on the characteristics of MITs in several materials systems, rather than the transport mechanisms operative in their metallic states and in their insulating states individually.

6.6.2

VO₂: 3d¹ System

VO₂ is a paramagnetic metal and adopts a rutile structure above $\sim 68^\circ\text{C}$ [60, 61]. The rutile unit cell is tetragonal; therefore, the electron states of VO₂ do not have the full symmetry of the octahedron building block. Specifically, edge sharing of VO_{6/3} octahedra along the *c* axis results in direct bonding of one of the t_{2g} orbitals as shown in Figure 6.2d. The tetragonal symmetry splits the t_{2g} orbitals into a lower energy a_{1g} state, representing the orbital involved in direct cation bonding along the *c* axis, and doubly degenerate e_g π -bonding states (Figure 6.14a,b). If the orbital splitting is larger than the energy dispersion of the relevant electronic bands, the orbital degree of freedom of this 3d¹ system would vanish, allowing for a pathway to stabilize an insulating state.

VO₂ undergoes a MIT that is concomitant with a structural transition from rutile to monoclinic at 68°C [62]. Below 68°C , VO₂ is a diamagnetic insulator, and its monoclinic structure is characterized by the formation of cation dimer pairs along the *c* axis of the pseudorutile structure. For simplicity, we will use the lattice vectors of the higher symmetry rutile phase in the discussion of all crystallographic axes and planes. There is much debate on whether the temperature-induced MIT of VO₂ is a Peierls transition [63] or a Mott transition [64]. If one were to assume spin degeneracy, the a_{1g} orbital can hold two electrons per formula unit of VO₂. However, the Peierls instability in half-filled electronic systems can result in the formation of V–V dimers (alternating short and long V–V bonds) along the *c* axis, a doubling of the unit cell, and a splitting of the a_{1g} band to two bands, each of which can hold just one electron per VO₂. The splitting of the a_{1g} state as a result of this lattice effect is schematically shown in Figure 6.14c. However, Mott argued that the Peierls effect would not lead to a large enough splitting of the a_{1g} states as compared to the bandwidths of the a_{1g} and e_g π bands to form an insulating state in VO₂, and the correlation *U* is in fact the primary origin of the splitting of the a_{1g} states of different spins (Figure 6.14d) [64].

While the exact nature of the temperature-induced MIT in VO₂ remains a topic of much investigation at present, the lifting of orbital degeneracy is widely thought

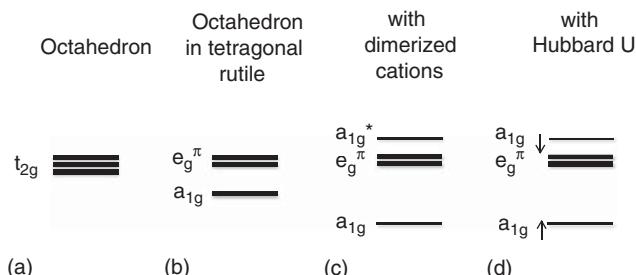


Figure 6.14 Energy splitting of the t_{2g} levels in a rutile structure. Thick (thin) lines represent levels that can hold two (one) electrons per cation.

to be necessary to achieve an insulating ground state. More recent cluster dynamical mean field theory calculations performed by Biermann *et al.* have shown that an insulating state of VO_2 indeed has a substantial degree of orbital polarization: 0.8 electrons in a_{1g} and 0.2 electrons in $e_g \pi$ orbitals. Note that the orbital polarization is not complete as is conceptualized in simplified pictures. For the rutile phase, the authors found that the population is 0.42 in a_{1g} and that of the e_g states is 0.29 each, which gives a rough metric of the effect of the tetragonal symmetry versus the full octahedral symmetry [65]. The precise occupation depends on the type of calculation performed and its assumptions. Haverkort *et al.* also found that orbital polarization is necessary to explain an insulating state of VO_2 and called the temperature-induced MIT an “orbital-assisted collaborative Mott–Peierls transition” [66]. The transition in orbital occupancy across the MIT is experimentally observed by X-ray absorption spectroscopy [67]. Theoretical attempts to model an insulating state without specifically accounting for the orbital degrees of freedom in VO_2 require unreasonably large U values, which often would also lead to the prediction of an insulating rutile phase [66, 68]. Experimentally in its metallic state, rutile VO_2 exhibits bad metallic behavior [69].

High-quality synthesis of epitaxial films [70–72] and nanowires [73] has led to much recent work on VO_2 . Ha *et al.* has shown that a planar electrical device (Figure 6.15a) of an epitaxial VO_2 film grown on sapphire exhibits threshold switching from an insulator to a metallic state upon application of a voltage (Figure 6.15b); furthermore, it was shown that the electrical response of VO_2 occurs on the order of 10 ns (Figure 6.15c,d) [74]. Similar timescales were reported by Chae *et al.*, while Joule heating would take roughly 1 μs in their structure [75]. These results of high-speed switching show that in electrical devices wherein insulating VO_2 is driven out of equilibrium, the metallic state appears to be accessible without a temperature increase, though the exact mechanism for switching is not completely understood. Threshold switching of polycrystalline VO_2 films grown on Si has also been observed (Figure 6.16) [76], and recent studies have demonstrated a \sim 2 ns switching time in a vertical device geometry [77]. Threshold switching in VO_2 has been exploited to address the sneak-path leakage problem in crossbar resistive memory structures [78].

6.6.3

Ti_2O_3 : 3d¹ System

Ti_2O_3 undergoes a temperature-induced MIT at \sim 420 K, which is coincident with a structural transition [79]. Ti_2O_3 is also a 3d¹ system similar to VO_2 but crystallizes into the corundum-type structure. Each cation in the corundum structure is part of a pair of face-sharing octahedra along the c axis. The Ti^{3+} cations participate in direct bonding, leading to the splitting of t_{2g} orbitals to an a_{1g} and two e_g orbitals. The rhombohedral symmetry results in the rehybridization of the d orbitals: $|a_{1g}\rangle = 1/\sqrt{3}(|xy\rangle + |yz\rangle + |zx\rangle)$, $|e_{g(1)}\rangle = 1/\sqrt{2}(|yz\rangle - |zx\rangle)$, and $|e_{g(2)}\rangle = 1/\sqrt{6}(2|xy\rangle - |yz\rangle - |zx\rangle)$, using the x , y , z convention defined in Figure 6.2a [80]. The a_{1g} orbital participates in direct bonding along corundum

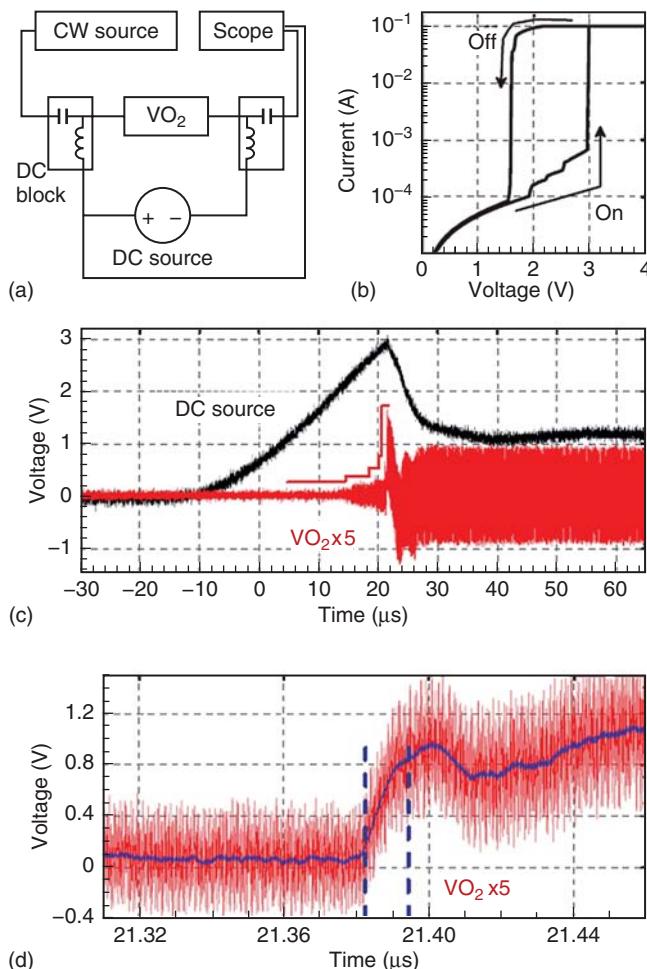


Figure 6.15 (a) Setup of planar electrical switching experiment of a VO₂ film deposited on *c*-plane sapphire. (b) IV trace showing threshold switching. (c) Time-resolved voltage ramp-up and response characteristics of the VO₂ film, magnified in (d) [74].

c axis. Similarly to VO₂, Ti cation pairs dimerize, and the insulating state of Ti₂O₃ is attributed to sufficient lifting of orbital degeneracy by both Peierls dimerization and electron correlation between the a_{1g} and e_g orbitals; both recent and older theoretical studies suggest that the insulating state is described by preferential occupation of the a_{1g} state [81, 82]. The 3d electrons are envisaged being localized within dimer pairs, and they cannot delocalize in the basal plane of the corundum structure. Currently, there is a lack of experimental verification of orbital ordering in the insulating state and a transition in relative orbital population across the MIT.

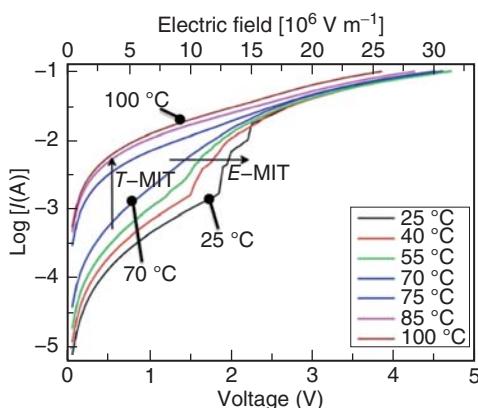


Figure 6.16 IV sweeps of a vertical n -Si/VO₂/Pd structure at different temperatures, showing voltage-induced threshold switching at temperatures below the transition temperature of VO₂ [76].

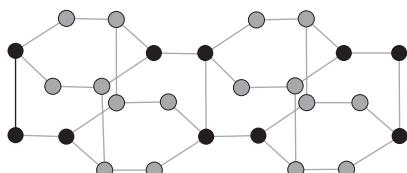


Figure 6.17 Antiferromagnetic structure of V₂O₃ consistent with neutron scattering, wherein the dark sites and light sites represent V³⁺ cations of opposite spin [84].

6.6.4 V₂O₃: 3d² System

Similarly to Ti₂O₃, V₂O₃ crystallizes into a corundum-type structure, but V³⁺ has a 3d² electron filling. V₂O₃ undergoes a phase transition from a high-temperature paramagnetic, metallic, corundum phase to a low-temperature antiferromagnetic, insulating, monoclinic phase at \sim 155 K [83]. For the sake of simplicity, the Miller indices of the corundum structure are used in reference to the low-temperature phase. It was determined that the antiferromagnetic structure of V₂O₃ is described by parallel moment alignment within the (1120) corundum planes while alternating planes are in antiparallel alignment (Figure 6.17) [85]. The relevant a_{1g} and e_g orbitals are as in the case of Ti₂O₃. It has been commonly assumed that the first of the two 3d electrons of each V³⁺–V³⁺ pair would occupy the a_{1g} orbital, forming spin singlets that do not contribute moment to the antiferromagnetic structure [86]. Experimentally, the moment per cation was found to be close to 1 μ_B [85]. Its insulating state would be described by the second electron on each V³⁺ site preferentially occupying one e_g orbital over another, creating an ordered state of e_g orbital occupation [81, 86], and the metallic state would be characterized by more comparable population of both orbitals (in principle, the occupation of the a_{1g} state can be different as well).

Resonant X-ray scattering evidence for orbital ordering in the insulating state consistent with the aforementioned scenario has been shown and claimed by Paolasini *et al.* [87], but the interpretation of the data is in debate. Particularly, in a theoretical study, Mila *et al.* argue about the case of a more complicated ground-state orbital wave function in which, at each site of the V dimer pair, the electron configuration resonates between $a_g^1 e_{g(1)}^1$ and e_g^2 states [84], which they show to be consistent with the known magnetic structure of V_2O_3 and appears to be corroborated by linear dichroism studies on X-ray absorption [88]. Still other calculations suggest the predominance of e_g occupation [89], and an $e_g^2 a_g^0$ ground-state electron configuration has been used explain the optical absorption features of V_2O_3 [90]. These conflicting claims may raise questions of about the orbital ground state of Ti_2O_3 as well.

6.6.5

Fe_3O_4 : Mixed-Valent System

While in all the previous sections only single-valent TMOs were considered, Fe_3O_4 is a mixed-valent system. In a formula unit, a Fe^{3+} cation resides in a tetrahedral site and a Fe^{2+} and a Fe^{3+} cation reside in the two octahedral sites. Fe^{2+} has a $3d^6$ ($t_{2g}^4 e_g^2$) configuration, and the (de)localization of the extra electron determines the electrical transport of Fe_3O_4 ; since this extra electron has a choice of two octahedral sites to occupy, Fe_3O_4 has a charge degree of freedom. Verwey was the first to report a temperature-induced MIT in Fe_3O_4 at ~ 120 K (Figure 6.18) [91] and, with coworkers, proposed a possible ordered arrangement of Fe^{2+} and Fe^{3+} cations, that is, charge (electron) ordering, to account for the low-temperature insulating state [92]. This Verwey charge-ordered structure is described by the alternate ordering of planes of Fe^{2+} and planes of Fe^{3+} along a lattice vector of the spinel structure, say, the c axis (see Figure 6.1 for the B-site

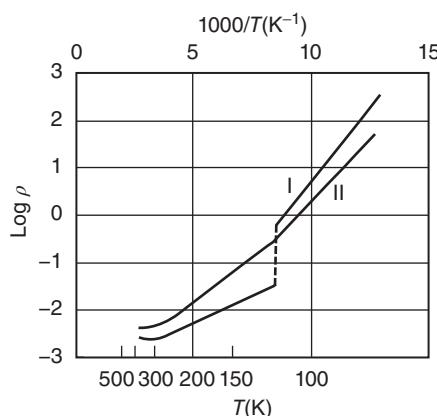


Figure 6.18 Temperature-dependent resistivity of nearly stoichiometric (I) and slightly overoxidized (II) Fe_3O_4 [91].

linked tetrahedra arrangement in the spinel structure). Anderson proposed other charge configurations [93]. He showed that all structures with each tetrahedron consisting of two Fe^{2+} and two Fe^{3+} cations, the “Anderson criterion,” have energies comparable to that of the Verwey structure, which in itself satisfies the Anderson criterion.

Over half a century after a charge-ordered state in Fe_3O_4 was proposed before there was any experimental evidence of such a phenomenon. Synchrotron diffraction studies showed at 90 K that Fe cations of two distinct valences (2.4+ and 2.6+) are identifiable, but the disproportionation of the extra electron is not complete, that is, 2+ and 3+ valences [94]. Because the temperature at which the experiments were performed is fairly near the transition temperature, it is unclear whether the extent of charge disproportionation will increase at even lower temperatures. Perhaps more surprising is that the observed structure does not satisfy the Anderson criterion and exhibits a charge density wave along the spinel *c* axis. The general understanding is that some level of charge ordering appears to be a defining signature of the insulating state of Fe_3O_4 , but other details remain topics of current research. Charge degrees of freedom exist in mixed valent systems, and only recently has the role of orbital physics been investigated in Fe_3O_4 [95, 96].

Above the Verwey transition temperature of 120 K and below the Curie temperature of 858 K, Fe_3O_4 is a ferrimagnetic metal, wherein all of the B-site moments are in parallel alignment. The extra electrons per formula unit assume the minority spin and are delocalized about all the B octahedral sites. In thin films, Fe_3O_4 has been used as an electrode for magnetic tunnel junctions and is shown to be negatively spin polarized [97–99]. Planar two-terminal electrical devices based on epitaxial Fe_3O_4 on MgO substrates (with Au top electrodes, see inset in Figure 6.19a) have suggested the possibility of an electric-field-induced transition from the insulator to the metallic state for devices tested at temperatures below 120 K [100], akin to the VO_2 -based devices tested at room temperature (Figure 6.19a). As the temperature is increased, approaching the metallic state, the switching voltage decreases. In their IV traces, Fursina *et al.* observed hysteretic threshold switching in a continuously rising staircase voltage sweep, but nonhysteretic switching in a pulsed voltage sweep (Figure 6.19b); they concluded that the threshold switching of Fe_3O_4 in itself is a field-induced phenomenon, while the hysteresis is a result of Joule heating [100]. Recent work on Fe_3O_4 thin films grown on ferroelectric substrates have also shown the possibility of electric-field modulation of resistivity owing to ferroelastic strain coupling [101].

6.6.6

Limitations

We must note that, although TMOs that exhibit temperature-induced MITs often possess orbital degrees of freedom, the converse is not true. Some materials that have orbital degrees of freedom turn out to be correlated insulators and others to be correlated metals. A temperature-induced MIT requires that the difference

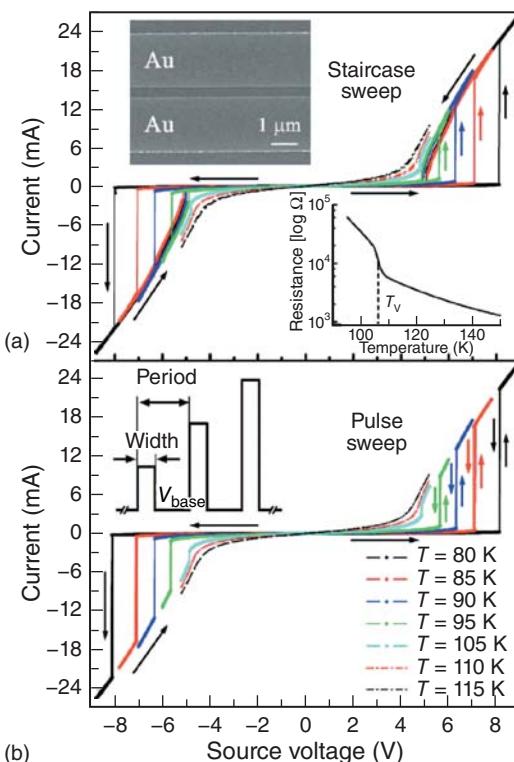


Figure 6.19 Current–voltage sweeps with (a) a staircase and (b) a pulsed voltage source on planar Fe_3O_4 -based devices grown on (001) MgO [100].

in energy between the metallic and insulating states to be on the order of $\sim k_{\text{B}}T$, which is a rather strict criterion.

There is also an interesting case of what appears to be a metallic, perfectly half-filled system: rock-salt VO ($t_{2g}^{3/2}$). As shown in Figure 6.13, Morin observed a MIT in VO, but because there have not been other reports of such a transition, it is possible that the sample was contaminated with another stoichiometric vanadium oxide phase [102]. It is more commonly accepted that VO is metallic in the entire temperature range measured. It has signatures of being near a Mott transition [102], though it cannot be induced by temperature. If a MIT can be achieved in VO by an external field other than temperature, it would be a Mott transition in an integer-valent system that does not feature an orbital degree of freedom.

6.6.7

Summary

This section introduces the concepts of orbital ordering and charge ordering in single- and mixed-valent systems, respectively, to account for the insulating states of correlated TMOs that do not have half-filled 3d shells. Temperature-induced

MITs in these systems are often related to orbital (charge) disorder–order transitions, that is, changes in relative orbital (charge) population; however, the details concerning particular materials systems are often intensely debated and unsettled. Further discussion can be found in Refs. [33, 103–105]. A good grasp of the relationship between the electron structure and the crystal structure is a starting point to delve into the vast literature on this topic. The intuition one can develop typically is based on simplified models assuming complete orbital polarization or complete charge ordering; however, rarely, if ever, have these extreme cases been observed in experiments. Furthermore, often, computational studies rely on adjustable parameters and experimental data require assumed models for fitting, so the interested reader must examine the literature critically and not presuppose that the understanding of these materials is absolute. Currently, there is no single accepted model for a given materials system, much less for all systems that undergo temperature-induced MITs. Regardless, we hope to have identified the relevant concepts and some terminology that are pervasive in the literature of MITs in TMOs and have given a glimpse on how researchers have attempted to exploit MITs in thin-film electrical devices.

References

1. Goodenough, J.B. and Zhou, J.S. (2001) *Struct. Bond.*, **98**, 17.
2. Tsuda, N., Nasu, K., Fujimori, A., and Siratori, K. (2000) *Electronic Conduction in Oxides*, 2nd edn, Springer, Berlin.
3. Muller, U. (2007) *Inorganic Structural Chemistry*, 2nd edn, John Wiley & Sons, Ltd, West Sussex.
4. West, A.R. (1984) *Solid State Chemistry and Its Applications*, John Wiley & Sons, Ltd, Chichester.
5. Rohrer, G.A. (2001) *Structure and Bonding in Crystalline Materials*, Cambridge University Press, Cambridge.
6. Cotton, F.A. (1993) *Chemical Applications of Group Theory*, 3rd edn, Wiley-Interscience, New Jersey.
7. Yu, P.Y. and Cardona, M. (2010) *Fundamentals of Semiconductors: Physics and Materials Properties*, 4th edn, Springer, Berlin.
8. Madelung, O. (1981) *Introduction to Solid-State Theory*, Springer-Verlag, Berlin.
9. Austin, I.G. and Mott, N.F. (1969) *Adv. Phys.*, **18**, 41.
10. Janotti, A., Franchini, C., Varley, J.B., Kresse, G., and Van de Walle, C.G. (2013) *Physica Status Solidi-RRL*, **7**, 199.
11. Peng, H.W. and Lany, S. (2012) *Phys. Rev. B*, **85** (5), 201202.
12. Deskins, N.A. and Dupuis, M. (2007) *Phys. Rev. B*, **75**, 195212.
13. Emin, D. (2013) *Polarons*, Cambridge University Press, Cambridge.
14. Friedman, L. and Holstein, T. (1963) *Ann. Phys.*, **21**, 494.
15. Emin, D. and Holstein, T. (1969) *Ann. Phys.*, **53**, 439.
16. Bosman, A.J. and Vandaal, H.J. (1970) *Adv. Phys.*, **19**, 1.
17. Bottger, H. and Bryksin, V.V. (1985) *Hopping Conduction in Solids*, Akademie-Verlag, Berlin.
18. Holstein, T. (1959) *Ann. Phys.*, **8**, 343.
19. Gantmakher, V.F. (2005) *Electrons and Disorder in Solids*, Oxford University Press, Oxford.
20. Mott, N.F. and Davis, E.A. (1979) *Electronic Processes in Non-crystalline Materials*, Oxford University Press, Oxford.
21. Ioffe, A.F. and Regel, A.R. (1960) *Prog. Semicond.*, **4**, 237.

22. Hussey, N.E., Takenaka, K., and Takagi, H. (2004) *Philos. Mag.*, **84**, 2847.
23. Gunnarsson, O., Calandra, M., and Han, J.E. (2003) *Rev. Mod. Phys.*, **75**, 1085.
24. Janotti, A. and Van de Walle, C.G. (2011) *Phys. Status Solidi B*, **248**, 799.
25. Robertson, J. and Clark, S.J. (2011) *Phys. Rev. B*, **83**, 075205.
26. Li, C., Li, J.B., Li, S.S., Xia, J.B., and Wei, S.H. (2012) *Appl. Phys. Lett.*, **100**, 262109.
27. Button, K.J., Fonstad, C.G., and Dreybrodt, W. (1971) *Phys. Rev. B*, **4**, 4539.
28. Yagi, E., Hasiguti, R.R., and Aono, M. (1996) *Phys. Rev. B*, **54**, 7945.
29. Breckenridge, R.G. and Hosler, W.R. (1953) *Phys. Rev.*, **91**, 793.
30. Hendry, E., Wang, F., Shan, J., Heinz, T.F., and Bonn, M. (2004) *Phys. Rev. B*, **69**, 081101.
31. Bogomolov, V.N., Kudinov, E.K., and Firsov, Y.A. (1968) *Sov. Phys.-Solid State*, **9**, 2502.
32. Yildiz, A., Iacomi, F., and Mardare, D. (2010) *J. Appl. Phys.*, **108**, 083701.
33. Khomskii, D.I. (2010) *Basic Aspects of the Quantum Theory of Solids*, Cambridge University Press, Cambridge.
34. Stohr, J. and Siegmann, H.C. (2006) *Magnetism: From Fundamentals to Nanoscale Dynamics*, Springer, Berlin.
35. Goodenough, J.B. (1963) *Magnetism and the Chemical Bond*, Interscience Publishers, New York.
36. Chou, H.H. and Fan, H.Y. (1974) *Phys. Rev. B*, **10**, 901.
37. Ali, M., Fridman, M., Denayer, M., and Nagels, P. (1968) *Phys. Status Solidi*, **28**, 193.
38. Jagadeesh, M.S. and Seehra, M.S. (1981) *Phys. Rev. B*, **23**, 1185.
39. Pai, M. and Honig, J.M. (1981) *J. Solid State Chem.*, **40**, 59.
40. Crevecoeur, C. and Dewit, H.J. (1970) *J. Phys. Chem. Solids*, **31**, 783.
41. Dewit, H.J. and Crevecoeur, C. (1967) *Phys. Lett. A*, **25**, 393.
42. Powell, R.J. and Spicer, W.E. (1970) *Phys. Rev. B*, **2**, 2182.
43. Roth, W.L. and Slack, G.A. (1960) *J. Appl. Phys.*, **31**, S352.
44. Morin, F.J. (1954) *Phys. Rev.*, **93**, 1199.
45. Bosman, A.J. and Crevecoeur, C. (1966) *Phys. Rev.*, **144**, 763.
46. Keem, J.E., Honig, J.M., and Vanzandt, L.L. (1978) *Philos. Mag. B*, **37**, 537.
47. Spear, W.E. and Tannhauser, D.S. (1973) *Phys. Rev. B*, **7**, 831.
48. Shinar, R. and Kennedy, J.H. (1982) *Sol. Energy Mater.*, **6**, 323.
49. Liao, P.L., Toroker, M.C., and Carter, E.A. (2011) *Nano Lett.*, **11**, 1775.
50. Iordanova, N., Dupuis, M., and Rosso, K.M. (2005) *J. Chem. Phys.*, **122** (10), 144305.
51. Nakau, T. (1960) *J. Phys. Soc. Jpn.*, **15**, 727.
52. Benjelloun, D., Bonnet, J.P., Doumerc, J.P., Launay, J.C., Onillon, M., and Hagenmuller, P. (1984) *Mater. Chem. Phys.*, **10**, 503.
53. van Daal, H.J. and Bosman, A.J. (1967) *Phys. Rev.*, **158**, 736.
54. Zhao, B., Kaspar, T.C., Droubay, T.C., McCloy, J., Bowden, M.E., Shuthanandan, V., Heald, S.M., and Chambers, S.A. (2011) *Phys. Rev. B*, **84**, 245325.
55. Sivula, K., Le Formal, F., and Gratzel, M. (2011) *ChemSusChem*, **4**, 432.
56. Fratini, S. and Ciuchi, S. (2003) *Phys. Rev. Lett.*, **91**, 256403.
57. Ortmann, F., Bechstedt, F., and Hannewald, K. (2009) *Phys. Rev. B*, **79**, 235206.
58. Morin, F.J. (1959) *Phys. Rev. Lett.*, **3**, 34.
59. Banus, M.D., Reed, T.B., and Strauss, A.J. (1972) *Phys. Rev. B*, **5**, 2775.
60. Andersson, G. (1956) *Acta Chem. Scand.*, **10**, 623.
61. Kawakubo, T. and Nakagawa, T. (1964) *J. Phys. Soc. Jpn.*, **19**, 517.
62. Goodenough, J.B. (1960) *Phys. Rev.*, **117**, 1442.
63. Goodenough, J.B. (1971) *J. Solid State Chem.*, **3**, 490.
64. Zylbersztejn, A. and Mott, N.F. (1975) *Phys. Rev. B*, **11**, 4383.
65. Biermann, S., Poteryaev, A., Lichtenstein, A.I., and Georges, A. (2005) *Phys. Rev. Lett.*, **94**, 026404.
66. Haverkort, M.W., Hu, Z., Tanaka, A., Reichelt, W., Streletsov, S.V., Korotin, M.A., Anisimov, V.I., Hsieh, H.H., Lin,

- H.J., Chen, C.T., Khomskii, D.I., and Tjeng, L.H. (2005) *Phys. Rev. Lett.*, **95**, 196404.
67. Koethe, T.C., Hu, Z., Haverkort, M.W., Schussler-Langeheine, C., Venturini, F., Brookes, N.B., Tjernberg, O., Reichelt, W., Hsieh, H.H., Lin, H.J., Chen, C.T., and Tjeng, L.H. (2006) *Phys. Rev. Lett.*, **97**, 116402.
68. Weber, C., O'Regan, D.D., Hine, N.D.M., Payne, M.C., Kotliar, G., and Littlewood, P.B. (2012) *Phys. Rev. Lett.*, **108** (5), 256402.
69. Qazilbash, M.M., Burch, K.S., Whisler, D., Shrekenhamer, D., Chae, B.G., Kim, H.T., and Basov, D.N. (2006) *Phys. Rev. B*, **74**, 205118.
70. Wu, Z.P., Yamamoto, S., Miyashita, A., Zhang, Z.J., Narumi, K., and Naramoto, H. (1998) *J. Phys.-Condens. Matter*, **10**, L765.
71. Wong, F.J., Zhou, Y., and Ramanathan, S. (2013) *J. Cryst. Growth*, **364**, 74.
72. Zhou, Y. and Ramanathan, S. (2012) *J. Appl. Phys.*, **112**, 074114.
73. Cao, J., Gu, Y., Fan, W., Chen, L.Q., Ogletree, D.F., Chen, K., Tamura, N., Kunz, M., Barrett, C., Seidel, J., and Wu, J. (2010) *Nano Lett.*, **10**, 2667.
74. Ha, S.D., Zhou, Y., Fisher, C.J., Ramanathan, S., and Treadway, J.P. (2013) *J. Appl. Phys.*, **113**, 184501.
75. Chae, B.G., Kim, H.T., Youn, D.H., and Kang, K.Y. (2005) *Physica B*, **369**, 76.
76. Ko, C. and Ramanathan, S. (2008) *Appl. Phys. Lett.*, **93**, 252101.
77. Zhou, Y., Chen, X.N., Ko, C.H., Yang, Z., Mouli, C., and Ramanathan, S. (2013) *IEEE Electron Device Lett.*, **34**, 220.
78. Son, M., Liu, X., Sadaf, S.M., Lee, D., Park, S., Lee, W., Kim, S., Park, J., Shin, J., Jung, S., Ham, M.H., and Hwang, H. (2012) *IEEE Electron Device Lett.*, **33**, 718.
79. Rao, C.N.R., Loehman, R.E., and Honig, J.M. (1968) *Phys. Lett. A*, **27**, 271.
80. Tanaka, A. (2004) *J. Phys. Soc. Jpn.*, **73**, 152.
81. Zeiger, H.J. (1975) *Phys. Rev. B*, **11**, 5132.
82. Poteryaev, A.I., Lichtenstein, A.I., and Kotliar, G. (2004) *Phys. Rev. Lett.*, **93**, 086401.
83. Dernier, P.D. and Marezio, M. (1970) *Phys. Rev. B*, **2**, 3771.
84. Mila, F., Shiina, R., Zhang, F.C., Joshi, A., Ma, M., Anisimov, V., and Rice, T.M. (2000) *Phys. Rev. Lett.*, **85**, 1714.
85. Moon, R.M. (1970) *Phys. Rev. Lett.*, **25**, 527.
86. Castellani, C., Natoli, C.R., and Ranninger, J. (1978) *Phys. Rev. B*, **18**, 4945.
87. Paolasini, L., Vettier, C., de Bergevin, F., Yakhou, F., Mannix, D., Stunault, A., Neubeck, W., Altarelli, M., Fabrizio, M., Metcalf, P.A., and Hoing, J.M. (1999) *Phys. Rev. Lett.*, **82**, 4719.
88. Park, J.H., Tjeng, L.H., Tanaka, A., Allen, J.W., Chen, C.T., Metcalf, P., Honig, J.M., de Groot, F.M.F., and Sawatzky, G.A. (2000) *Phys. Rev. B*, **61**, 11506.
89. Ezhov, S.Y., Anisimov, V.I., Khomskii, D.I., and Sawatzky, G.A. (1999) *Phys. Rev. Lett.*, **83**, 4136.
90. Stewart, M.K., Brownstead, D., Wang, S., West, K.G., Ramirez, J.G., Qazilbash, M.M., Perkins, N.B., Schuller, I.K., and Basov, D.N. (2012) *Phys. Rev. B*, **85**, 205113.
91. Verwey, E.J.W. (1939) *Nature*, **144**, 327.
92. Verwey, E.J., Haayman, P.W., and Romeijn, F.C. (1947) *J. Chem. Phys.*, **15**, 181.
93. Anderson, P.W. (1956) *Phys. Rev.*, **102**, 1008.
94. Wright, J.P., Attfield, J.P., and Radaelli, P.G. (2001) *Phys. Rev. Lett.*, **87**, 266401.
95. Zhou, F. and Ceder, G. (2010) *Phys. Rev. B*, **81**, 205113.
96. Senn, M.S., Wright, J.P., and Attfield, J.P. (2012) *Nature*, **481**, 173.
97. Hu, G. and Suzuki, Y. (2002) *Phys. Rev. Lett.*, **89**, 276601.
98. Nelson-Cheeseman, B.B., Chopdekar, R.V., Alldredge, L.M.B., Bettinger, J.S., Arenholz, E., and Suzuki, Y. (2007) *Phys. Rev. B*, **76**, 220410.
99. Nelson-Cheeseman, B.B., Wong, F.J., Chopdekar, R.V., Arenholz, E., and

- Suzuki, Y. (2010) *Phys. Rev. B*, **81**, 214421.
100. Fursina, A.A., Sofin, R.G.S., Shvets, I.V., and Natelson, D. (2009) *Phys. Rev. B*, **79**, 245131.
101. Liu, M., Hoffman, J., Wang, J., Zhang, J.X., Nelson-Cheeseman, B., and Bhattacharya, A. (2013) *Sci. Rep.*, **3**, 1876.
102. Rivadulla, F., Fernandez-Rossier, J., Garcia-Hernandez, M., Lopez-Quintela, M.A., Rivas, J., and Goodenough, J.B. (2007) *Phys. Rev. B*, **76**, 205110.
103. Tokura, Y. and Nagaosa, N. (2000) *Science*, **288**, 462.
104. Khomskii, D.I. (2005) *Phys. Scr.*, **72**, CC8.
105. Goodenough, J.B. and Zhou, J.S. (2007) *J. Mater. Chem.*, **17**, 2394.

7**Quantum Point Contact Conduction**

Jan van Ruitenbeek, Monica Morales Masis, and Enrique Miranda

7.1**Introduction**

The high-conductance state of resistive switching (RS) systems is often attributed to the formation of a conductive filament (CF) with metallic conductance. When the size of the weakest connection in this CF is of atomic size, the conductance is expected to be the result of transport through a limited number of quantized conductance channels. In recent years, several reports have appeared describing the observation of quantized conductance in the high-conductance state of RS devices [1–7]. However, stepwise variation in the conductance may simply arise from the discrete atomic structure of the contact. Such contacts would have a limited number of quantum channels for conductance, but this does not need to appear directly as quantization of the conductance. In this chapter, we discuss the properties of CFs in the high-conductance state of resistive switches in the framework of the Landauer theory of conductance, and we critically evaluate the evidence for the formation of atomic-size constrictions and quantization of the conductance.

7.2**Conductance Quantization in Metallic Nanowires**

Historically, conductance quantization was first discovered in two-dimensional (2D) electron gas devices similar to the ones used for studies of the quantum Hall effect. Also pedagogically, it is attractive to start from 2D systems. Nowadays, many interesting new 2D systems have been discovered for which conductance quantization has been observed, including graphene [8], topological insulators [9], and interfaces between oxide insulators such as LaAlO_3 and SrTiO_3 [10] or $\text{ZnO}/\text{Mg}_x\text{Zn}_{1-x}\text{O}$ [11]. However, the original discovery was made for a 2D electron gas formed at the interface between two semiconductors, GaAs and AlGaAs [12, 13]. The advantage of this 2D electron gas is that it behaves to a large extent as a free-electron gas, for which the confining potential can be adjusted by the electrostatic field of a gate electrode. This is how Van Wees and coworkers succeeded

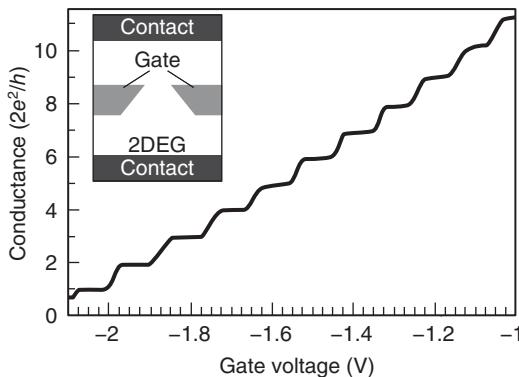


Figure 7.1 Discovery of conductance quantization: conductance as a function of the width of a point contact in a two-dimensional electron gas, formed by means

of control of the gate potential. The inset shows the arrangement of the contacts and the gate electrodes. (Reproduced with permission from Ref. [12].)

in forming a constriction, or point contact, in the 2D electron gas. As illustrated in the inset of Figure 7.1, two metallic gate electrodes were deposited on top of the GaAs–AlGaAs heterostructure. By connecting these electrodes to a sufficiently large negative potential with respect to that of the 2D electron gas, the electrochemical potential for the electrons below the gate electrodes is raised so high that the electrons effectively see the region below the gates as a barrier. The barrier extends beyond the edges of the metallic gate structures according to the profile of the electrostatic potential, and this potential drops gradually back to zero for large distances from the edge of the gates. Thus, the width of the region over which the electron gas is completely expelled depends on the strength of the gate potential. By exploiting this dependence of the effective width of the gate region on the gate potential, it is possible to continuously adjust the width of the contact connecting the top half of the 2D electron gas to the bottom half (see inset in Figure 7.1). The main panel in Figure 7.1 shows the variation observed when measuring the conductance between the top and bottom halves of the 2D electron gas while changing the strength of the gate potential. The remarkable discovery reported in Refs. [12, 13] is that the conductance does not vary smoothly with the width of the point contact, that is, with the gate potential, but it increases stepwise when the point contact smoothly becomes wider. Moreover, the height of the steps is given by a fundamental unit, the quantum of conductance, $G_0 = 2e^2/h$, where e is the electron charge and h is Planck's constant. In terms of resistance, this unit is $R_Q = 1/G_0 \approx 12.9\text{ k}\Omega$. For an extensive review, we refer to [14, 15].

The theoretical framework for description of conductance quantization is provided by the Landauer–Büttiker theory of conductance [16, 17]. The starting point is a free and noninteracting electron gas (which we can consider as 2D or 3D, depending on the problem at hand) for which all scattering lengths are much larger than the system size. In other words, we assume that the mean free path between

collisions with defects is much longer than the width and length of the conductor and that relaxation of energy and that of spin occur only in the contacts. Such a system is called a *ballistic* conductor, because the electrons pass through unhindered once they are launched from the contacts. We will also consider the contacts to the outside world to serve as ideal electron reservoirs. Ideal electron reservoirs perfectly absorb all electrons that fall onto them and emit electrons according to the equilibrium Fermi distribution with the local temperature T and electrochemical potential μ . Let us first consider a noninteracting electron gas in 2D within a confining potential that defines straight parallel walls. The energy dispersion relation for the electrons in such a wire of width W can be written as

$$E(k_x, n_y) = \frac{\hbar^2 k_x^2}{2m} + \frac{\hbar^2}{2m} \left(\frac{n_y \pi}{W} \right)^2 \quad (7.1)$$

with m being the electron mass. The first term is the kinetic energy for motion parallel to the wire (here taken along the x -direction), while the second term gives the quantized kinetic energy levels in the perpendicular direction y . Next, we consider a wire for which the width changes, but only very gradually, on the scale of the Fermi wavelength. In this case, we can use the adiabatic approximation, which allows us to replace the constant W in Eq. (7.1) simply by a position-dependent width $W(x)$. For a given subband index n_y , the second term in Eq. (7.1) can be viewed as a smoothly varying potential, which has a maximum at the constriction of a point contact, as illustrated in Figure 7.2.

Indeed, far from the point contact, the electron has a fixed amount of kinetic energy along the direction of the wire, but as it approaches the point contact confinement claims a growing amount of the available energy for the motion in the perpendicular direction. This energy is deduced from the forward motion kinetic energy and slows the electron down just as a true potential barrier would do. Each subband has a maximum in the effective barrier given by $E_{b,n} = \frac{1}{2m}(n\pi\hbar/W_{\min})^2$, where W_{\min} is the minimal cross section of the wire at the point contact. Electrons

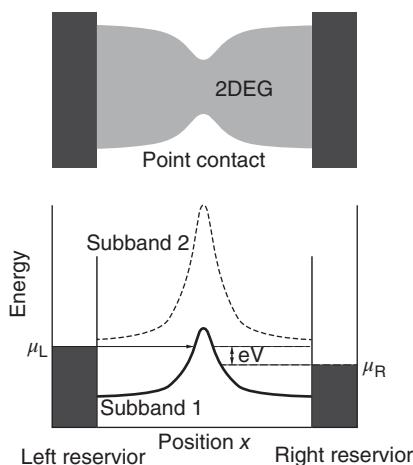


Figure 7.2 Schematic representation of an adiabatic constriction in 2D (top) and the resulting profile of the bottom of the subbands (below).

moving from one of the reservoirs with energy much smaller than $E_{b,n}$ will be reflected at the effective barrier, while electrons with energy much higher than $E_{b,n}$ will be transmitted to the other reservoir. Near the top of the barrier there is a finite tunneling probability, which can be obtained from the shape of the barrier in the Wentzel–Kramers–Brillouin (WKB) approximation. Similarly, for energies just above the barrier, there will be a finite probability of backscattering. In total, the effect of the barrier can be described in terms of an energy-dependent transmission probability $T(E)$, which is 0 for $E \ll E_{b,n}$ and smoothly rises around the top of the barrier to $T(E) = 1$ for $E \gg E_{b,n}$.

The total current resulting from an applied voltage bias can be obtained by summing over all states traveling left and right [14, 15, 18],

$$I = \frac{2e}{h} \int_{-\infty}^{\infty} dE \ T(E)(f_L(E) - f_R(E)), \quad (7.2)$$

where $T(E) = \sum_n T_n(E)$ is now the sum of the transmissions at a given energy E for all subbands, each forming a separate conductance channel. The Fermi–Dirac distribution functions in Eq. (7.2) describe the equilibrium occupation of states in the left and right contacts, $f_{L,R}(E) = 1/(\exp((E - \mu_{L,R})/kT) + 1)$. When the variation of $T(E)$ with energy is weak on the scales of the temperature, kT , and the applied voltage difference, $\mu_L - \mu_R = eV$, the integral in Eq. (7.2) becomes straightforward and we obtain

$$G = \frac{dI}{dV} = \frac{2e^2}{h} \sum_n T_n = G_0 \sum_n T_n \quad (7.3)$$

This is the celebrated Landauer–Büttiker result, which shows that conductance is a measure of the transmission probability for the electrons. The properties of such a ballistic system differ from a classical conductor in many important ways. The occupation of states in the scattering region, that is, in the region between the contacts to the outside world, is not described by the Fermi–Dirac distribution but is strongly out of equilibrium. Although it can be shown that the voltage drop occurs over a small range near the point contact, there is no dissipation in the scattering region. The power $P = IV$ is completely converted into kinetic energy of the electrons, which is finally deposited in the form of heat in the outside contacts.

Let us discuss the application of these ideas, successively relaxing some of the approximations that we made. First, an extension to finite temperature and to finite bias simply requires performing the full integral in Eq. (7.2). If the voltage bias window $(\mu_L - \mu_R)/e$ overlaps with one or more of the subband barriers $E_{b,n}$, the current–voltage dependence becomes strongly nonlinear [19]. For a single channel, in the limit that eV is much larger than the distance between the Fermi energy and the top of the subband barrier, the conductance approaches $0.5G_0$ because only the top half of the electrons within the window eV are transmitted. Second, the extension from 2D to 3D is obvious and only requires proper labeling of the conductance channels, by taking into account the two perpendicular dimensions. Third, the assumption of an adiabatic constriction is a good approximation for a point contact in a 2D electron gas, with the point contact defined

by an electrostatic gate potential. In general, notably, when we are considering 3D point contacts of atomic size, the approximation will break down. However, it can be shown [14, 15, 18] that, for any shape of contact, one can define a basis of independent eigenchannels such that Eqs. (7.2) and (7.3) remain valid. For such channels, the simple plane-wave picture is no longer appropriate, and the transmission probabilities can assume any value between 0 and 1.

In the 2D experiment illustrated in Figure 7.1, the width of the point contact is about half the Fermi wavelength at the moment when the conductance reaches the first plateau, at $G = 1G_0$. In order to observe conductance quantization for metallic systems, one needs to produce point contacts with a size comparable to the Fermi wavelength, which is only ~ 0.5 nm for most metals. In other words, the contact size needs to be of the order of a single atom. Several experimental techniques have been developed for reaching such a small size. The most widely used techniques are mechanically controllable break junctions and scanning tunneling microscopes [18]. In both cases, a nanocontact between two metallic electrodes can be made by means of a piezoelectric actuator. These contacts can be repeatedly made and broken, and in the process of making and breaking, the contact often passes through a stage of just a single atom connecting the two electrodes.

Two other techniques are also relevant in the context of the formation of CFs in ionic conductors. One of these is the use of electromigration [20–22], where one exploits the force exerted by the electron current on the atoms in a conductor. Here, one starts from a microfabricated metallic wire, which is then reduced to an atomic contact by passing a large current through the wire. The current is usually pulsed while monitoring the changes in the conductance. The other technique that needs to be mentioned here is the use of wet electrochemistry [23–28]. Here, two metal working electrodes are produced with a small gap between them, by mechanical means or by microfabrication. This assembly is immersed in a solution of a metal salt, the electrolyte, and is placed in contact with a counter electrode. In many cases, there is also a reference electrode for monitoring and controlling the electrochemical potential of the electrolyte. By controlling the potential difference between the working electrodes and the counter electrode, it is possible to switch between deposition and dissolution of the metal, thus allowing repeated cycles of making and breaking a metallic contact between the two working electrodes.

Figure 7.3a shows examples of break junction experiments for three different metals. The curves show the evolution of the conductance as a function of the control voltage across the piezoelement by which the contact is gradually broken. For each of the three metals, Au, Al, and Pt, we observe an overall decrease in the conductance, with sections of small and smooth variation interrupted by sharp steps. The steps have been demonstrated, through many experiments, to be the result of sudden atomic structural rearrangements [18]. Slightly less precise, one could say that the contact decreases in size in an atom-by-atom manner. The three curves are just random examples, because each time a new contact is formed, the atomic structure of the contacts will be different. In order to extract the generic properties for the atomic contacts, one usually collects data of many contact-breaking curves to construct a conductance histogram. Figure 7.3b shows conductance histograms

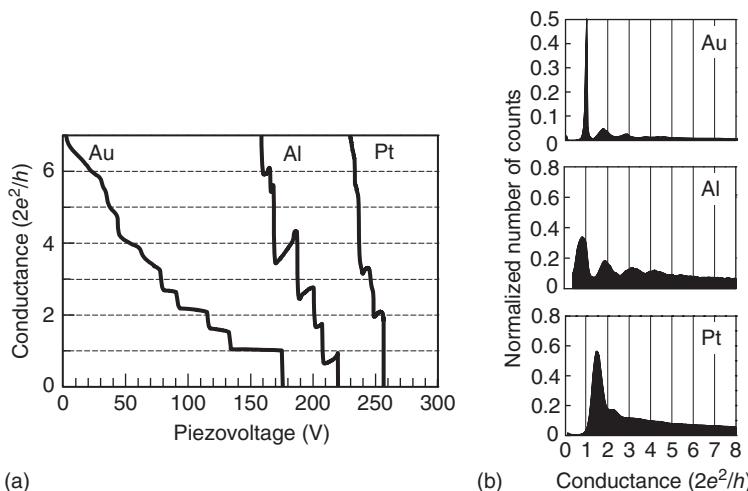


Figure 7.3 (a) Examples of conductance curves for Au, Al, and Pt taken while breaking contacts at 4.2 K. (b) Conductance histograms formed from hundreds of breaking curves.

for Au, Al, and Pt obtained from typically 1000 breaking curves in each case. We observe a characteristic series of peaks in the histogram and note that there is a sharp peak at $1G_0$ in the histogram for Au. None of the other peaks for Au, or those for Al and Pt, coincide convincingly with integer quantum numbers. This conveys the important message that these peaks, and therefore the conductance at the plateaus in the conductance curves shown in Figure 7.3a, cannot be directly attributed to conductance quantization. The difficulty in the interpretation of the experiments is that the step size due to a change in the contact by a single atom is of similar magnitude as a conductance quantum.

This raises the question of what role true conductance quantization plays in atomic contacts. This problem has been addressed from the theoretical point of view and through a number of targeted experiments, which have been reviewed in Ref. [18]. The picture that has emerged is that contacts of atomic size at low temperature and low voltage bias are very well described by the Landauer–Büttiker theory outlined earlier. On the other hand, a model of a smooth adiabatic contact is not very accurate, and one generally needs to include several conductance channels in the description, with each having a transmission probability between 0 and 1. This implies that the conductance is determined by channels that are due to lateral size quantization, but the total conductance given by Eq. (7.3) in general is *not* an integer multiple of the conductance quantum. Only for free-electron-like monovalent metals a description by plane-wave electrons and adiabatic constriction provides a reasonable approximation, as was observed for Na contacts [29], and can also be seen in the sharp peak at $1G_0$ for Au as shown in Figure 7.3b.

It is possible to analyze the transmission properties for atomic contacts well beyond the mere observation of their conductance. A single-atom contact for Al

can have a total conductance of $0.9G_0$, and this could, in principle, be the result of a single conductance channel. In practice, we find that Al atomic contacts have a conductance resulting from a combination of the channels. The most powerful experimental tool for demonstrating this is the measurement and analysis of the superconducting subgap structure [30, 18]. In order to explain how this can be useful, imagine we have a superconducting metallic wire and we use one of the break junction techniques to break it and form a vacuum tunnel junction. The tunnel current will be linear at high voltage bias in accordance with a fixed tunnel resistance, but at low bias, the current will drop to zero due to the gap in the density of states for the superconductors. The drop occurs for $eV = 2\Delta$, where Δ is the superconducting energy gap, and a factor of 2 enters because we have a superconductor on both sides. Now, when we move the electrodes closer together, but still remaining in the tunneling regime, we can see a small additional step in the current within the gap region at $eV = \frac{1}{2}(2\Delta)$. By reducing the vacuum gap further, this additional step grows and a new step becomes visible at $eV = \frac{1}{3}(2\Delta)$, and so on. The additional current steps are the result of multiple Andreev reflections. The theory that includes Andreev reflection to all powers is capable of describing the tunnel current curves with only one free parameter: the transmission probability T [31], see the inset in Figure 7.4. This property can be exploited for analyzing the transmission also in the contact regime when more than just a single channel is involved, [32]. Since the shape of the I - V curves varies strongly depending on T , one can use this as a fitting parameter and combine several curves for a set of $\{T_1, T_2, T_3, \dots\}$, as shown in Figure 7.4. As long as the numbers of channels involved are not too large (up to about 6) a well-defined answer is obtained on the number of conductance channels and their

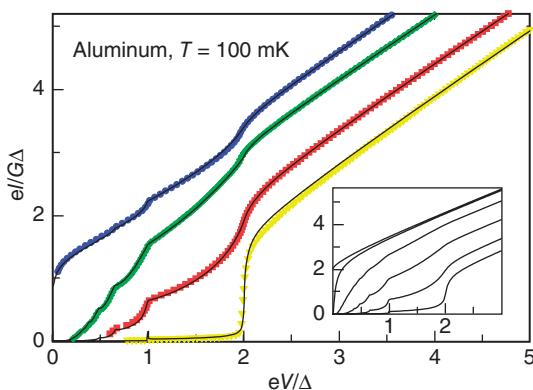


Figure 7.4 Superconducting subgap structure for Al atomic-size contacts. Al has a critical temperature of about 1 K, and measurements were performed at 0.1 K. Inset: Theoretical I - V curves for a junction formed

between two superconductors by a single conductance channel, for transmission probabilities ranging from 1 (top) to 0.1 (bottom curve). (Reproduced with permission from Ref. [32].)

transmission probabilities. Extending this approach to several metals and combining it with tight-binding calculations has revealed a simple rule [30]: the number of conductance channels for a single atom is determined by the number of valence orbitals.

For metals that are not superconductors, other techniques that provide partial information on the conductance channels have been developed [18]. The analysis of shot noise is the most important of these techniques [33, 34]. Shot noise is the intrinsic noise in the current due to the discrete character of the electron charge. In the limit of low temperatures ($k_{\text{B}}T \ll eV$), the current noise power is given by (see reviews [35, 18]),

$$S_I = 2eI \frac{\sum_n T_n(1 - T_n)}{\sum_n T_n} = 2eIF \quad (7.4)$$

In this expression, the Fano factor F depends only on the transmission probabilities. In the simple case of a single conductance channel, one can read from Eq. (7.4) that the noise disappears when the transmission probability approaches unity. For two channels, the combination of the measured conductance G and noise power S_I allows uniquely to determine the set $\{T_1, T_2\}$. For larger numbers of channels the information is incomplete but still permits drawing boundaries on the possible solutions. In the following sections we review the experimental evidence for quantized conductance in CFs formed after switching of a resistive layer. Since the mechanisms are somewhat different, we first discuss cation-based RS materials, which have highly mobile cations as part of the pure lattice structure, and where these mobile cations are responsible for the action of switching. In a subsequent section, we review the experiments for binary oxides, where CFs are often attributed to the formation of oxygen vacancy paths or to the migration of metal atoms from the metallic electrode materials.

7.3

Conductance Quantization in Electrochemical Metallization Cells

In electrochemical metallization (ECM) cells, also called CBRAM, the *on* and *off* states are defined by a reversible electrochemical formation of a metallic CF embedded in a resistive material. The metallic CF acts as a conductive bridge or “short” between the electrodes of the device. For more details on this type of memories, see Chapters 17 and 18.

In order to obtain an ECM cell, one needs a resistive material that allows the transport of mobile metal ions, a solid electrolyte, for example, sandwiched between two electrodes: an active electrode working as the reservoir of the metal ions (e.g., M = Ag, Cu) and an inactive electrode made of an inert metal (e.g., Pt, W). The switching process is commonly described as follows: at the initial stage,

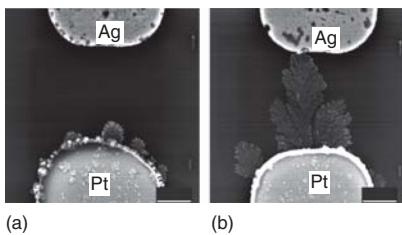
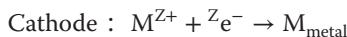


Figure 7.5 Scanning electron microscopy (SEM) images for a Pt/H₂O/Ag cell showing (a) the *off* state of the device with shorter and smaller Ag dendrites and (b) the same cell in the *on* state with longer and larger Ag dendrites obtained by applying a positive voltage on the Ag electrode. (Reproduced from Ref. [36].)

the device has a high resistance, defined as the *off* state. When a sufficiently large voltage is applied between the electrodes, and when negative polarity is applied at the inert electrode (i.e., the cathode), the metal ions (M^{Z+}) migrate toward the cathode and are reduced to metal atoms,



On the anode side (the active electrode), the metal atoms are oxidized to metal ions, which dissolve into the electrolyte, maintaining the ion concentration in the electrolyte constant,



As long as the voltage is applied, ion migration in the electrolyte continues, and further reduction of the ions leads to the formation of a CF, which grows toward the anode. A conducting bridge between the electrodes is then formed, switching the device to the low-resistance state (LRS) or *on* state (see Figure 7.5). When the voltage polarity is reversed, the metal atoms forming the CF will be dissolved back into the solid electrolyte and deposited at the active electrode [37]. The filamentary nature of the metallic deposit is ascribed to a nonuniform deposition of the metal at the surface of the cathode. Any protrusion in the deposited metal enhances the electric field at its tip, causing a faster growth of the wire at one specific point. A more controlled way to form the metallic nanowires is the use of an atomic force microscopy (AFM) or scanning tunneling microscopy (STM) tip as the cathode.

For the purpose of discussion, we distinguish two types of ECM switches: those with the cathode in full contact with the solid electrolyte [1, 38, 2, 39] and those where a tunneling gap is maintained between the electrolyte and the cathode [40–42]. The *off* state of the devices will be defined by the resistance of the solid electrolyte layer and the contact resistance (electrode/electrolyte interface) in the former case or the resistance of the solid electrolyte layer plus the tunneling gap resistance in the latter experimental configuration. The *on* state, in both cases, is usually defined by the resistance of the CF shorting the device. When, during the process of forming or breaking the metallic nanowire, a contact is formed by just one atom, or just a few atoms, one may observe step-like conductance traces. These conductance steps would then be expected to appear either at the last stages of breaking of a wire or at the first stages of contact.

7.3.1

Current–Voltage Characteristics and Definition of Initial Device Resistance

In order to make a distinction between device resistance and quantized resistance, we will start by analyzing the initial states of switching, that is, just before the first CF forms. In the off state or preswitching state, the I - V characteristics of ECM switches of mixed ion and electron conductors (e.g., solid electrolytes) are well described by the following expression [39]:

$$I(V) = K\sigma_0 \frac{k_B T}{e} \left(e^{eV/k_B T} - 1 \right), \quad (7.5)$$

where k_B is the Boltzmann's constant, σ_0 is the conductivity of the layer at zero bias, and K is a constant with the dimension of length representing the geometry of the contact. This equation describes the response observed in the conductivity of the electronic semiconductor as a result of self-doping in the applied electric field due to the mobile ionic charges. In the limit close to zero bias, that is, $eV \ll k_B T$, Eq. (7.5) reduces to

$$I = K\sigma_0 V \quad (7.6)$$

The geometry of the electrodes and sample enters through the constant K , which is the only fitting parameter if the conductivity (σ_0) of the layer is known. In writing Eq. (7.6), it is also assumed that one of the electrodes (the anode) maintains the chemical potential for the metal ions constant, that is, it provides the reservoir for the ions. The I - V curves presented in Figure 7.6a have been fitted using Eq. (7.5). The I - V curves were measured for an Ag_2S device with an Ag bottom electrode and a Pt-coated AFM tip as the top electrode. For the fitting, it is assumed that contact point of the tip has approximately a spherical shape, $K = 2\pi a$, and $\sigma_0 = 7.8 \times 10^{-2} \Omega^{-1} \text{ m}^{-1}$ for Ag_2S . With the fitting, the calculated AFM tip radius is 12 nm, which agrees well with the AFM tip radius observed by scanning electron microscopy (SEM) [39].

The exponential increase in the electronic current at forward bias observed in Figure 7.6a (negative polarity at the cathode) is explained by the initial accumulation of positive ions close to the nanoscale contact. The accumulating positive ions act as n-type donors and therefore locally increase the conductance of the device. This occurs before switching is observed.

In the case of gap-type switches, when a large tunneling gap is formed, the resistance is dominated by vacuum tunneling (i.e., resistance in the order of gigaohms) and the I - V curve is nearly linear. Notice that for gap-type switches, the control of the top electrode–sample distance requires biasing well below or well above the band gap of the mixed conductor. A total device resistance of the order of megaohms or less, as employed by Ref. [41], already indicates an electrode in mechanical contact with the sample [42]. If one performs the same experiment as illustrated in Figure 7.6a and continues to apply higher bias voltages, the device starts presenting hysteresis, that is, electrical switching behavior (Figure 7.6b). Here is where we expect the CFs to start forming (forward bias) and dissolving

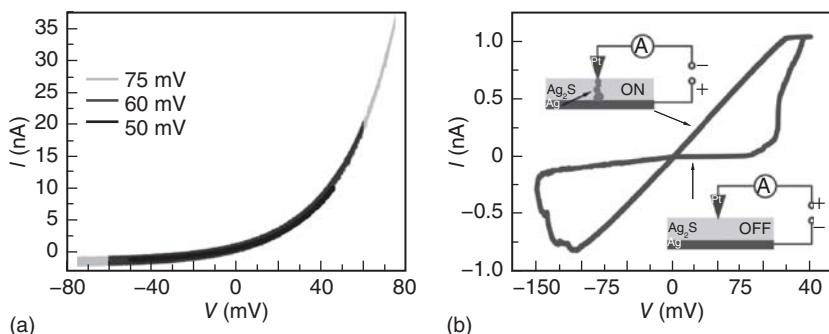


Figure 7.6 Preswitching (a) and full bipolar switching (b) current–voltage characteristics of a $\text{Ag}/\text{Ag}_2\text{S}/\text{Pt}$ (nanocontact) system. In the full bipolar switching case, the transition

from off to on state and back is explained by the formation and dissolution of a metallic nanowire. (Reference [39].)

(reverse bias). We will now discuss whether these CFs are metallic and to what extent the conductance in the CFs is quantized.

7.3.2

Stepwise Conductance Changes in Metallic Filaments

Step-like conductance traces have been observed in ECM memories during the formation or dissolution of the CFs, by using well-defined device architectures [43, 1, 2] or under He atmosphere at low temperatures [44]. Figure 7.7 presents traces in conductance measured by Tappertzhofen *et al.* in $\text{Ag}/\text{AgI}/\text{Pt}$ devices while increasing the bias current. The authors note that the observation of conductance steps during formation occurs only if sufficiently low bias currents are used to switch the device to the *on* state (in this specific case, from 0 to $20 \mu\text{A}$). For higher currents, the switching occurs so fast that the conductance steps cannot be resolved and the traces look continuous rather than step-like. The steps in the conductance traces shown in Figure 7.7 are attributed to the discrete atomic character of the contact. Although many authors also use the term conductance quantization in this context, we want to caution against this: the steps will be visible when the contact grows in an atom-by-atom manner, independent of whether the conductance channels per se are forming one by one. In fact, as mentioned before, only for simple *s*-electron metals, one expects the transmission probability for the conductance channels to be close to integer values [18].

Although the guidelines illustrated in Figure 7.7a are very suggestive, they are also somewhat misleading. There is a fitting parameter that takes care of the series resistance by the more bulky part of the CF. While this allows to position the multiples of conductance quanta close to the observed plateaus, one needs to remember that the levels are always within $0.5G_0$ of any level, so that the apparent match is easily mistaken for conductance quantization. This is also apparent in Figure 7.7b, where the results of measurements of 65 cells have been collected into a

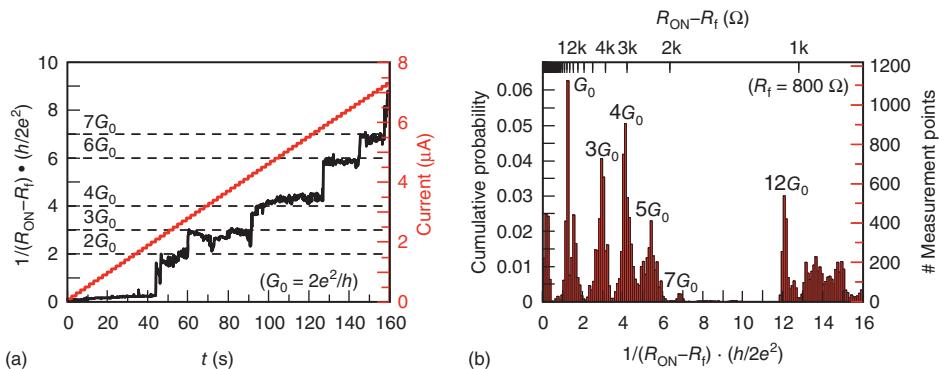


Figure 7.7 (a) Steps in conductance measured in an Ag/AgI/Pt cell during a current sweep. (b) Conductance histogram constructed from cumulative data of 65 devices. (Reproduced from Ref. [43].)

histogram. It is remarkable to observe how well the results reproduce, as evidenced by the sharp peaks in the histograms. Yet, one also observes that the deviations in the peak positions from the ideal integer conductance values are significant, up to the maximum of $0.5G_0$. This indicates reproducibility in terms of atomic structure, rather than strict conductance quantization. Yet, conductance should still be carried by only a few channels (each with a finite transmission probability), and in this sense, the observations are in the quantum (rather than quantized) regime of conductance. The important conclusion from the observations is that the structure of the CF has a limited number of recurrent configurations. These are most likely discrete atomic structures, of just a few atoms at the narrowest section of the wire. Possibly, the matrix provided by the solid electrolyte imposes constraints on the accessible atomic configurations.

Also, during the breaking of metallic CF, two types of conductance breaking traces have been observed: continuous and step-like traces. Wagenaar *et al.* reported conductance traces for Ag_2S defined by a step-like pattern, mainly at the last stages of breaking of the wire, and a second type of traces showing only a slow and continuous decay of the conductance [38]. Once a CF dominates the conductance, as evidenced by a linear current–voltage relation, the CF can be broken by applying a negative voltage. During this process, the traces of conductance with a step-like pattern indicate that, when the CF is at its weakest point, it becomes reduced to only a few bridging atoms. In this way, quantum properties of the conductance of the silver CF will show up [18]. Figure 7.8 shows atomic conductance steps observed in conductance traces recorded as a function of the time. In Figure 7.8a, one observes that the conductance spontaneously decays almost linearly until arriving at about $5G_0$, when steps close to $1G_0$ in height start to become visible. Very pronounced plateaus and steps of approximately $1G_0$ can also be seen in the plots in Figure 7.8b, with a very long plateau of 0.4s in the middle trace. The upper trace also shows two-level fluctuations with amplitude close to $1G_0$, which is frequently observed in atomic-scale contacts [18]. Such

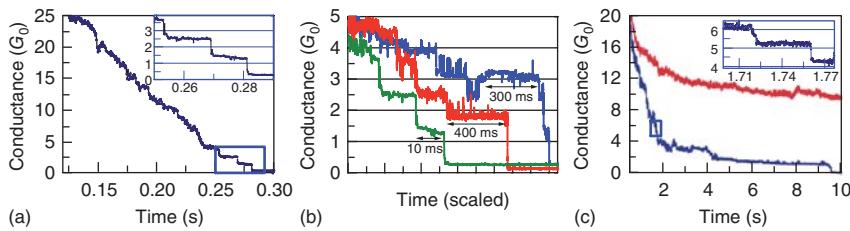


Figure 7.8 Steps in conductance traces of Ag_2S devices become visible when breaking the conductive path at low bias voltages (<100 mV). (a) Breaking trace with three clear conductance steps of approximately $1G_0$ observed at the last stages of breaking. The inset shows a zoom of the steps. (b) Three breaking traces with atomic

conductance steps having different lengths in time. The upper trace shows two-level fluctuations that are typical for atomic-size contacts and are attributed to single atoms oscillating near the contact. (c) Continuous-conductance traces and mixed-conductance traces observed in the same devices. (Reproduced from Ref. [38].)

two-level fluctuations are attributed to two configurations at nearly the same energy, allowing the atomic structure to randomly switch between the two states by thermal activation.

Apart from these stepwise variations, there is a second mechanism active, characterized by a slower and nearly continuous decrease in the conductance. This is illustrated by the upper trace in Figure 7.8c. This behavior was attributed to a modification of the local lattice structure, giving rise to a region of increased conductance. This modification is induced by the electric field and the resulting increased concentration of silver in the region of switching. After applying a negative voltage, silver diffuses back to the Ag bottom reservoir and the lattice slowly relaxes to its equilibrium structure. Occasionally, the two processes were observed together, as illustrated in the blue trace in Figure 7.8c: around $6G_0$, atomic conductance steps are visible, while later on, there is a continuous decrease from 2 to $1G_0$. In terms of the two mechanisms described, this may be explained as being due to two parallel conductance paths. Indeed, parallel RS paths in Ag_2S switching devices were first reported by Xu *et al.* [45] based on the analysis of transmission electron micrographs.

In summary, in the *off* state ($G \ll G_0$) of the mixed ionic conductors, the device should be characterized by a nonlinear I - V curve (see Eq. (7.5)), and in the *on* state, once $G > G_0$, the device is characterized by a linear I - V curve. In this case, the current flows through the metallic CF, leading to ohmic behavior. Usually, in the *on* state, the conductance of the device can reach values from 10 to 100 times G_0 , within few micro- or nanoseconds (depending on the amplitude of the bias voltage), due to the large number of atoms participating in the switching. Continuously sweeping the voltage from negative to positive polarity gives rise to hysteretic I - V characteristics, as observed in Figure 7.8b. It is important to emphasize here that the condition of $G > G_0$ is not sufficient to prove the presence of a metallic nanowire shorting the device, nor is a measurement of a conductance close to 1

G_0 or a resistance close to $12.7\text{ k}\Omega$ direct evidence of the presence of a point contact. The total conductance of the device depends on the size of the contact and the conductivity of the solid electrolyte ($G = \sigma A/L$, A : area of the contact and L : length or thickness of the layer). The conductivity (σ) of the solid electrolyte can be strongly modified after several switching cycles, due to either lattice deformation, or high doping, or just accumulation of metal clusters in the resistive layer [38, 45].

At room temperature, the parallel conductance through the locally modified Ag_2S lattice structure decays more slowly than the atomic contacts formed by the CF. Therefore, the stable conductance steps at integers of G_0 , achieved by applying specific bias [1], are most likely due to the modified layer properties and not due to the formation of atomic contacts [38]. The work of Terabe *et al.* in 2005 [1] triggered a lot of further research and raised the possibility of multilevel conductance switching. The fabrication of “atomic switches” showing highly stable conductance steps close to multiples of G_0 is now widely found in literature. However, in view of the two mechanisms involved in switching Ag_2S junctions, the conductance values shown in [1] were likely set by the particular choice of the switching voltages. In general, a pure quantization effect is usually masked by parallel conductance paths through the electrolyte matrix and through several partially transmitting channels in the CF. Nevertheless, we have now ample evidence that CFs can be formed that make contact by one atom, or just a few atoms, and this atomic discreteness is clearly visible as conductance steps.

7.4

Filamentary Conduction and Quantization Effects in Binary Oxides

In binary oxides, the CF formation during the redox-based RS proceeds through the field-induced migration of oxygen vacancies or metal ions [46, 47]. This type of switching is called valence change mechanism (VCM) as explained in Chapter 1. Many recent experimental results on these types of RS materials have also been interpreted in terms of conduction through atom-sized filamentary structures. This is the case of a wide variety of binary oxides such as SiO_2 [48, 3], HfO_2 [6, 49–51], Ta_2O_5 [4, 5], and ZnO [7, 52], and their suboxides, for which the conductance in the LRS exhibits values close to the quantum conductance unit G_0 . In fact, early observations of steps in the CF conductance of the order of the conductance quantum have already been presented by Hajto *et al.* for a-Si:H [53] and by Yun *et al.* for V_2O_5 [54] thin films. The issue of CF formation is closely connected with the dielectric breakdown phenomenon in ultrathin ($t_{\text{ox}} < 5\text{ nm}$) oxide layers and the observation of two distinctive conduction modes called soft breakdown (SBD) and hard breakdown (HBD) in clear allusion to the magnitude of the leakage current [55].

Filamentary conduction in SiO_2 has been extensively investigated since the 1960s [56–58], and in the last two decades, several conduction mechanisms were proposed to account for the electron transport characteristics under such circumstances. The details of the I - V curves in metal-oxide-semiconductor (MOS)

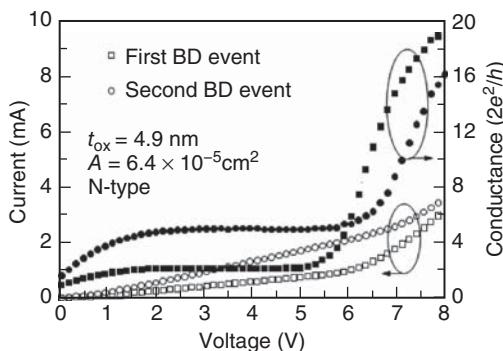


Figure 7.9 I - V characteristics and differential conductance as a function of applied voltage corresponding to two consecutive HBD events detected on the same MOS device. (Reproduced from Ref. [59].)

devices after the occurrence of an HBD event were not investigated in depth until recent years [59]. As illustrated in Figure 7.9, the HBD I - V characteristics in a MOS device are far from linear in the whole bias range. The I - V curves exhibit interesting features both at low and at high voltages, associated with the band bending in the semiconductor and with the prevalence of the Fowler–Nordheim tunneling current, respectively. Particularly noteworthy is the fact that the HBD I - V curve in ultrathin SiO_2 films displays conductance plateaus of the order of G_0 , which is a property typical of mesoscopic systems. This conduction mode is often related to the LRS of RS devices. In contrast, the functional form of the SBD I - V characteristics in MOS devices has been studied thoroughly. The SBD current can be fitted by a power-law model ($I = \alpha V^\beta$) or by an exponential function of the applied bias ($I = A \exp(BV)$). A point that has been much debated is the dependence of the SBD current on the oxide thickness and the device area. Several reports indicate that the SBD conduction is essentially independent of such factors as shown in Figure 7.10 [70]. These observations confirm that the SBD mode corresponds to the formation of a locally enhanced conduction path and that the current is dominated by a small filamentary structure. This is totally consistent with the HRS and LRS I - V characteristics of many RS devices [63–66].

Most of the proposed models that have been designed to account for the filamentary conduction in SiO_2 are to large extent adaptations of well-known electron transport mechanisms in dielectric films. Among others, these include junction-like models, variable range hopping, percolation in nonlinear conductor networks, and direct tunneling. A comprehensive review about this subject can be found in Ref. [59]. The majority of these approaches deal with the SBD mode exclusively, completely disregarding HBD. Interestingly, only three models consider explicitly the atom-sized dimensions of the CF. In this regard, Ting considered the crowding of the electron wave functions into nanowires embedded in the oxide layer connecting partially (SBD) or totally (HBD) the opposite electrodes [67]. This model succeeded in accounting for both regimes at once consistently, including

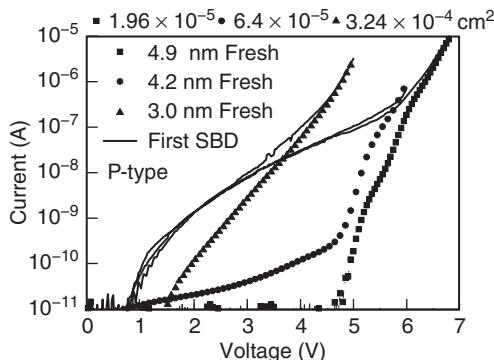


Figure 7.10 Fresh (thin solid lines) and post-SBD $I(V)$ characteristics (symbols) of three MOS devices with different thicknesses and areas. (Reproduced from Ref. [59].)

the temperature effect. On the other hand, Nigam *et al.* [68] proposed that a single-electron tunneling (SET) mechanism can explain both the voltage and temperature dependence of the leakage current after SBD. In this model, the conducting sites forming the break-down (BD) path behave as isolated islands embedded in the oxide layer. To travel between the electrodes, the electrons must pass through such islands by means of successive uncorrelated tunnel events, with these transitions being dictated by the electrostatic potentials due to the charges within the islands. When there is only room for a single electron within the junction, the change of the system's electrostatic energy leads to the so-called Coulomb blockade effect. The conduction process as a whole can be viewed as a tunneling problem through the Coulomb energy barrier, and for this reason, the process has been referred to as macroscopic quantum tunneling of charge. The $I-V$ curves associated with this model are polynomials with powers ranging from 1 to $2N - 1$, with N being the number of tunnel junctions ($N - 1$ traps along the SBD path).

The third approach is the quantum point contact (QPC) model for dielectric breakdown in which the CFs are considered as atom-sized structures spanning the oxide layer [48, 59, 69]. The model does not give details about the microscopic nature of the atomic bridge but provides an interpretation of the $I-V$ characteristic in terms of a transmission probability and a confining potential according to the concepts presented in Section 7.2. The application of the Landauer formulation is influenced by the observation that the $I-V$ characteristic is independent of the oxide thickness and device area and that the HBD conductance exhibits values close to G_0 . In addition, it is found that the SBD $I-V$ characteristics in MOS devices are independent of the injection polarity and substrate type [70]. This has two remarkable consequences: first, knowledge about the particular features of the injecting electrode must not, *a priori*, be explicitly included in the modeling, and second, the voltage, and not the field, is the relevant variable governing the electron transport. The QPC model assumes that SBD and HBD share a unique physical root so that a common conceptual framework is suitable for the description of both modes. According to the QPC model, the effective tunneling

barrier formed at the narrowest point along the constriction (for illustration, see Figure 7.2) determines whether SBD, HBD, or nonlinear hard breakdown (NL-HBD) is observed. An analysis of BD statistics shows that the CFs have cross-sectional areas in the range from 10^{-14} to 10^{-12} cm^2 [71]. For this range of lateral dimensions, quantum effects can be certainly expected. However, contrary to what is observed in typical RS materials (HfO_2 , NiO , ZnO , etc.), the switching process in SiO_2 from LRS to HRS is in most of the cases irreversible, which indicates that some fundamental change in the CF's structure takes place. Recall that ballistic transport and energy quantization are the two conditions required for the observation of quantization effects. However, ballistic transport would not be necessary along the whole BD path. An inelastic mean free path smaller than the lateral size of the constriction's bottleneck is only required. Importantly, as mentioned in Section 7.2, conductance levels close to integer values of G_0 are indicative of electron transport through atom-sized constrictions, and the question whether we are dealing with true conductance quantization effects in these materials remains to be demonstrated.

For the purpose of discussion of the experiments, it is useful to rewrite Eq. (7.2) as [72],

$$I = \frac{2e}{h} \int_{-\infty}^{\infty} T(E, V_c) [f(E - \beta eV_c) - f(E + (1 - \beta)eV_c)] dE \quad (7.7)$$

in order to account for asymmetries in the voltage drop profile. Here, V_c is the voltage drop across the constriction, $0 < \beta < 1$ is the fraction of V_c that drops on the source side of the constriction, and T is the transmission coefficient of the system. In the case of semiconductor electrodes or in the presence of a series resistance, we can assume that only part of the applied bias drops across the constriction, that is, $V_c = V - V_0(V)$, with V and $V_0(V)$ being the applied voltage and a threshold voltage, respectively. For the sake of simplicity, we consider here $V_0(V) = V_0$ a constant, where V_0 is a fitting parameter, but more elaborate models can be found in the literature [73]. If we assume that an inverted parabolic barrier of height E_0 describes the bottom of the first subband at its narrowest point (see Figure 7.11), then the tunneling probability reads,

$$T(E) = \frac{1}{1 + \exp[-\alpha(E - E_0)]}, \quad (7.8)$$

where α is a parameter related to the shape of the barrier and, more specifically, to the second derivative of the potential profile [74]. Moreover, if the smearing of the Fermi functions at the electrodes (zero-temperature limit) is neglected, expression Eq. (7.7) yields

$$I(V) = \frac{2e}{h} \left\{ e(V - V_0) + \frac{1}{\alpha} \ln \left[\frac{1 + \exp \{ \alpha [\Phi - \beta e(V - V_0)] \}}{1 + \exp \{ \alpha [\Phi + (1 - \beta)e(V - V_0)] \}} \right] \right\}, \quad (7.9)$$

where Φ is the effective tunneling barrier height measured from the equilibrium Fermi energy. This expression is valid for a single-mode ballistic conductor. As shown in Figure 7.11, depending on the cross-sectional area of the CF's bottleneck,

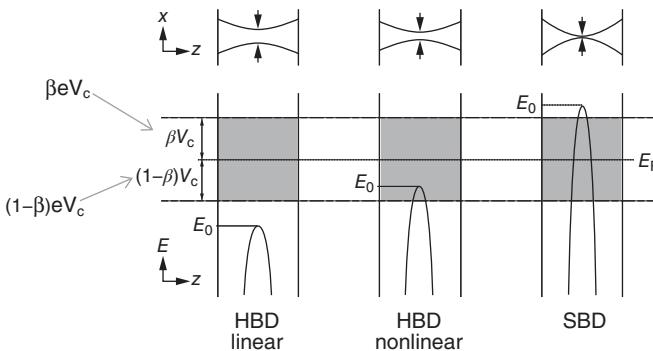


Figure 7.11 Schematic energy diagram of the constriction and its associated potential barrier profile.

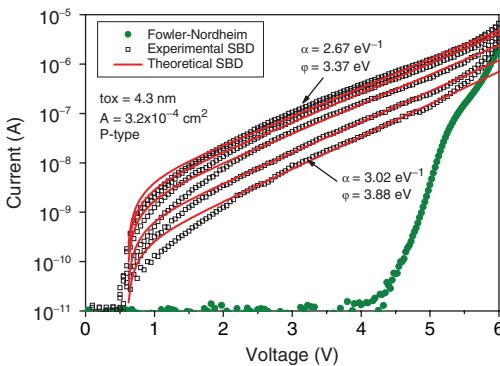


Figure 7.12 SBD I - V characteristics corresponding to first events measured on different MOS devices. Φ is the barrier height associated with the first subband level and α is a parameter related to the shape of the barrier. (Reproduced from Ref. [59].)

E_0 may be below or above the current-carrying energy states. In the first case, the conduction mode corresponds to HBD, in the second case to SBD. When the subband level is in between the quasi-Fermi levels, we can have a mixture of both modes, which is identified, in terms of the physics of mesoscopic conductors [19], as the NL-HBD mode. Experimental and model results for the SBD I - V characteristics are illustrated in Figure 7.12. Notice that neither the oxide thickness nor the device area appears in Eq. (7.9), in agreement with the experimental observations. The deviations of the theoretical curves in the low voltage range may be ascribed to a deficient modeling of the potential drops distribution (semiconductor electrode). If the barrier vanishes ($\Phi \ll 0$) in Eq. (7.9), the linear I - V characteristic associated with HBD is obtained.

Furthermore, it can be demonstrated that, under certain limitations, the I - V characteristic in the LRS can be approximated by [19, 75],

$$I \approx G_0(\beta N_{\rightarrow} + (1 - \beta)N_{\leftarrow})(V - V_0), \quad (7.10)$$

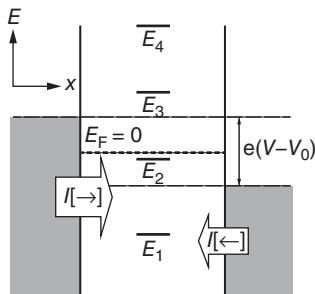


Figure 7.13 Schematic energy diagram for a filamentary structure with multiple subbands. E_i represents the top of each subband.

where N_{\rightarrow} is the number of right-going conduction modes, obtained by counting the number of subbands having $E_i \leq e\beta(V - V_0)$, and N_{\leftarrow} is the number of left-going conduction modes, obtained by counting the number of subbands with $E_i \leq -e(1 - \beta)(V - V_0)$ (see Figure 7.13). For symmetric potential drops at the two ends of the constriction ($\beta = 1/2$), Eq. (7.10) reduces to:

$$I \approx G_0 N(V - V_0), \quad (7.11)$$

where $N = (N_{\rightarrow} + N_{\leftarrow})/2$. Notice that Eq. (7.11) is formally equivalent to the Landauer formula, Eq. (7.3), in the linear regime for multiple conduction modes [15]. However, notice that if the difference between N_{\rightarrow} and N_{\leftarrow} is an even number, Eq. (7.11) predicts integer multiples of G_0 , whereas if N_{\rightarrow} and N_{\leftarrow} is an odd number, half-integer multiples of G_0 are obtained. These are referred to as the linear and nonlinear conduction modes in quantum point contacts. Some recent experimental observations that have been interpreted in this way are illustrated next.

The QPC model was further investigated by Cester *et al.* [76], who were able to extract, by means of inverse modeling, information about the barrier potential profiles associated with SBD. The model was also invoked to explain the heavy-ion-induced SBD conduction [77], and it was extended to incorporate the temperature dependence of the phenomenon [78]. More recently, the QPC model has been used to explain the experimental high-resistance state (HRS) and LRS I - V characteristics of HfO_2 -based metal-insulator-metal devices as well as their temperature dependence [49, 79]. Degraeve *et al.* proposed an extension of the model to account for the switching dynamics in HfO_2 as well [50, 80]. In their approach, the oxygen vacancies are the mobile defect species responsible for the formation and dissolution of the CF. They assumed that the ion diffusion takes place only inside the CF, which according to their viewpoint acts as a container while the set/reset process is controlled by modulation of the constriction width. The random telegraph noise (RTN) fluctuations in the HRS current in ReRAMs with very small area has been attributed to stochastic vacancy movements in and out of the CF in the context of the QPC model [81]. Prócel *et al.* [51] have also investigated the statistical dispersion of the model parameters for HfO_2 . From the theoretical point of view, it is worth mentioning that the formation of CFs in monoclinic and amorphous HfO_2 was investigated from the first principles by Cartoixa *et al.* [82]. The filamentary paths were built from oxygen vacancies, and by using a Green's function formalism coupled to a density functional theory

code, the conductance of vacancy filaments of different lengths were calculated. According to their results, even the thinnest CFs can sustain conductive channels exhibiting signs of conductance quantization.

There is now substantial evidence that the RS effect in binary oxides is associated with the formation and dissolution of filamentary current pathways spanning the dielectric films. According to experimental studies, the CFs can be formed by oxygen vacancies or metal atoms [37, 47]. Remarkably, contrary to what happens in SiO_2 , RS materials are able to commute reversibly between a high and a low conduction state, which allows for obtaining a great deal of information on a single device. This statistical information is often represented using conductance histograms as in metallic nanowires (Section 7.2). Several recent experimental results are consistent with the idea of conductance quantization or at least with electron transport through atom-sized constrictions. Observation of discrete conductance steps of the order of G_0 in Ta_2O_5 layers has been reported by Tsuruoka *et al.* [83] and by Chen *et al.* [5]. The steps were generated by electrical pulses, but it is not entirely clear from this work to what extent the results depend on the choice of the pulse height. Although the conductance levels presented are quite suggestive of quantization, it is very likely that the atomic discreteness plays a dominant role leading to the stepwise variation. Discrete changes in conductance of the order of G_0 were also reported by Long *et al.* during the unipolar reset transitions of $\text{Pt}/\text{HfO}_2/\text{Pt}$ structures [6] (see Figure 7.14). The conductance changes were interpreted as the signature of atomic-size variations of

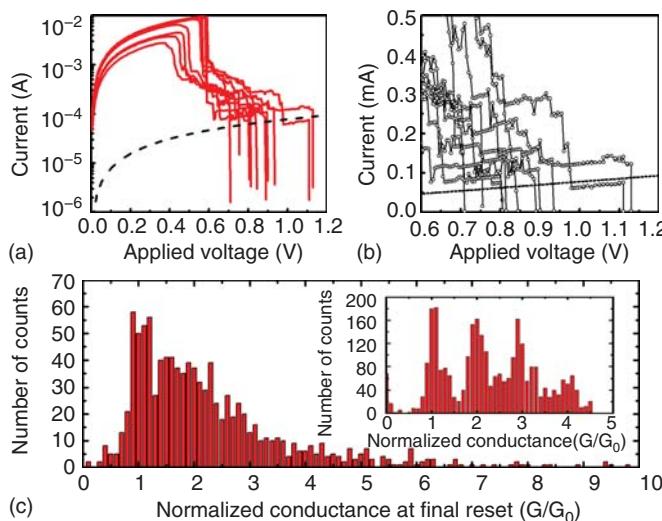


Figure 7.14 (a) I - V curves measured during the application of reset voltage ramps in $\text{Pt}/\text{HfO}_2/\text{Pt}$ structures. The dashed line corresponds to a linear I/V with conductance equal to $G_0 = 2e^2/h$. (b) Detail of the current–voltage evolution during the

last phase of the reset transients. (c) Histogram of conductance at the final reset point; the inset shows the histogram of conductance readings during 100 successive conductance-time traces. (Reproduced from Ref. [6].)

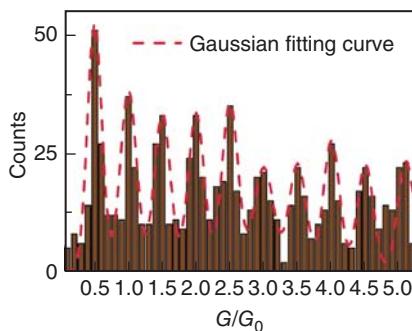


Figure 7.15 Histogram of the conductance values for ITO/ZnO/ITO devices obtained by sweeping the bias to successive levels of maximum voltage. Similar results were obtained for Nb/ZnO/Pt. (Reproduced from Ref. [7].)

the CF nanostructure. According to these authors, the reset process occurs in two phases: first, a progressive narrowing of the CF to the limit of a quantum wire, and second, the opening of a spatial gap that exponentially reduces the transmission probability. The histogram of all data does not display a clear peak structure, but a selection of 100 successive conductance break traces does again show up to four peaks close to integer conductance values.

Zhu *et al.* have reported conductance quantization effects in Nb/ZnO/Pt and ITO/ZnO/ITO structures as well [7]. The existence of nanoscale CFs in these structures has been confirmed by conductive atomic force microscopy (C-AFM) imaging. As illustrated in Figure 7.15, the ITO/ZnO/ITO devices exhibit conductance values close to half-integer multiples of G_0 . Similar results are exhibited by Nb/ZnO/Pt devices. Note that the conductance in this case is measured after the switching has been induced, at a low voltage level of only 0.03 V. This rules out an interpretation of the half-integers as due to the high-bias mechanism of Eq. (7.10). For a pure free-electron-like quantization model, this observation poses serious challenges.

We end this overview with results reported recently by Mehonic *et al.* [3], showing similar effects in silicon-rich silica (SiO_x) resistive switches (see Figure 7.16). In this case, the authors explain the half-integer quantization explicitly in terms of nonlinear effects and subbands within the constriction. Nonlinear effects in HfO_2

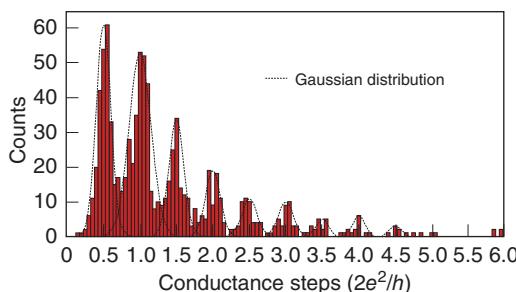


Figure 7.16 Histogram of conductance changes collected for 1000 conductance steps observed in silicon-oxide resistive memory devices. Peaks are obtained at half-integer multiples of G_0 . (Reproduced from Ref. [3].)

were also reported by Degraeve *et al.* [80] and in $\text{CeO}_x/\text{SiO}_2$ -based structures by Miranda *et al.* [73]. Many of these latest results need further confirmation.

7.5

Conclusion and Outlook

The question regarding the origin of the stepwise variation in CFs formed upon breakdown of the dielectric is of considerable interest since it is closely related to an understanding of the mechanism of breakdown. The important conclusion from the observations discussed is that the structure of the CF in many cases has a limited number of recurrent configurations. These are most likely discrete atomic structures, of just a few atoms at the narrowest section of the wire.

The high probability of frequently recurring conductance values is the first remarkable property in these measurements. The second is the fact that the high-probability values appear to be close to integer numbers of conductance quanta, while in other experiments, half-integers are also found. The obvious question that has not been posed up to this point is how one could imagine having 10 or more quantum levels in these systems, while this is normally not observed for pure metal contacts formed by break junctions [18]. Possibly, the matrix provided by the solid electrolyte or dielectric imposes constraints on the accessible atomic configurations. This may explain why specific conductance values appear very frequently, but this would still not explain the proximity to integer values, let alone the half-integer values. Integer series of conductance values are expected for simple metal nanowires with s-electron orbitals only. As soon as p or d electrons are involved, experiments on atomic contacts demonstrate that several partly transmitted channels appear, spoiling the pure quantization features. For most of the CFs that are believed to be formed upon breakdown in the various systems discussed, simple metals are not expected to be relevant. Another possible explanation would be the parallel formation of many single-atom contacts. This would require a limiting mechanism for further growth of CFs beyond a single-atom contact, such as an external current compliance, but then for each CF separately.

Clearly, the nature of the CF remains to be elucidated, but there is a perspective for exploiting techniques that have been developed for the analysis of atomic contacts, such as superconducting subgap structure, application of magnetic fields, and measurement of shot noise, to shed light on these problems. A first step in this direction was taken very recently by Geresdi *et al.* [84]. The time is ripe for more refined experiments on the possible quantum nature of the conductance that will help elucidate the nature of the CFs.

References

1. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005) Quantized conductance atomic switch. *Nature*, **433**, 47.
2. Jameson, J.R., Gilbert, N., Koushan, F., Saenz, J., Wang, J., Hollmer, S., Kozicki, M., and Derhacobian, N. (2012) Quantized conductance in Ag/GeS2/W

- conductive-bridge memory cells. *IEEE Electron Device Lett.*, **33**, 257.
3. Mehnici, A., Rajitoarea, A., Cueff, S., Hudziak, S., Howe, H., Labbe, C., Rizk, R., Pepper, M., and Kenyon, A.J. (2013) Quantum conductance in silicon oxide resistive memory devices. *Sci. Rep.*, **3**, 2708.
 4. Tsuruoka, T., Hasegawa, T., Terabe, K., and Aono, M. (2012) Conductance quantization and synaptic behavior in a Ta_2O_5 -based atomic switch. *Nanotechnology*, **23**, 435705.
 5. Chen, C., Gao, S., Zeng, F., Wang, G., Li, S., Song, C., and Pan, F. (2013) Conductance quantization in oxygen-anion-migration-based resistive switching memory devices. *Appl. Phys. Lett.*, **103**, 043510.
 6. Long, S., Lian, X., Cagli, C., Cartoixa, X., Rurali, R., Miranda, E., Jiménez, D., Perniola, L., Liu, M., and Suñé, J. (2013) Quantum-size effects in hafnium-oxide resistive switching. *Appl. Phys. Lett.*, **102**, 183505.
 7. Zhu, X., Su, W., Liu, Y., Hu, B., Pan, L., Lu, W., Zhang, J., and Li, R.-W. (2012) Observation of conductance quantization in oxide-based resistive switching memory. *Adv. Mater.*, **24**, 3941.
 8. Tombros, N., Veligura, A., Junesch, J., Guimaraes, M.H.D., Vera-Marun, I.J., Jonkman, H.T., and van Wees, B.J. (2011) Quantized conductance of a suspended graphene nanoconstriction. *Nat. Phys.*, **7**, 697.
 9. König, M., Wiedmann, S., Brüne, C., Roth, A., Buhmann, H., Molenkamp, L.W., Qi, X.-L., and Zhang, S.-C. (2007) Quantum spin hall insulator state in HgTe quantum wells. *Science*, **318**, 766.
 10. Ron, A. and Dagan, Y. (2014) One-dimensional quantum wire formed at the boundary between two insulating $LaAlO_3/SrTiO_3$ interfaces. *Phys. Rev. Lett.*, **112**, 136801.
 11. Tsukazaki, A., Ohtomo, A., Kita, T., Ohno, Y., Ohno, H., and Kawasaki, M. (2007) Quantum Hall effect in polar oxide heterostructures. *Science*, **315**, 1388.
 12. van Wees, B.J., van Houten, H., Beenakker, C.W.J., Williamson, J.G., Kouwenhoven, L.P., van der Marel,
 - D., and Foxon, C.T. (1988) Quantised conductance of point contacts in a two-dimensional electron gas. *Phys. Rev. Lett.*, **60**, 848.
 13. Wharam, D.A., Thornton, T.J., Newbury, R., Pepper, M., Ahmed, H., Frost, J.E.F., Hasko, D.G., Peacock, D.C., Ritchie, D.A., and Jones, G.A.C. (1988) One-dimensional transport and the quantisation of the ballistic resistance. *J. Phys. C*, **21**, L209.
 14. van Houten, H., Beenakker, C.W.J., and van Wees, B.J. (1992) *Semiconductors and Semimetals*, vol. 35, Academic Press, pp. 9–112.
 15. Datta, S. (1997) *Electronic Transport in Mesoscopic Systems*, Cambridge University Press.
 16. Landauer, R. (1957) Spatial variation of currents and fields due to localized scatterers in metallic conduction. *IBM J. Res. Dev.*, **1**, 223.
 17. Buttikier, M. (1986) Four-terminal phase coherent conductance. *Phys. Rev. Lett.*, **57**, 1761.
 18. Aguirre, N., Levy Yeyati, A., and van Ruitenbeek, J.M. (2003) Quantum properties of atomic-sized conductors. *Phys. Rep.*, **377**, 81.
 19. KouwenPRB89
 20. Park, H., Park, J., Lim, A.K.L., Anderson, E.H., Alivisatos, A.P., and McEuen, P.L. (2000) Nanomechanical oscillations in a single-C₆₀ transistor. *Nature*, **407**, 57.
 21. Keane, Z.K., Yu, L.H., and Natelson, D. (2006) Magnetoresistance of atomic-scale electromigrated nickel nanocontacts. *Appl. Phys. Lett.*, **88**, 062514.
 22. Heersche, H.B., Lientschnig, G., O'Neill, K., van der Zant, H.S.J., and Zandbergen, H.W. (2007) In situ imaging of electromigration-induced nanogap formation by transmission electron microscopy. *Appl. Phys. Lett.*, **91**, 072107.
 23. Xie, F.Q., Nittler, L., Obermair, C., and Schimmel, T. (2004) Gate-controlled atomic quantum switch. *Phys. Rev. Lett.*, **93**, 128303.
 24. Xie, F.Q., Obermair, C., and Schimmel, T. (2004) Switching an electrical current with atoms: the reproducible

- operation of a multi-atom relay. *Solid State Commun.*, **132**, 437.
25. Li, C.Z. and Tao, N.J. (1998) Quantum transport in metallic nanowires fabricated by electrochemical deposition/dissolution. *Appl. Phys. Lett.*, **72**, 894.
 26. Li, C.Z., Bogozi, A., Huang, W., and Tao, N.J. (1999) Fabrication of stable metallic nanowires with quantized conductance. *Nanotechnology*, **10**, 221.
 27. Li, C.Z., He, H.X., and Tao, N.J. (2000) Quantized tunneling current in the metallic nanogaps formed by electrodeposition and etching. *Appl. Phys. Lett.*, **77**, 3995.
 28. Kiguchi, M., Konishi, T., Miura, S., and Murakoshi, K. (2005) Mechanical fabrication of nano-contacts showing conductance quantization under electrochemical potential control. *Physica E*, **29**, 530.
 29. Krans, J.M., van Ruitenbeek, J.M., Fisun, V.V., Yanson, I.K., and de Jongh, L.J. (1995) The signature of conductance quantization in metallic point contacts. *Nature*, **375**, 767.
 30. Scheer, E., Agrait, N., Cuevas, J.C., Levy Yeyati, A., Ludoph, B., Martin-Rodero, A., Rubio Bollinger, G., van Ruitenbeek, J.M., and Urbina, C. (1998) The signature of chemical valence in the electrical conduction through a single-atom contact. *Nature*, **394**, 154.
 31. van der Post, N., Peters, E.T., Yanson, I.K., and van Ruitenbeek, J.M. (1994) Subgap structure as function of the barrier in atom-size superconducting tunnel junctions. *Phys. Rev. Lett.*, **73**, 2611.
 32. Scheer, E., Joyez, P., Esteve, D., Urbina, C., and Devoret, M.H. (1997) Conduction channel transmissions of atomic-size aluminum contacts. *Phys. Rev. Lett.*, **78**, 3535.
 33. van den Brom, H.E. and van Ruitenbeek, J.M. (1999) Quantum suppression of shot noise in atom-size metallic contacts. *Phys. Rev. Lett.*, **82**, 1526.
 34. Cron, R., Goffman, M.F., Esteve, D., and Urbina, C. (2001) Multiple-charge-quanta shot noise in superconducting atomic contacts. *Phys. Rev. Lett.*, **86**, 4104.
 35. Blanter, Y.M. and Büttiker, M. (2000) Shot noise in mesoscopic conductors. *Phys. Rep.*, **336**, 2–166.
 36. Guo, X., Schindler, C., Menzel, S., and Waser, R. (2007) Understanding the switching-off mechanism in Ag⁺ migration based resistively switching model systems. *Appl. Phys. Lett.*, **91**, 133513.
 37. Waser, R., Dittmann, R., Saikov, G., and Szot, K. (2009) Redox-based resistive switching memories nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21**, 2632.
 38. Wagenaar, J.J.T., Morales-Masis, M., and van Ruitenbeek, J.M. (2012) Observing “quantized” conductance steps in silver sulfide: two parallel resistive switching mechanisms. *J. Appl. Phys.*, **111**, 014302.
 39. Morales-Masis, M., Wiemhöfer, H.D., and van Ruitenbeek, J.M. (2010) Towards a quantitative description of solid electrolyte conductance switches. *Nanoscale*, **2**, 2275–2280.
 40. Terabe, K., Nakayama, T., Hasegawa, T., and Aono, M. (2002) Ionic/electronic mixed conductor tip of a scanning tunneling microscope as a metal atom source for nanostructuring. *J. Appl. Phys.*, **91**, 10110.
 41. Nayak, A., Tamura, T., Tsuruoka, T., Terabe, K., Hosaka, S., Hasegawa, T., and Aono, M. (2010) Rate-limiting processes determining the switching time in a Ag₂S atomic switch. *J. Phys. Chem. Lett.*, **1**, 604–608.
 42. Morales-Masis, M., van der Molen, S.J., Hasegawa, T., and van Ruitenbeek, J.M. (2011) Bulk and surface nucleation processes in Ag₂S conductance switches. *Phys. Rev. B*, **84**, 115310.
 43. Tappertzhofen, S., Valov, I., and Waser, R. (2012) Quantum conductance and switching kinetics of AgI-based micro-crossbar cells. *Nanotechnology*, **23**, 145703.
 44. Geresdi, A., Halbritter, A., Gyenis, A., Makk, P., and Mihály, G. (2011) From stochastic single atomic switch to nanoscale resistive memory device. *Nanoscale*, **3**, 1504–1507.

45. Xu, Z., Bando, Y., Wang, W., Bai, X., and Golberg, D. (2010) Real-time in situ HRTEM-resolved resistance switching of Ag₂S nanoscale ionic conductor. *ACS Nano*, **4** (5), 2515–2522.
46. Kim, K.M., Jeong, D.S., and Hwang, C.S. (2011) Nanofilamentary resistive switching in binary oxide system: a review on the present status and outlook. *Nanotechnology*, **22**, 254002.
47. Wu, X., Cha, D., Bosman, M., Raghavan, N., Migas, D., Borisenko, V., Zhang, X., Li, K., and Pey, K.L. (2013) Intrinsic nanofilamentation in resistive switching. *J. Appl. Phys.*, **113**, 114503.
48. Suñé, J., Miranda, E., Nafría, M., and Aymerich, X. (1998) Point contact conduction at the oxide breakdown of MOS devices. IEEE International Electron Device Meeting (IEDM), p. 191.
49. Miranda, E., Walczyk, C., Wenger, C., and Schroeder, T. (2010) Model for the resistive switching effect in HfO₂ MIM structures based on the transmission properties of narrow constrictions. *IEEE Electron Device Lett.*, **31**, 609.
50. Degraeve, R., Roussel, Ph., Goux, L., Wouters, D., Kittl, J., Altimime, L., Jurczak, M., and Groeseneken, G. (2010) Generic learning of TDDB applied to RRAM for improved understanding of conduction and switching mechanism through multiple filaments. IEEE International Electron Device Meeting, p. 28.4.1.
51. Prócel, L.M., Trojman, L., Moreno, J., Crupi, F., Maccaronio, V., Degraeve, R., Goux, L., and Simoen, E. (2013) Experimental evidence of the quantum point contact theory in the conduction mechanism of bipolar HfO₂-based resistive random access memories. *J. Appl. Phys.*, **114**, 074509.
52. Zhu, X., Shang, J., and Li, R.-W. (2012) Resistive switching effects in oxide sandwiched structures. *Front. Mater. Sci.*, **6**, 183.
53. Hajto, J., Rose, M.J., Snell, A.J., Osborne, I.S., Owen, A.E., and Lecomber, P.G. (1991) Quantised electron effects in metal/a-Si:H/metal thin film structures. *J. Non-Cryst. Solids*, **137–138**, 499.
54. Yun, E.-J., Becker, M.F., and Walser, R.M. (1993) Room temperature conductance quantization in V|amorphous-V₂O₅|V thin film structures. *Appl. Phys. Lett.*, **63**, 2493.
55. Depas, M., Nigam, T., and Heyns, M. (1996) Soft breakdown of ultra-thin gate oxide layers. *IEEE Trans. Electron Devices*, **43**, 1499.
56. O'Dwyer, J. (1964) *The Theory of Dielectric Breakdown in Solids*, Oxford University Press, London.
57. Klein, N. (1966) The mechanism of self-healing electrical breakdown in MOS structures. *IEEE Trans. Electron Devices*, **13**, 788.
58. Solomon, P. (1977) Breakdown in silicon oxide – A review. *J. Vac. Sci. Technol.*, **14**, 1122.
59. Miranda, E. and Suñé, J. (2004) Electron transport through broken down ultra-thin SiO₂ layers in MOS devices. *Microelectron. Reliab.*, **44**, 1.
60. Halimaoui, A., Brière, O., and Ghibaudo, G. (1997) Quasi-breakdown in ultrathin gate dielectrics. *Microelectron. Eng.*, **36**, 157.
61. Okada, K., Kawasaki, S., and Hirofushi, Y. (1994) New experimental findings on stress induced leakage current of ultra thin silicon dioxides. Proceedings of Extended Abstracts of the SSDM'94, p. 565.
62. MiraED00
63. Ye, J., Li, Y., Gao, J., Peng, H., Wu, S., and Wu, T. (2010) Nanoscale resistive switching and filamentary conduction in NiO thin films. *Appl. Phys. Lett.*, **97**, 132108.
64. Ielmini, D. (2011) Filamentary-switching model in RRAM for time, energy and scaling projections. IEEE International Electron Device Meeting (IEDM), p. 17.2.1.
65. Bersuker, G., Gilmer, D., Veksler, D., Yum, J., Park, H., Lian, S., Vandelli, L., Padovani, A., Larcher, L., McKenna, K., Shluger, A., Iglesias, V., Porti, M., Nafría, M., Taylor, W., Kirsch, P., and Jammy, R. (2010) Metal oxide RRAM switching mechanism based on conductive filament microscopic properties.

- IEEE International Electron Device Meeting (IEDM), p. 19.6.1.
66. Kim, Y., Kim, J., Choi, J., Hwang, I., Hong, S., Kang, S., and Park, B. (2011) Resistive switching behaviors of NiO films with controlled number of conducting filaments. *Appl. Phys. Lett.*, **98**, 192104.
 67. Ting, D. (1999) An embedded quantum wire model of dielectric breakdown. *Appl. Phys. Lett.*, **74**, 585.
 68. Nigam, T., Martin, S., and Abusch-Magder, D. (2003) Temperature dependence and conduction mechanism after analog soft breakdown. IEEE International Reliability Physics Symposium (IRPS), p. 417.
 69. Suñé, J., Miranda, E., Nafría, M., and Aymerich, X. (1999) Modeling the breakdown spots in silicon dioxide films as point contacts. *Appl. Phys. Lett.*, **75**, 959.
 70. Miranda, E., Suñé, J., Rodríguez, R., Nafría, M., Aymerich, X., Fonseca, L., and Campabadal, F. (2000) Soft breakdown conduction in ultrathin (3-5 nm) gate dielectrics. *IEEE Trans. Electron Devices*, **47**, 82.
 71. Suñé, J. (2001) New physics-based analytic approach to the thin-oxide breakdown statistics. *IEEE Electron Device Lett.*, **22**, 296.
 72. Hu, P. (1987) One-dimensional quantum electron system under a finite voltage. *Phys. Rev. B*, **35**, 4078.
 73. Miranda, E., Kano, S., Dou, C., Kakushima, K., Suñé, J., and Iwai, H. (2012) Nonlinear conductance quantization effects in CeO_x/SiO₂-based resistive switching devices. *Appl. Phys. Lett.*, **101**, 012910.
 74. Büttiker, M. (1990) Quantized transmission of a saddle-point constriction. *Phys. Rev. B*, **41**, 7906.
 75. Xu, H. (1993) Theory of nonlinear ballistic transport in quasi-one-dimensional constrictions. *Phys. Rev. B*, **47**, 15630.
 76. Cester, A., Bandiera, L., Suñé, J., Boschiero, L., Ghidin, G., and Paccagnella, A. (2001) A novel approach to quantum point contact for post soft breakdown conduction. IEEE International Electron Device Meeting (IEDM), p. 13.6.1.
 77. Suehle, J., Wang, B., Miyahara, T., Vogel, E., Bernstein, J., and Conley, J. (2001) Heavy-ion-induced soft breakdown of thin gate oxides. *IEEE Trans. Nucl. Sci.*, **48**, 1913.
 78. Avellán, A., Miranda, E., Schroeder, D., and Krautschneider, W. (2005) Model for the voltage and temperature dependence of the soft-breakdown current in ultrathin gate oxides. *J. Appl. Phys.*, **97**, 14104.
 79. Walczyk, C., Walczyk, D., Schroeder, T., Bertaud, T., Sowinska, M., Lukosius, M., Fraschke, M., Wolansky, D., Tillack, B., Miranda, E., and Wenger, C. (2011) Impact of temperature on the resistive switching behavior of embedded HfO₂-based RRAM devices. *IEEE Trans. Electron Devices*, **58**, 3124.
 80. Degraeve, R., Fantini, A., Clima, S., Govoreanu, B., Goux, L., Chen, Y.Y., Wouters, D., Roussel, Ph., Kar, G., Pourtois, G., Cosemans, S., Kittl, J., Groeseneken, G., Jurczak, M., and Altimime, L. (2012) Dynamic 'hour glass' model for SET and RESET in HfO₂ RRAM. Symposium on VLSI Technology, p. 75.
 81. Raghavan, N., Degraeve, R., Fantini, A., Goux, L., Strangio, S., Govoreanu, B., Wouters, D., Groeseneken, G., and Jurczak, M. (2013) Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability. IEEE International Reliability Physics Symposium (IRPS), p. 5E.3.1.
 82. Cartoixa, X., Rerali, R., and Suñé, J. (2012) Transport properties of oxygen vacancy filaments in metal/crystalline or amorphous HfO₂/metal structures. *Phys. Rev. B*, **86**, 165445.
 83. Tsuruoka, T., Hasegawa, K., Terabe, M., Aono (2012) Conductance quantization and synaptic behavior in a

- Ta₂O₅-based atomic switch, *Nanotechnology*, **23**, 435705
84. Geresdi, A., Csontos, M., Gubicza, A., Halbritter, A., and Mihály, G. (2014) A fast operation of nanometer-scale metallic memristors: highly transparent conductance channels in Ag₂S devices. *Nanoscale*, **6**, 2613.

8**Dielectric Breakdown Processes**

Jordi Suñé, Nagarajan Raghavan, and K. L. Pey

8.1**Introduction**

The dielectric breakdown (BD) of a thin insulating film involves the creation of a conducting filament (CF), which effectively suppresses its insulating property. In CMOS circuits; BD is an undesired event, which might cause circuit failure, so that it has mainly been studied in the framework of CMOS reliability. In this context, much has been learned about the BD of thin SiO_2 -based films (the natural oxide in silicon microelectronics) and more recently of the high permittivity (high- k/HK) and SiO_2 gate stacks required to reduce the gate leakage for equivalent oxide thickness (T_{ox}) below 1 nm [1–6].

In this work, we deal with the BD of thin oxide films in metal–insulator–metal (MIM) or Metal Oxide Semiconductor (MOS) structures, which show resistive switching (RS) effects. Most resistive memories (ReRAM) operate by the alternate creation and partial destruction of a CF. In these filamentary ReRAM structures, BD does not cause failure since it is intrinsic to the device operation and data storage.

Oxide reliability and ReRAM scientific communities have evolved separately for years. However, it is very important to bridge these two communities for a synergic improvement of the understanding of the BD phenomenon in both fields [7–9]. In this chapter, we focus on how to improve our understanding of ReRAM by learning from oxide reliability studies. First, we know that BD is related to the creation of the CF during electroforming, often related to a kind of soft breakdown (SBD). On the other hand, it has been suggested that the SET transition and, consequently, the retention of the high-resistance state (HRS) might also be related to the BD [10, 11]. The stability of the BD location and the generation of multiple CFs might also affect ReRAM performance, variability, and endurance [7]. In this regard, the theory of multiple BD statistics [12] might be useful to deal with the generation of multiple CFs in ReRAM [8]. Understanding the conduction properties of BD paths can help unveil the nature of CF in ReRAM. However, this issue is already considered in detail in Chapter 7.

While electrical and statistical studies provide a wealth of insight into the kinetics of the BD (CF formation) phenomenon, they are insufficient to understand its physical and chemical origin, except for some speculations using information such as activation energy. Without the knowledge of the chemistry involved in defect-generation-induced BD, it would not be possible to identify the key process and material parameters governing the BD/electroforming process and control them so as to reduce variability at the nanoscale either through material/process design/device architecture changes or a combination of them.

In this chapter, we first review the basics of thin-oxide BD, including the description of the different BD modes, and the techniques used to measure the time-to-BD and BD voltage distributions. Given that BD is related to the generation of defects in the oxide, we will also discuss the physics of defect generation and the BD statistics in the framework of the percolation model [13]. The statistics of multiple BD events, the impact of Progressive Breakdown (PBD) on the final failure distribution, and the BD distribution in bilayer oxide stacks will also be discussed. This will be followed by a detailed microscopic study on MOS stacks to analyze the elemental makeup of the BD region and its dependence on the current compliance for different material stacks (dielectric and electrode materials) ranging from SBD ($I_{\text{comp}} \sim 1 \mu\text{A}$) to hard breakdown (HBD) ($I_{\text{comp}} \sim 100 \mu\text{A}$ to 1 mA), where I_{comp} refers to the system predefined compliance setting. The root cause of BD and the ionic transport/bond-dissociation mechanisms helping the BD process will be identified using these microscopic investigations. Finally, we will conclude this chapter by summarizing the key findings that will be relevant from BD studies in improving the controllability of the forming process.

8.2

Basics of Dielectric Breakdown

Before oxide BD, conduction through the insulator is uniformly distributed in area, and it is dominated either by tunneling injection (as in the case of high-quality SiO_2 films) or by defect (trap)-assisted mechanisms such as Poole–Frenkel emission or inelastic trap-assisted tunneling in the case of more defective HK insulators. After oxide BD, conduction is strongly localized in a nanoscale CF in spite of the fact that its area is much smaller than the total oxide area. Based on the analysis of the BD statistics, the BD path lateral size has been estimated to range from 1 to 10 nm. Due to this extreme localization, several BD paths can nucleate at different spatial locations, thus reducing the post-BD resistance. Under operating conditions, the BD of a MOS device must be an extremely unlikely event. Thus, oxide BD testing requires highly accelerated stress conditions (higher voltages and temperatures) in order to induce failure within a practical measurement time. In the case of ReRAM, all the devices need to be electroformed within a reasonable time, so that the forming procedure is usually very similar to a BD reliability test. Since BD has a stochastic character, a statistically significant number of fresh samples need to be stressed to measure a statistical BD distribution. The simplest method for characterizing the BD consists in the application of a constant voltage stress

(CVS) while periodically reading the current until BD is detected as a current jump and the time-to-breakdown, t_{BD} , is registered as the natural BD variable.

Another widely used method to characterize BD is the application of a ramp voltage stress (RVS). In this case, the BD is detected as a current jump during an I - V sweep that determines the BD voltage, V_{BD} . The value of V_{BD} is known to increase with the ramp rate because there is insufficient time for defects to nucleate and evolve at lower voltages for high ramp rates. The voltage ramp is also widely used to electroform ReRAM structures and for *quasistatic* SET/RESET cycling. The SET and the forming voltages are also known to depend on the ramp rate for analogous reasons.

Being related to the random generation of defects in the insulator, the BD has a random nature and requires a statistical description. The Weibull distribution is generally accepted as the best distribution model to describe BD [14]. The Weibull distribution is a type of extreme-value distribution with cumulative failure function given by:

$$F(x) = 1 - \exp \left\{ - \left(\frac{x}{x_{63\%}} \right)^{\beta} \right\}, \quad (8.1)$$

where x is the statistical variable, $x_{63\%}$ is the scale factor, and β is the shape factor (indicative of spread in data) also known as *Weibull Slope*. The graphical representation of this type of distribution is usually carried out in the Weibull plot, in which the Weibit parameter, $W(x) \equiv \ln(-\ln(1 - F(x)))$, is represented versus $\ln(x)$ to yield a straight line with equation $W(x) = \beta \ln(x) - \beta \ln(x_{63\%})$. The Weibull distribution is compatible with the weakest-link property of BD. According to this property, the BD distributions of structures with areas A_1 and A_2 are related to the Weibit values by $W_1 = W_2 + \ln(A_1/A_2)$, provided that defect generation is uniform in area. This is also known as *Poisson area scaling*, which has been experimentally demonstrated for the distributions of t_{BD} and V_{BD} (see Figure 8.1) and which is routinely used for reliability extrapolation from the device to the circuit level.

The statistical results obtained from CVS and RVS methods can be directly related to one another by considering the relation of the BD to the cumulative defect generation [15]. Let us consider a CVS with stress voltage V_s and a RVS with ramp rate R . In the case of CVS, the result is the t_{BD} Weibull distribution with scale factor $t_{63\%}$ and shape factor β_{CVS} . By using the RVS method, we obtain the Weibull distribution of V_{BD} , with scale factor $V_{63\%}$ and shape factor β_{RVS} . By assuming that $t_{63\%}$ scales with the stress voltage according to a power law, $t_{63\%}(V) \propto V^{-n}$ [16], it can be shown that for any cumulative failure percentile, we can relate V_{BD} and t_{BD} by:

$$V_{\text{BD}} = [(n+1)RV_s^n t_{\text{BD}}]^{1/(n+1)}, \quad (8.2)$$

so that we can obtain the distribution of V_{BD} from the measured t_{BD} distribution or vice versa. Since this relation is valid for any failure percentile, we can apply it to the point $W = 0$ to write the relation between the scale factors $t_{63\%}$ and $V_{63\%}$, thus explicitly demonstrating that $V_{63\%}$ depends on the ramp rate. Moreover, it is also easy to demonstrate that the relation between the Weibull slopes is

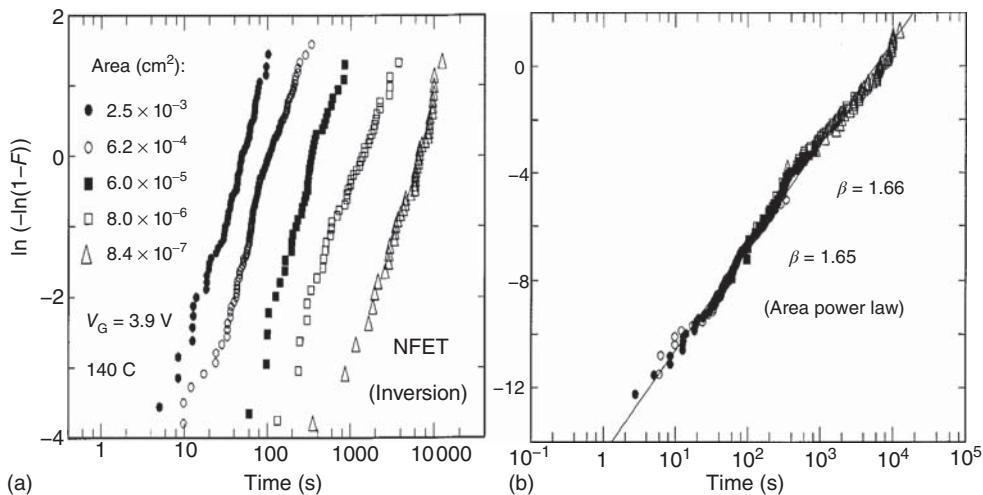


Figure 8.1 (a) Experimental BD distributions for SiO₂-based MOS structures with $T_{\text{ox}} = 2.7 \text{ nm}$ and five different areas (see legend). (b) Normalization of the same t_{BD} statistical data to a reference area of $8.4 \times 10^{-7} \text{ cm}^2$, according to the scaling law derived from the weakest-link property.

Straight lines in the Weibull plot demonstrate that the Weibull model is adequate for the time-to-BD distribution. Overlapping of the data for structures with different areas demonstrates the area scaling property. (Reprinted with permission from Ref. [14] © 2002 IEEE.)

$\beta_{\text{RVS}} = \beta_{\text{CVS}}(n + 1)$, thus showing that β_{RVS} is related to the voltage acceleration exponent of BD (n) and that it does not depend on the ramp rate. Since n is usually very large ($n = 40 - 50$ in the case of thin SiO₂ gate oxides [16]), the slope of the V_{BD} distribution is much larger than that of t_{BD} . The relation between the CVS and RVS statistics has been recently shown to apply to the distributions of the time-to-SET (t_{set}) measured under (constant voltage) pulsed SET operation and the SET voltage V_{set} obtained under *quasistatic* (voltage ramp) SET switching [10].

In the field of oxide reliability, a key issue is the projection of the BD statistics obtained during accelerated stress experiments to forecast the failure statistics of large circuits operated at much lower voltages. The main steps involved in reliability extrapolation are illustrated in Figure 8.2. First, since the stress data are obtained at high voltages and temperatures, an extrapolation to operating conditions is required. Second, since the data are obtained on small-area test structures, an extrapolation to total chip oxide area is needed. Finally, since a limited number of test structures (typically ~ 100) are stressed in BD experiments, an extrapolation to low percentiles is also needed. For lifetime extrapolation from stress to operating conditions, voltage and temperature acceleration models are required. In the case of thick oxides, electric field acceleration was usually considered. However, in the thin oxides of present interest, for advanced CMOS ($T_{\text{ox}} < 5 \text{ nm}$), the energy of the carriers at the anode interface (roughly equal to the applied voltage due to quasi-ballistic transport) is considered to be the fundamental BD variable. Acceleration models usually deal only with the scale factors because the Weibull slope

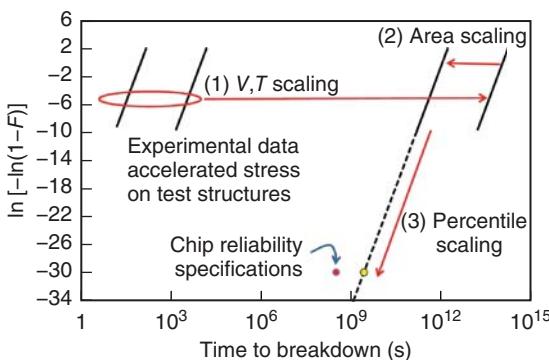


Figure 8.2 A schematic representation of the projection procedures used for gate oxide breakdown reliability assessment. First, data from strongly accelerated BD experiments are extrapolated to operating conditions using voltage and temperature

acceleration models. Second, using the weakest-link property, the results are extrapolated to chip total oxide area. Finally, the results are extrapolated to low failure percentiles using the Weibull distribution model.

is found to be independent of voltage and temperature [14]. The three most commonly used acceleration models are the exponential field (voltage) model, $t_{63\%} \sim \exp(-\gamma_V V)$; the reciprocal exponential model, $t_{63\%} \sim \exp(C/V)$; and the power-law model, $t_{63\%}(V) \propto V^{-n}$. Different aspects of the defect generation physics provide support to different acceleration models, and the choice might also depend on the oxide thickness. In the case of ultrathin SiO_2 -based insulators ($T_{\text{ox}} < 5 \text{ nm}$) stressed under low voltages ($V < 5 \text{ V}$), the power-law model [16] is quite widely accepted. Recently, the power-law model has also been found to be adequate to describe the SET voltage dependence in ReRAM structures based on other transition metal oxides such as TiO_2 and HfO_2 [10]. For extrapolation to total oxide area, the weakest-link property is used. Finally, for extrapolation to low failure percentiles, different distributions might fit the data rather well in the limited high-failure-percentile observation window while giving very different projection to low percentiles. Although some arguments have been made to justify other models such as the lognormal distribution, many physical arguments and experimental observations favor the assumption of the Weibull model for the first BD distribution. In the case of ReRAM, extrapolation to low (and also to high) failure percentiles has been shown to be useful in dealing with the SET speed-disturb trade-off [10], and the Weibull model has also been proposed [10, 11].

The oxide thickness drastically influences the BD phenomenology. Along with the scaling of T_{ox} from hundreds of nanometers to $\sim 1 \text{ nm}$, even the physical mechanisms responsible for the degradation and BD have drastically changed [1, 3]. From the point of view of reliability, the two main effects of the reduction of T_{ox} are a drastic reduction of $t_{63\%}$ (if the stress voltage is not scaled down) and a decrease in the Weibull slope, which is found to be proportional to T_{ox} [17]. The reduction of low-percentile lifetime margin can be estimated to be ~ 25 orders of magnitude (~ 20 orders due to $t_{63\%}$ reduction and ~ 5 orders due to the decrease in β) when

scaling the oxide from 3.5 to 1 nm. This drastic decrease in the reliability margin is only partially compensated by scaling the operating voltage and by the increase in the voltage acceleration factor ($\gamma_V = n/V$, in the power-law acceleration model). The reduction in T_{ox} also has consequences in the post-BD properties. In the 3–5 nm range, the BD events can be classified in terms of their severity as SBD or HBD [18], with the post-BD current being 3–4 orders of magnitude smaller for SBD than for HBD. The situation is even more complex in ultrathin oxides (below ~2 nm) because the HBD becomes progressive [19] and the SBD stage is unstable (noisy). Progressive breakdown (PBD) has a very important consequences for thin-oxide reliability because a certain time (the residual time) is required for the evolution of the BD current to a value that causes the device to malfunction. This extra reliability margin might be very long at low percentiles and low operating voltages. Moreover, two different phases can be distinguished in the evolution of the current during PBD [20, 21]. In the very early stages after the BD (digital-SBD phase), the current remains rather constant (on average) and shows large multi-level random discrete fluctuations. In the second phase (analog-SBD), the current suddenly evolves much faster in a monotonic fashion toward quasi-saturation. On the other hand, a critical voltage (which is, in most cases, greater than the operating voltage of 1 V for $T_{ox} \geq 1.2$ nm) below which the transition from digital to analog BD phase hardly occurs has also been reported [22, 23], implying that the lifetime can be prolonged significantly more than what the standard extrapolation models predict for the PBD stage [23].

In general, it can be concluded that the BD consists of two different phases. First, there is a nucleation phase in which defects are generated randomly in the bulk of the oxide (as opposed to earlier theories on very thick SiO_2 dielectrics (10–100 nm) by Wolters and van der Schoot [24, 25] and Fischetti *et al.* [26] when defect generation is believed to begin at anodic interfaces due to conduction-band offset energy of the arriving electrons and the defect cluster gradually propagates toward the cathode forming a “tree”-like structure) and finally triggers the formation of a percolation path. Once the percolation path is triggered, the local current increases significantly and causes lateral wear-out and growth of the CF. This evolution depends on many factors related to the stress (mainly applied voltage and temperature) and to the structure (oxide thickness, device area, insulator material, nature and thickness of the electrodes, etc.) and also on stochastic effects at the microscopic scale (such as electrode–dielectric interface roughness, dielectric microstructure (role of grain boundaries (GBs)) [27]) that determine the initial shape of the CF. For example, it has been demonstrated that the PBD growth rate depends exponentially on T_{ox} and the stress voltage [19], that the stress temperature might change the post-BD CF properties from a stable SBD to a continuously growing PBD mode [28], and that changing from polysilicon to metal electrode might completely eliminates the PBD phase [29, 30]. Under some circumstances, if the post-BD current is not limited by external means (i.e., a current compliance limiter), severe thermal effects can cause irreversible damage to the stack. The two-phase nature of oxide BD (creation of the percolation path + lateral growth) is very similar to what has been suggested to occur in the

SET transition of filamentary ReRAM, where a nucleation time was considered to precede the lateral growth of the CF [31]. The analogy between the PBD and the SET process might be useful to clarify some aspects of the physics and statistics of the SET transition. It is also remarkable that there is a technology need to operate ReRAM in the ultralow-power regime for both embedded and stand-alone applications. This means that the BD event occurring during forming must necessarily be softer and that the SET/RESET operation of commercially viable ReRAM must be based on very narrow or even incomplete CFs. Understanding the path formation and its subsequent lateral growth is of paramount importance to deal with the stability, noise, variability, reliability, and performance of ReRAM, and the evidences accumulated in the field of dielectric BD over the past couple of decades will help.

8.3

Physics of Defect Generation

Under electrical stress, defects are generated in the oxide bulk and at the interfaces. When a critical density of defects is reached, oxide BD is triggered. The atomic-scale nature of the defects that cause the final oxide BD are now rather well known after the systematic use of physical analysis tools and techniques [32]. However, there is more than one model that has been proposed to explain the phenomena of defect generation and their driving forces. The signature of defect generation has also been evidenced by a variety of macroscopic electrical observations such as positive charge trapping, trapping of electrons in native and newly generated neutral traps, generation of interface states, and increase in the low-field leakage current. All these effects have been used to monitor the oxide degradation under different stress conditions, and the following conclusions have been derived: (i) the average density of defects required to trigger the BD (N_{BD}) decreases with the oxide thickness and is independent of the stress voltage (at least to the first order), (ii) the statistical variations of N_{BD} increase for thinner oxides, and (iii) the defect generation rate increases very strongly with the stress voltage and also depends on the stress temperature.

The importance of the oxide electric field (F_{ox}) in the oxide degradation process has been well known from the early times of BD research. Some authors have considered that defect generation takes place by the process of field-assisted thermally activated bond breakage [6, 33], so that oxide BD is considered to be a field- and time-driven process. However, there is strong experimental evidence demonstrating that both electron fluency and electron energy also play a crucial role [24, 34], as considered in defect generation models such as the Anode Hydrogen Release (AHR) model [35] and the anode hole injection (AHI) model [36]. In these models, defect generation is described as a two-step process: (i) electrons are injected through the oxide, lose their energy at the anode interface, and release some positively charged species (holes or protons) and (ii) the released species are injected into the oxide, travel backward to the cathode, and generate defects in the

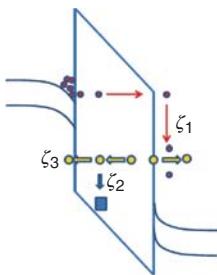


Figure 8.3 Schematic representation of the different processes involved in defect generation according to current energy-based models such as the AHR model. ζ_1 represents the efficiency of H release by incoming electrons, ζ_2 is the defect creation efficiency by the released species, and ζ_3 is the probability of the released species escaping to the electrodes without causing damage in the oxide.

oxide by reaction with some unspecified precursor sites. This two-step process is schematically depicted in Figure 8.3, where ζ_1 represents the efficiency of the injected electrons to release species from the anode interface, ζ_2 the efficiency of the generated species to generate defects in the oxide bulk, and ζ_3 the probability to exit from the oxide bulk toward the electrodes. By assuming that the injected carriers are responsible for the generation of defects in the oxide, the dynamics of defect generation can be written as $dN(Q)/dQ = \zeta/qT_{\text{ox}}$, where Q is the total charge fluency per unit area, ζ the defect generation efficiency, and N the volume density of generated defects. Under some reasonable assumptions, it follows that $\zeta = k\zeta_1$ with $k \equiv (\zeta_2/(\zeta_2 + \zeta_3))$. This means that defect generation efficiency can be understood as the combination of *release* (ζ_1) and *reaction* (k) processes. If ζ can be considered to be independent of N and Q , the defect density is proportional to the injected charge $N = \zeta Q/qT_{\text{ox}}$, although some experiments suggest a time power law $N \propto Q^\alpha$ with $\alpha < 1$ [37, 38]. By assuming a linear relation ($\alpha = 1$) and provided that the current density, J , remains constant during CVS, the time to BD can be calculated as $t_{63\%} = qT_{\text{ox}}N_{\text{BD}}/J\zeta$. This is a simple equation that summarizes our current understanding of oxide BD, stating that a BD model requires the combination of three elements: (i) the relation between defect density and BD probability to provide $N_{\text{BD}} = N_{\text{BD}}(A_{\text{ox}}, T_{\text{ox}})$; (ii) a physics-based model for the defect generation efficiency as a function of the stress conditions $\zeta = \zeta(V, T)$; and (iii) a model for the voltage, temperature, and thickness dependence of the current density $J = J(V, T, T_{\text{ox}})$.

In this section, we review two different models for the physics of defect generation. These are examples of the two main approaches to oxide degradation: the thermochemical model, which assumes an electric-field-assisted thermal generation of oxide defects [6, 33], and the AHR model [35], which is an example of defect generation initiated by the energy of the injected electrons.

8.3.1

Thermochemical Model of Defect Generation

One of the most popular models used to describe defect generation is the thermochemical model [39–42]. This model is commonly advocated because it is well aligned with the physical evidence (based on analysis of the SBD region using transmission electron microscope (TEM) and electron energy loss spectroscopy

(EELS)) showing that the defects (traps) are fundamentally *oxygen vacancies* that arise from breaking of the Si–O bonds (in SiO_2) and/or Hf–O bonds (in HK HfO_2) [43]. In the thermochemical model, the defect generation rate (k_{GEN}) is given by $k_{\text{GEN}} = \nu \cdot \exp((E_{\text{aG}} - \gamma\xi)/(k_{\text{B}}T))$ where ν , E_{aG} , γ , ξ , k_{B} , and T refer to the lattice vibration frequency, field-free activation energy for Si–O/Hf–O bond breaking, field acceleration factor, applied electric field, Boltzmann constant, and temperature in Kelvin scale, respectively. The parameter γ , which determines the field-induced activation barrier lowering, in turn depends on the permanent dipole moment (p_0) and relative permittivity (κ) of the dielectric by $\gamma = p_0 \left(\frac{2+\kappa}{3} \right)$. The advantage of this model is that all the parameters governing defect generation are material properties, which are well known for most common dielectrics. With the absence of empirical and phenomenological factors, this model is a useful tool for improved reliability design of transistors at the front end from a material's perspective by tuning the critical parameters using a systematic design of experiment and CMOS process parameter modifications (such as annealing temperature and inclusion of dopants).

In spite of its simplicity and association to classical thermodynamics, the thermochemical approach has its own shortcomings when applied to study dielectric BD. It fails to explain the polarity dependence of BD, which is shown by electrical tests with the breakdown voltage (V_{BD}) being higher in the accumulation mode than in the inversion mode for NMOS transistors [44]. This model also fails in explaining the role of the injected current and the energy of the injected carriers. Another issue is that the relevant parameters of the real dielectrics (such as the permittivity and the activation energy) are far off from the ideal values calculated from the first principle studies. Therefore, the model is only as good as the accuracy of the material parameter values. A reliable use of this model would require a detailed calibration exercise on every fabricated stack (taking into account even minor process modifications) to estimate the values of E_{aG} (from Time Dependent Dielectric Breakdown (TDDB) tests at different temperatures), p_0 , and κ (from C - V measurement).

8.3.2

Anode Hydrogen Release Model of Defect Generation

Hydrogen is known to form defects in SiO_2 , and many evidences suggest that hydrogen has a strong impact on defect generation and oxide BD. From a theoretical viewpoint, first-principles calculations have identified the hydrogen bridge as an electron trap with the properties required to explain the stress-induced leakage current, a forerunner for oxide BD [45]. On the other hand, many different experiments have correlated the presence of hydrogen with $t_{63\%}$ (see a short review in Chapter 3 in Ref. [3]). These results suggest that defect generation might be based on the release of H species from the anode interface [46]. By comparing the BD results with experimental results of H desorption by scanning tunneling microscopy (STM) [47], some details of the physics of the H-release process are revealed. In particular, an electronic excitation (EE) mechanism in which an

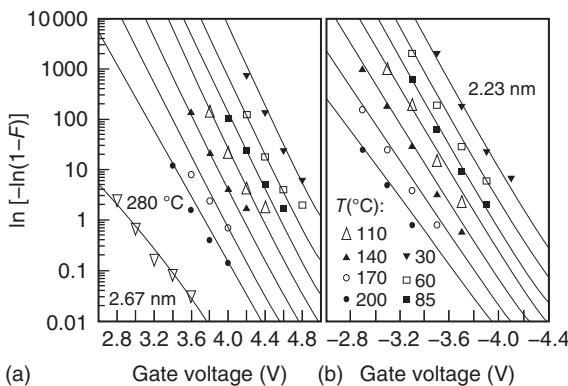


Figure 8.4 Comparison of charge-to-BD data as a function of voltage and temperature with a complete BD model combining percolation statistics and anode hydrogen release model for defect generation. (a) 2.67 nm SiO_2 in p^+ -poly/ SiO_2 /n-Si structures

under substrate injection (positive gate voltage). (b) 2.23 nm SiO_2 in n^+ -poly/ SiO_2 /p-Si structures under gate injection (negative gate voltage). Oxide area is 10^{-4} cm^2 in both cases. (Reprinted with permission from Ref. [46] (© 2013 AIP).)

incident carrier causes direct transition of one electron from the bonding 6σ state to the $6\sigma^*$ antibonding state of the Si–H bond is identified above an electron energy threshold of ~ 6 – 7 eV. Below this energy, H release is assumed to take place via the coupling of the tunneling electrons with the vibrational modes of the Si–H bond, that is, by Si–H vibrational excitations (VEs) [46–48]. If the electron current is high, then incoherent multiple excitation of the bond by N different electrons is considered to dominate. This is known as *incoherent thermal heating mechanism* [47, 48]. On the contrary, if the electron current is small (as the tunnel current in MOS structures under BD stress), the coherent AHR mechanism, which requires only one electron for bond disruption will more probably occur [49]. In both the coherent and incoherent release mechanisms, the H-desorption rate depends on voltage as a power law with an exponent $n \sim 40$, in full agreement with the BD acceleration data [46]. Recently, a general compact AHR model including the temperature dependence of both the hydrogen release process (ζ_1) due to the thermal occupation of excited vibrational states and the thermal activation of the defect generation reaction process (k) has been presented [46]. Coupling this model to the percolation model of BD (see Section 8.4) has allowed tracking of the experimental BD data accurately as a function of stress voltage and temperature as shown in Figure 8.4.

The AHR model might be perceived as a very specific model applicable only to insulating stack structures with an SiO_2 –Si anodic interface but not for other insulators in MIM stacks. However, though the involved physics might be different for different material systems, AHR might be an inspiring example of a defect generation process controlled by the energy of the injected carriers and their interaction with the anode interface.

Although there is still no clear consensus on the best model to describe dielectric BD, it is clear that both field (thermochemical model) and carrier fluence (AHR model) play a role in the bond-breaking process. Therefore, a more robust approach to describe the defect generation mechanism would be to combine these two models under a Markovian framework wherein the field- and fluence-driven mechanisms may be acting in series or parallel with their corresponding state transition rates and unique dependence on the applied voltage and temperature. It remains to be determined whether these two models dominate at different voltage ranges and, if so, how significant their effects are relative to each other in the (V, T) space domain.

The defect generation models discussed have a direct implication on the physics of forming/SET operation. By considering the forming process to be analogous to dielectric BD (with the bottom electrode being the only difference (MIM vs. MIS)) and also by considering the SET process to be the reconnection of a partially broken filament requiring additional defect generation, we can imply that effective control of the forming process requires careful optimization of the dielectric bulk in terms of its bond activation energy, permittivity, and dipole moment (which determine the defect generation rate) as well as the electrode–dielectric interface, which is governed by the interface roughness. Although the CMOS process has been very well optimized to achieve atomically smooth interfaces, this still remains a major bottleneck for ReRAM where the metal–dielectric interfaces tend to exhibit roughness in the range of 1–2 nm [50]. In addition to controllability of the forming process, the reduction of the mean forming voltage (V_{FORM}) is also a key performance metric. This can be achieved by taking advantage of the models described. For example, the dielectric can be purposely annealed to conditions that make it polycrystalline with GBs, so as to reduce the activation energy (E_{aG}) near these dislocation/fault lines. Similarly, the SET process and its variability in terms of the tunnel gap distance during different stochastic reset events requiring different number of additional defects to re-form a percolation path can also be described using the aforementioned same models with changes such as replacing the oxide electric field by the local electric field across the tunnel gap. Another important feature of ReRAM is the shape and size of the CF, wherein the shape would depend on the relative defect generation rates in the bulk and interfaces of the MIM stack and the size depends on the positive feedback process of vacancy generation (analogous to PBD phase) prior to reaching the compliance limit.

8.4 Breakdown and Oxide Failure Statistics

Since there is a broad consensus in relating the BD to defect generation, only the percolation model of the BD statistics is described in this section. In this model, defects are modeled as three-dimensional (3D) objects, which are randomly generated in the oxide and trigger the BD when they overlap to form a percolation path that short-circuits the electrodes [13, 51, 52]. The main achievement of the

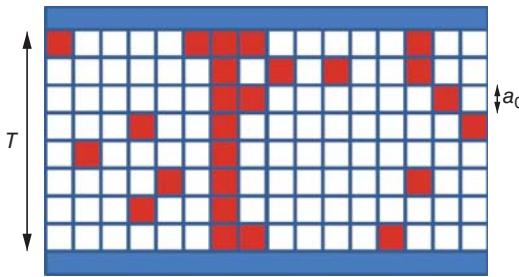


Figure 8.5 Schematic representation of the formation of the BD path in the cell-based percolation model of Ref. [52].

percolation model is the explanation of the T_{ox} dependence of the BD statistics. In this section, we briefly review an analytical version of the percolation model, which provides analytical equations that are mandatory for practical reliability assessment methodology [52].

The oxide volume is divided into 3D cells of area σ and thickness a_0 . Defects are generated at random in these cells. The BD occurs when a complete vertical path of $n_{\text{BD}} = T_{\text{ox}}/a_0$ defective cells is formed, as shown in Figure 8.5. By assuming a uniform density of defects in the oxide volume, $N(t)$, the probability of a cell being defective is $\lambda(t) = 1 - \exp(-N\sigma a_0)$, according to the Poisson distribution. The probability that a percolation path of n_{BD} defective cells is fully formed is $\lambda^{n_{\text{BD}}}$ and, by assuming the weakest-link property, the cumulative survival probability is $1 - F_{\text{BD}} = [1 - \lambda^{n_{\text{BD}}}]^{A_{\text{ox}}/\sigma}$ so that the Weibull of the BD distribution becomes $W_{\text{BD}}(\lambda) = \ln(A_{\text{ox}}/\sigma) + \ln(-\ln(1 - \lambda^{n_{\text{BD}}}))$. Under the conditions of interest for oxide BD reliability (i.e., large oxide areas and low failure percentiles), $\lambda \ll 1$ and the BD distribution is a Weibull distribution:

$$W_{\text{BD}}(N) = \ln\left(\frac{A_{\text{ox}}}{\sigma}\right) + \frac{T_{\text{ox}}}{a_0} \ln(a_0 \sigma N) \quad (8.3)$$

with the scale factor and Weibull slope given by

$$N_{\text{BD}} = \frac{1}{a_0 \sigma} \exp\left[-\frac{a_0}{T_{\text{ox}}} \ln\left(\frac{A_{\text{ox}}}{\sigma}\right)\right] \quad (8.4)$$

and

$$\beta = \frac{T_{\text{ox}}}{a_0}. \quad (8.5)$$

Since the density of defects increases with the injected charge or the stress time as a power law $N \sim Q^\alpha \sim t^\alpha$, the percolation model predicts that the charge-to-BD and time-to-BD distributions are also Weibull distributions. In this case, the Weibull slope scales linearly with the oxide thickness as $\beta = \alpha T_{\text{ox}}/a_0$, in very good agreement with the experiments. The scaling of $t_{63\%}$ with T_{ox} is also well captured by this analytical percolation model [52]. Moreover, successful generalizations including multiple types and arrangement of defects [53] and multilayer dielectrics have also been recently documented [54–56].

Percolation models have also been modified to replicate the role played by the dielectric microstructure to simulate the effect of columnar GBs, which are dislocations that could exist when the dielectric becomes polycrystalline as a result of high annealing temperatures. STM studies have shown [57, 58] that these GBs serve as favorite spots for early percolation due to preferential defect generation and lower activation energy for vacancy generation and diffusion [59]. Upon incorporation of this nonideality into the percolation model by having GB and grain columns in the cell matrix with different defect generation rates, the stochastics of failure becomes non-Weibull [42, 60] with convex trends (high β at low failure percentiles). This is also confirmed by electrical device level [60] and STM-localized stress BD tests [61] on polycrystalline HK stacks. Such non-Weibull trends are also observed for CF formation in dual-layer dielectric stacks where BD of the first layer localizes the percolation of the second layer, going against the area scaling rule.

8.5

Implications of Breakdown Statistics for ReRAM

The BD statistics can be a source of inspiration to deal with different aspects of ReRAM performance and reliability. In this regard, there is wide agreement in considering the forming of filamentary ReRAM as a current-limited SBD event. Thus, the percolation model is likely to be adequate to describe the forming statistics. It is worth noting that, while only the low-percentile region of the failure distribution (where the Weibull model is asymptotically valid) is of interest for oxide reliability, both the low- and high-percentile regions need to be well modeled in the case of ReRAM forming. Highly efficient forming is required to ensure that “all” the devices are formed under the forming stress, and this is related to the high percentile region of the distribution. On the other hand, aiming at low-disturb probability under normal operation, that is, to avoid the creation of new CFs during SET/RESET cycling, is related to the low percentiles. Voltage and temperature acceleration of BD might also be a useful background to improve the forming methods. Since there is a trend toward using thinner dielectrics in ReRAM devices, understanding the boundary between stable SBD and PBD modes might also be an important piece of information to design the ReRAM structures and the forming/SET procedures.

Based on the idea that in the HRS, the CF can be described by two stumps with a gap of insulating material in between (Figure 8.6), BD has also been related to the SET transition in filamentary ReRAM [10, 11]. Departing from the percolation model of oxide BD, a model was proposed for the distribution of SET voltage [11]. By taking into account the condition that the CF gap thickness after RESET changes in a wide range, as reflected in the variations of R_{OFF} during cycling, the data needs to be screened as a function of R_{OFF} , and hence, a very large statistical sample size is required. Fortunately, this is possible because a large number of SET voltage readings can be obtained in a single sample by repetitive

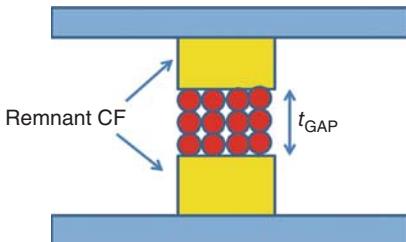


Figure 8.6 Application of the percolation model to the SET transition in ReRAM. A cell-based geometrical model is proposed to describe the insulator gap in the broken CF. SET transition is modeled as the BD of the insulator gap. With the percolation approach, the R_{OFF} dependence on the scale and shape factors of the SET voltage distribution are well captured.

SET/RESET cycling. On the contrary, in the case of studying the BD of an oxide film (or forming in ReRAM), a different sample is needed for each BD (forming) data point. By relating the resistance of the HRS to the CF gap thickness through the quantum point contact (QPC) model of conduction, both the scale factor (V_{BD}) and the Weibull slope (β_{RVS}) were predicted to be proportional to $\log(R_{OFF})$, in full agreement with the experimental data. This result strongly supports the application of the oxide BD percolation model to the SET statistics in ReRAM. A similar methodology of cell-based description of the CF constriction was used in a model for unipolar RESET [62]. In this case, however, the analogy with oxide BD is by far less evident, although certain symmetry can be established between BD and recovery processes. The application of the percolation model to the SET transition has also been thoroughly explored in ReRAM structures based on TiO_2 and HfO_2 [10]. These authors showed that a Weibull model perfectly fits the t_{set} and V_{set} distributions under pulsed (CVS) and quasistatic (RVS) SET conditions, respectively. By combining a pulse setup and a DC CVS setup, they explored the distribution of t_{set} and demonstrated a constant Weibull slope across 8 orders of magnitude in time. By changing the pulse voltage amplitude, they also demonstrated the validity of the power-law voltage acceleration model across 10 orders of magnitude in t_{set} , reporting a power-law exponent $n \sim 26$ for $Ni/HfO_2/Si$ and $n \sim 49$ for $Ti/TiO_2/Pt$ structures. Moreover, they checked the validity of the relations between the distributions of t_{set} (CVS) and V_{set} (RVS), exactly as established previously for the BD statistics [15].

In Luo's work [10], not only the feasibility of the percolation model was explored in detail, but it was also used as a prediction/design methodology to approach the SET speed–disturb issue. The disturb of the HRS is analogous to the SET transition except that it occurs at a lower disturb voltage (V_{dis}) rather than the higher SET voltage (V_{set}). In practice, V_{dis} can be the read voltage or, in a crossbar ReRAM, the write disturb voltage, which is at least $V_{set}/3$ if a $V/3$ scheme is used [12]. The desired SET speed ($t_{set} < 1\mu s$, for example) and disturb immunity ($t_{dis} > 1\text{ s}$) can be theoretically achieved, thanks to the highly nonlinear voltage–time dependence of the SET process. However, exploration of the design space needs to take into account the significant statistical spread of the SET variables. As shown in Figure 8.7, the time distributions corresponding to V_{set} and V_{dis} (obtained from the measured SET data using the power-law voltage acceleration) are required to meet a 1 ppm specification for efficient SET and low disturb, respectively. By using this methodology, the influence of n , $V_{63\%}$, and β_{RVS} on the design space of V_{set}

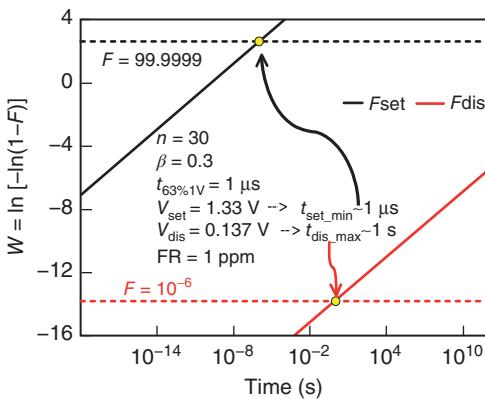


Figure 8.7 Application of the BD statistical methodology to deal with the SET speed/READ disturb trade-off in ReRAM. From the measured time-to-SET distribution, and having characterized the voltage acceleration power law, the distribution of time-to-SET (t_{set}) and time-to-disturb (t_{dis}) can be

obtained for any SET and READ voltages, respectively. The high percentile region of the t_{set} distribution is important to ensure a highly efficient SET process, while the low-percentile tail of the t_{dis} distribution is crucial to ensure low disturb probability, as discussed in detail in Ref. [10].

and V_{dis} at any failure rate can be explored [10]. As an example, Figure 8.8 shows the minimum Weibull slope of the V_{set} distribution required for an efficient SET (< 1 ppm failure rate) with 1 μ s pulses at an amplitude of $V_{\text{set}} > 2$ V and a low disturb probability with read pulses of 1 s and $V_{\text{dis}} < 0.25$ V. The Weibull slope requirements are plotted against the SET voltage scale factor $V_{63\%}$ measured under RVS with a ramp rate of 1 V s^{-1} . To obtain a reasonably wide design space,

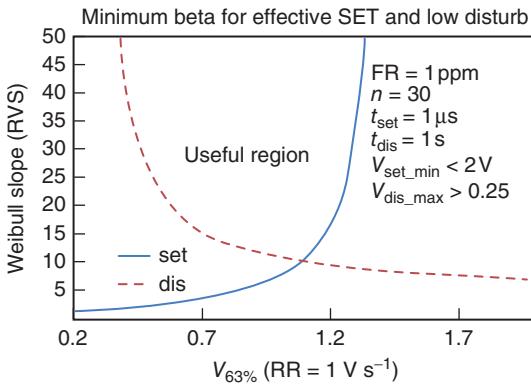


Figure 8.8 Application of the BD statistical methodology allows for establishing the design requirements to ReRAM trade-offs [10]. In this figure, the minimum Weibull slope of the SET voltage distribution required for an efficient SET with low READ

disturb is plotted versus the scale factor of the V_{set} distribution under VRS conditions. Only above both the solid and the dashed lines is it possible to simultaneously meet the specifications for reliable SET and low disturb, thus defining a useful design area.

this plot shows that $\beta_{RVS} > 20$ is required for the assumed voltage acceleration exponent ($n = 30$). Luo showed that $\beta_{RVS} > 20$ and $n > 20$ are usually required by the SET speed–read disturb trade-off, and his review of the published data revealed that the first requirement is very rarely achieved in common switching oxides.

The successive BD statistics has also been suggested to be of interest for ReRAM endurance [7, 8]. During the forming process, one aims at creating only one CF, which is expected to be switched on and off during a large number of SET/RESET cycles. The generation of a new CF during any SET cycle might cause device failure by deeply changing the switching properties or worsen the ReRAM performance by degrading the R_{ON}/R_{OFF} window and introducing additional variability issues. During the forming process, all the devices need to be formed (ideally) with only one active CF. The statistics of successive BD events will help in determining the probability of generation of more than one CF during the forming process. On the other hand, Degraeve *et al.* [7] suggested that one of the limits of ReRAM endurance might be related to the generation of a second CF during SET. Of course, since the maximum SET voltage is usually much smaller than the forming voltage, the probability to generate new CFs is expected to be low. However, since the number of cycles and the number of ReRAM cells in large nonvolatile memory arrays are very large, even a very low CF generation probability under SET conditions might cause a certain percentile of failure. In this regard, the theory of successive BD statistics is also relevant to this ReRAM disturb (reliability) problem. However, we must be very careful on how we apply this theory because the boundary conditions that define the problem are radically different when dealing with the ReRAM endurance limit. In the case of oxide reliability, we are interested in the distribution of the K th BD event at very low failure percentiles, that is, when most of the samples have not suffered any BD event. On the contrary, in the case of ReRAM, all the devices have been stressed to the first BD event during the forming process because at least one CF is required to activate the switching phenomena. Thus, the oxide might be substantially degraded before it is effectively used for ReRAM application. If we assume that the formation of the second CF causes failure, estimation of the lifetime (endurance) of the ReRAM structures requires the calculation of the total stress time elapsed between electroforming and the unwanted formation of a second CF during the application of successive SET pulses. The distribution of time elapsed between the first and the second SBD event cannot be calculated analytically, but an approximate model was developed [8], which effectively tracks the results of Monte Carlo (MC) simulation (see Figure 8.9). This model allows for establishing the constraints imposed by this failure mode to the SET voltage distribution and to the SET speed.

The percolation model developed for defective dielectric with GBs can be very useful for modeling the forming distribution. This is because polycrystalline dielectrics are being explored as a useful approach to reduce the forming voltage and power needed to create the CF [63]. While V_{FORM} has to be reduced, its distribution has to be steepened to ensure adequate immunity to read disturb.

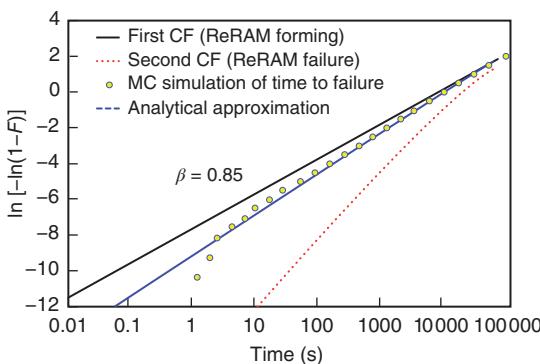


Figure 8.9 Application of the successive BD theory to consider ReRAM failure due to the formation of a second CF during SET/RESET cycling. The relevant time to failure is the time between the first BD (under forming at $V_{forming}$) and the second one (during cycling at V_{set}).

Tuning the distribution can be accomplished by combining the GB-based percolation model along with a physical model for defect generation. The GB model may also be applied for SET stage modeling if the filament does not rupture from one end to the other, and instead, the defect recovery also occurs by a random or pseudorandom process resulting in fragmented and disconnected defect clusters that can resemble a GB in the RESET (OFF) state. Note that the traditional percolation model is sufficient enough to describe the forming event in “amorphous” dielectric films with defect generation being spatially uniform.

The multilayer percolation model [54, 56] also finds application in the variability modeling of forming and SET distributions for ReRAM studies that have recently considered such multiple layers for the MIM stack (possibly enabling multibit storage). The presence of multiple layers of dielectric along with a careful compliance capping technique (e.g., using a transistor in series with the MIM, called as the 1-transistor-1-resistor (1T-1R structure)) can enable realization of ultralow-power switching devices [64] because the percolation can be confined to just one of the dielectrics (as demonstrated recently in MOS devices by Raghavan *et al.* [65]), keeping the other one intact by careful engineering of the material properties and thickness of both the dielectric layers.

8.6

Chemistry of the Breakdown Path and Inference on Filament Formation

While all the previous sections dealt with analysis of BD from a purely statistical and electrical perspective, there is still a need for use of advanced microscopic tools and techniques that can enable us to decipher the atomic phenomenon that governs the formation of a percolation path in MOS stacks (or, equivalently, the nucleation of a CF in MIM stacks). Although very challenging and having a low success rate in locating the nanoscale BD spot, a physical analysis of BD can

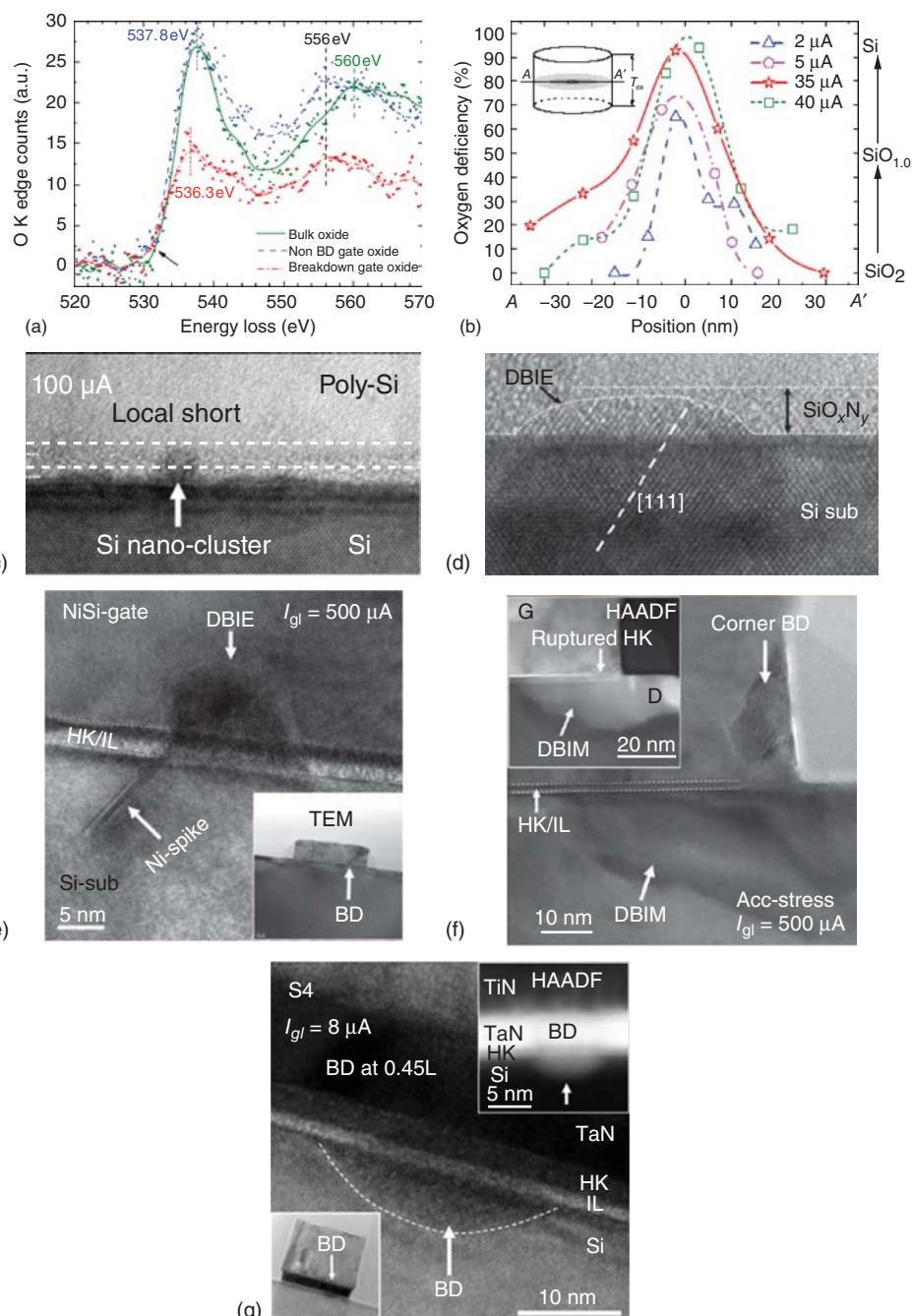


Figure 8.10 (a,b) Oxygen deficiency count in the percolated region of the dielectric in poly-Si–SiON based gate stack. With higher compliance, the percolation region laterally dilates in size and the central core of the percolation path becomes increasingly Si-rich with the oxygen stoichiometry shifting from $x \sim 2$ to 0 [67, 68]. (c,d) Physical TEM image showing the formation of the Si nanowire [23] and DBIE epitaxial defect due to silicon protrusion from the substrate (maintaining the same substrate orientation) [70]. (e) TEM image of HBD ($I_{\text{comp}} \sim 500 \mu\text{A}$) for a NiSi–HfSiON– SiO_x (bilayer)–Si stack showing Ni spiking into the dielectric and all the way into the substrate in the (111) direction [71]. Note that the size of the filament is as small as 2–3 nm. (f) TEM image

showing another failure mechanism in the NiSi gate stack involving the lateral diffusion of Ni from the source/drain contacts into the channel region for $I_{\text{comp}} \sim 500 \mu\text{A}$ [71]. This lateral diffusion is more clearly detected in the high-angle annular dark field (HAADF) image in the inset of (f). (g) TEM image of filamentation in the TaN (TiN)– HfO_2 –Si stack showing migration of Ta into the substrate, forming a bowl-shaped defect. Note that the size of the Ta filament is much larger with a diameter of ~ 20 –30 nm [73]. In this case, it is to be noted that, although the chosen I_{comp} is as low as 8 μA , there is a compliance overshoot effect as the TDDB transient was too fast to be adequately controlled by the external compliance setting.

provide us with a wealth of information and address any speculative conclusions that our electrical and statistical results convey. The typical methodology for the physical analysis of the MOS transistor (after percolation) involves identification of the BD spot location (s_{BD}) along the length of the transistor using the weighted ratio of the source and drain currents ($s_{\text{BD}} = I_d/(I_d + I_s)$) in the accumulation mode as presented by Degraeve *et al.* [66] (this is a major advantage in MOS transistor stacks, as compared to MIM capacitors where such electrical estimation of the BD/CF location is not possible). Following this, focused-ion beam (FIB) milling is performed to provide access to the failed device and TEM investigations are carried out. The elemental constitution of the BD region is analyzed using the EELS for light elements such as O, N, Si, and energy-dispersive X-ray (EDX) spectroscopy for heavy elements such as Ni, Ta, and Ti. We shall now present the defect chemistry for MOS devices based on different gate electrode materials (polysilicon, Ni-based, and Ti/Ta-based) for two vastly different regimes of low (1–5 μA) and high current compliance (100 μA –1 mA).

Table 8.1 presents a summary of the various gate stacks investigated, the different defect signatures observed, as well as the driving forces governing the observed morphology for these failure mechanisms. Figure 8.10 shows the corresponding TEM and EELS/EDX images for the various stacks (polysilicon, Ni-based, and Ta-based, respectively) with the dielectric being SiO_2 or HK with SiO_x (interfacial layer).

For all the three gate stacks, when the compliance is capped to the SBD regime with $I_{\text{comp}} < 1$ –2 μA , the percolation path is a cylindrical volume of oxygen vacancies (or oxygen deficiency, in general, which could also be due to cation interstitials in various metal oxides) (Figure 8.10a,b) [67, 68], without any major morphological changes to the dielectric. The cylindrical shape of the percolated region indicates that defect generation until SBD is uniform all along the dielectric between the gate and substrate (supporting our Poisson approximation during

Table 8.1 Summary of the failure defects observed in various gate material–SiON/high- κ stacks and their driving forces.

Gate stack	Failure defect in oxide	Driving forces
Poly-Si–SiON/HK [67–70]	Dielectric-breakdown-induced epitaxy (DBIE) Si nanowire filament Percolation path dilation Grain-boundary-assisted breakdown (high- κ)	Current density Thermomigration Grain boundary – faster diffusion and enhanced conduction path
NiSi gate–SiON/HK [71, 72]	Dielectric-breakdown-induced epitaxy (DBIE) DBIM (metal migration in Ni/Co/Ti silicides) Nickel spiking	Temperature gradient Concentration gradient
TaN gate–SiON/HK [73]	Metal filamentation Metal migration into substrate	Ionic conduction Electron wind force Ionic conduction Hole migration

statistical modeling). The initial size of the percolation path for a very SBD is $\sim 10\text{--}15\text{ nm}$ [68], in remarkable agreement with the spot size inferred from the analysis of the BD statistics with the percolation model [74].

In the case of the polysilicon stack (which was used for 0.18 and 0.13 μm CMOS technology node), with increasing compliance for the BD, we observe a lateral dilation of the BD path from 10 nm diameter ($I_{\text{comp}} \sim 1\text{--}2\text{ }\mu\text{A}$) up to 40–50 nm for $I_{\text{comp}} \sim 30\text{--}50\text{ }\mu\text{A}$ (Figure 8.10a,b) [68]. Additionally, the central core of the percolation path becomes increasingly silicon-rich (depletion of more oxygen) with $x \rightarrow 0$ in SiO_x . In other words, the vacancy concentration (indicated by the sub-stoichiometry) gradually increases from the outer shells to the inner core of the percolated region. One of the defects in the harder BD stage for poly-Si stack is the silicon nanowire that is formed in the oxide due to complete depletion of oxygen (Figure 8.10c) [23]. Another defect that is commonly observed is the bulging of the dielectric–substrate interface wherein the silicon starts to protrude into the oxide at the BD spot. This defect was first observed by Pey *et al.* [69, 70] about 10 years back and was termed as “dielectric-breakdown-induced epitaxy (DBIE)” [69] (Figure 8.10d) because the silicon continued to maintain its crystallographic (100) orientation while protruding into the oxide. The cause of this unanticipated failure mechanism is the electro–thermal migration of silicon through the percolation path, driven by the high current density (as high as 10 mA cm^{-2}) and joule heating temperature. As a result of this defect signature, the dielectric thickness is effectively reduced by the protruding silicon. This further enhances the electric field, current density, and temperature setting up a positive feedback loop, finally

resulting in a thermal runaway and very severe HBD of the dielectric. The DBIE defect has also been observed and confirmed by industrial research groups based at Toshiba® [75], ST Microelectronics® [76], and Intel® [77].

Moving on to NiSi-based electrode stacks (used in 90 nm CMOS technology node), for $I_{\text{comp}} < 10 \mu\text{A}$, we observe the Ni to be spiking (punch-through) through the dielectric in the percolation path region (Figure 8.10e) and diffusing all the way into the substrate [71]. The diameter of the Ni filament in the dielectric is as low as 2–3 nm. [71]. Although the DBIE defect from the Si substrate may still exist here, the dominant mechanism is the migration of atoms/ions from the metal gate electrode. This spiking nature of Ni was also reported and confirmed by silicide process studies under high annealing conditions [78]. In the post-percolation stage, the local temperature and field are sufficient enough (analogous to conditions of process anneal conditions) for highly mobile Ni to diffuse all the way into the dielectric. Additionally, we have also detected cases where the Ni from the silicide drain and source contacts happened to “laterally diffuse” through the interfaces of silicon with the dielectric and into the channel region (Figure 8.10f), resulting in a “channel short” as an extreme scenario leading to complete transistor malfunction. This defect is referred to as dielectric-breakdown-induced metal migration (DBIM) [72], and the same has also been observed [79] as a phenomenon that occurs in a nonoptimized silicide contact process flow for CMOS technology.

Finally, we consider the most advanced gate stack corresponding to the 65, 45, and 32 nm technology nodes comprising TaN or TiN as the gate electrode. Analysis of HBD in this material stack also reveals migration of the Ta into the oxide and the substrate. What is worth noting is that the filaments here are bowl-shaped [73] (Figure 8.10g) and large in size with a diameter of 15–20 nm. This is contrast to the sharp and thin Ni filament with larger surface-area-to-volume ratio. The direction of metal ion migration for the Ni and Ta filaments seems to originate at the anode side, opposite to the electron wind force. Therefore, we attribute this to be due to positive ion (Ni^+) and/or hole migration [73].

To summarize the findings from the TEM-EELS-EDX study, it is clear that all SBD events up to $I_{\text{comp}} \leq 5 \mu\text{A}$ involve only the formation of a percolation path of oxygen vacancies, irrespective of the material stack. However, the failure mechanism becomes highly electrode-dependent for harder stages of BD from $I_{\text{comp}} \geq 20–50 \mu\text{A}$. Therefore, it is the electrode material that governs the failure mechanism at HBD [80], not the dielectric. Our studies have shown that the failure defects remain the same when the electrode is fixed, but the dielectric is changed from SiO_2 to HK. It is worth noting that the local temperature in these metal filaments (HBD stage) can be as high as 1000–1500 °C [81] due to the very high current density and lack of effective heat sink.

Before we conclude, the role of the dielectric microstructure in the BD (forming) phenomenon deserves some attention. When a localized TDDB stress is induced on various grain and GB sites using the STM nanoprobe, we observe a drastic reduction in the TDDB lifetime. Although this is detrimental for logic devices where a higher V_{BD} is always desired, the presence of GB (which can

be intentionally induced by choosing high annealing temperature after HK deposition) in ReRAM helps reduce the V_{FORM} . It is, therefore, favorable to have GB for enabling low-power resistive switching. However, difficulty arises in our ability to control the grain size of the HK. STM studies on HfO_2 have shown the grain diameter to have a mean of 22 nm with a standard deviation of 5–10 nm [27]. As the ReRAM device area is reduced from $10 \times 10 \mu\text{m}^2$ to $10 \times 10 \text{ nm}^2$, there is a higher probability of finding zero GB in some devices and more than 1 GB in others. Due to this large spread in the number of GB fault lines in the small-area devices, the variability in V_{FORM} is expected to become larger. Further research on precise control of grain size distribution is necessary before the potential of this concept can be realized.

8.7

Summary and Conclusions

In this chapter, a detailed insight into the dielectric BD process has been presented from three different perspectives – electrical, statistical, and physical. The results from all these BD investigations were examined, and their link and implication to ReRAM were established. The use of different techniques to study BD has enabled us to unveil the atomistic phenomena governing the CF formation. The percolation model was used as the backbone to help explain the distributions for forming and SET events. It may also help in describing the endurance and retention failure distributions for ReRAM. The different stages of BD were also discussed, and the electrical leakage signatures provided valuable insight into the necessity for adequate control of the CF formation process using internal/external compliance setting. The difficulty to achieve low-power switching in thick dielectrics was explained in the absence of the PBD stage (as SBD for these cases is already quite catastrophic). The different physical models proposed for defect generation were also highlighted, and their applicability to modeling the CF formation process was introduced. We hope that this chapter has served its purpose of bridging the two very different research domains of logic (where BD causes a reliability problem) and memory ReRAM technology (where the formation and partial destruction of a CF are intrinsic to device operation). There are many issues that still need to be addressed in ReRAM, and understanding the link between BD and the switching filament formation would always help transfer knowledge from one domain to the other. However, it is necessary to exercise caution in this process because the bottom electrode for logic (semiconducting Si) and memory (metal film) devices are completely different. The bottom electrode could play a critical role depending on the polarity of device operation. There is still tremendous potential for applying the existing percolation model (in combination with ionic transport models) and carefully tuning it to phenomenologically describe many features of ReRAM including endurance degradation, retention loss, SET–RESET stochastics, and read disturb phenomena.

References

1. Lombardo, S., Stathis, J.H., Linder, B.P., Pey, K.L., Palumbo, F., and Tung, C.H. (2005) Dielectric breakdown mechanisms in gate oxides. *J. Appl. Phys.*, **98**, 121301.
2. Robertson, J. (2006) High dielectric constant gate oxides for metal oxide Si transistors. *Rep. Prog. Phys.*, **69**, 327–396.
3. Strong, A., Wu, E.Y., Vollertsen, R.-P., Suñé, J., La Rosa, G., Rauch, S.E. III., and Sullivan, T.D. (2009) *Reliability Wearout Mechanisms in Advanced CMOS Technologies*, IEEE Press Series on Microelectronic Systems, Wiley-IEEE Press.
4. Wu, E.Y. and Vollertsen, R.P. (2002) On the Weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part I: theory, methodology, experimental techniques, *IEEE Trans. Electron Devices*, **49**, 2131–2140.
5. Suehle, J.S. (2009) in *Defects in Micro-electronic Materials and Devices* (eds D.M. Fleetwood, S.T. Pantelides, and R.D. Schrimpf), CRC Press, pp. 437–463.
6. McPherson, J.W. (2012) Time dependent dielectric breakdown physics – models revisited. *Microelectron. Reliab.*, **52**, 1753–1760.
7. Degraeve, R., Roussel, P.H., Goux, L., Wouters, D., Kittl, J., Altimime, L., Jurczak, M., and Groeseneken, G. (2010) Generic learning of TDDB applied to RRAM for improved understanding of conduction and switching mechanism through multiple filaments. IEEE International Electron Devices Meeting (IEDM), pp. 633–636.
8. Suñé, J., Miranda, E., Jiménez, D., Long, S., and Liu, M. (2011) From dielectric failure to memory function: learning from oxide breakdown for improved understanding of resistive switching memories. 11th Non-Volatile Memory Technology Symposium (NVMTS'11), Shanghai, China, November 2011, pp. 1–6.
9. Pey, K.L., Raghavan, N., Wu, W., Liu, W., and Bosman, M. (2012) Dielectric breakdown-recovery in logic and resistive switching in memory-bridging the gap between the two phenomena, 11th International Conference on Solid State and Integrated Circuit Technology (ICSICT), Xian, China, Invited, pp. 1–6.
10. Luo, W.-C., Liu, J.-C., Feng, H.-T., Lin, Y.-C., Huang, J.-J., Lin, K.-L., and Hou, T.-H. (2012) RRAM SET speed-disturb dilemma and rapid statistical prediction methodology. IEEE International Electron Devices Meeting (IEDM), pp. 215–218.
11. Long, S., Lian, X., Cagli, C., Perniola, L., Miranda, E., Liu, M., and Suñé, J. (2013) A model for the set statistics of RRAM inspired in the percolation model of oxide breakdown. *IEEE Electron Device Lett.*, **34**, 999–1001.
12. Alam, M.A. (2002) Statistically independent soft-breakdowns redefine oxide reliability specifications. IEEE International Electron Devices Meeting (IEDM), pp. 151–154.
13. Stathis, J.H. (1999) Percolation models for gate oxide breakdown. *J. Appl. Phys.*, **86**, 5757–5766.
14. Wu, E.Y., Suñé, J., and Lai, W. (2002) On the Weibull shape factor of dielectric films and its accurate experimental determination, part II. Experimental results and the effects of stress conditions. *IEEE Trans. Electron Devices*, **49**, 2141–2150.
15. Kerber, A., Pantisano, L., Veloso, A., Groeseneken, G., and Kerber, M. (2007) Reliability screening of high- γ dielectrics based on voltage ramp stress. *Microelectron. Reliab.*, **47**, 513–517.
16. Wu, E. and Suñé, J. (2005) Power-law voltage acceleration: a key element for ultra-thin gate oxide reliability. *Microelectron. Reliab.*, **45**, 1809–1834.
17. Vollertsen, R.P. and Wu, E.Y. (2003) Gate oxide reliability parameters in the range 1.6 to 10 nm. IEEE International Integrated Reliability Workshop, pp. 10–15.
18. Miranda, E. and Suñé, J. (2001) Analytic modeling of leakage current through multiple breakdown paths in SiO₂ films. IEEE International Reliability Physics Symposium (IRPS), pp. 367–379.

19. Linder, B.P., Lombardo, S., Stathis, J.H., Vayshenker, A., and Frank, D. (2002) Voltage dependence of hard breakdown growth and the reliability implication in thin dielectric. *IEEE Electron Device Lett.*, **23**, 661–663.
20. Sahhaf, S., Degraeve, R., Roussel, P.J., Kauerlauf, T., Kaczer, B., and Groeseneken, G. (2007) TDDB reliability prediction based on the statistical analysis of hard breakdown including multiple soft breakdown and wear-out. IEEE International Electron Devices Meeting (IEDM), pp. 501–504.
21. Lo, V.L., Pey, K.L., Tung, C.H., and Li, X. (2007) Multiple digital breakdowns and its consequence on ultrathin gate dielectrics reliability prediction. IEEE International Electron Devices Meeting (IEDM), Washington, DC, pp. 497–500.
22. Lo, V.L., Pey, K.L., Tung, C.H., and Ang, D.S. (2007) A critical gate voltage triggering irreversible gate dielectric degradation. IEEE International Reliability Physics Symposium (IRPS), pp. 576–577.
23. Raghavan, N., Padovani, A., Li, X., Wu, X., Lo, V.L., Bosman, M., Larcher, L., and Pey, K.L. (2013) Resilience of ultrathin oxynitride films to percolative wear-out and reliability implications for high- κ stacks at low voltage stress. *J. Appl. Phys.*, **114**, 094504.
24. Wolters, D.R. and van der Schoot, J.J. (1985) Dielectric breakdown in MOS devices, part II: conditions for the intrinsic breakdown. *Philips J. Res.*, **40**, 137.
25. Wolters, D.R. and van der Schoot, J.J. (1985) Dielectric breakdown in MOS devices, part III: conditions for the intrinsic breakdown. *Philips J. Res.*, **40**, 164.
26. Fischetti, M.V., DiMaria, D.J., Brorson, S.D., Theis, T.N., and Kitley, J.R. (1985) Theory of high-field electron transport in silicon dioxide. *Phys. Rev. B*, **31** (12), 8124–8142.
27. Ong, Y.C., Ang, D.S., Pey, K.L., Oshea, S.J., Goh, K.E.J., Troadec, C., Tung, C.H., Kawanago, T., Kakushima, K., and Iwai, H. (2007) Bilayer gate dielectric study by scanning tunneling microscopy. *Appl. Phys. Lett.*, **91**, 102905.
28. Suñé, J. and Wu, E. (2005) Temperature-dependent transition to progressive breakdown in thin silicon dioxide based gate dielectrics. *Appl. Phys. Lett.*, **86**, 193502.
29. Kauerlauf, T., Degraeve, R., Zahid, M.B., Cho, M., Kaczer, B., Roussel, P., Groeseneken, G., Maes, H., and De Gendt, S. (2005) Abrupt breakdown in dielectric/metal gate stacks: a potential reliability limitation? *IEEE Electron Device Lett.*, **26**, 773–775.
30. Pey, K.L., Raghavan, N., Li, X., Liu, W.H., Shubhakar, K., Wu, X., and Bosman, M. (2010) New insight into TDDB and post breakdown reliability of novel high- κ gate dielectric stacks. IEEE International Reliability Physics Symposium (IRPS), Invited Paper, Anaheim, CA, pp. 354–363.
31. Ielmini, D., Nardi, F., and Balatti, S. (2012) Evidence for voltage-driven set/reset processes in bipolar switching RRAM. *IEEE Trans. Electron Devices*, **59**, 2049–2056.
32. Pey, K.L., Tung, C.H., Ranjan, R., Lo, V.L., MacKenzie, M., and Craven, A.J. (2007) Nano-characterisation of dielectric breakdown in the various advanced gate stack MOSFETs. *Int. J. Nanotechnol.*, **4**, 347–376.
33. McPherson, J.W. and Mogul, H.C. (1998) Disturbed bonding states in SiO₂ thin-films and their impact on time-dependent dielectric breakdown. IEEE International Reliability Physics Symposium (IRPS), pp. 47–56.
34. Vogel, E.M., Suehle, J.S., Edelstein, M.D., Wang, B., Chen, Y., and Bernstein, J.B. (2000) Reliability of ultra-thin silicon dioxide under combined substrate hot-electron and constant voltage tunneling stress. *IEEE Trans. Electron Devices*, **47**, 1183–1191.
35. Wu, E.Y., Stathis, J.H., and Liang-Kai, H. (2000) Ultra-thin oxide reliability for ULSI applications. *Semicond. Sci. Technol.*, **15**, 425–435.
36. Schuegraf, K.F. and Hu, C. (1994) Hole injection SiO₂ breakdown model for very low voltage lifetime extrapolation. *IEEE Trans. Electron Devices*, **41**, 761–767.

37. Degraeve, R., Groeseneken, G., Bellens, R., Ogier, J.L., Depas, M., Roussel, P.J., and Maes, H.E. (1995) A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides. IEEE International Electron Devices Meeting (IEDM), pp. 866–869.
38. Nicollian, P.E., Krishnan, A.T., Chancellor, C.A., and Khamankar, R.B. (2006) The traps that causes breakdown in deeply scaled SiON Dielectrics. IEEE International Electron Devices Meeting (IEDM), pp. 743–746.
39. McPherson, J.W. (2007) Quantum mechanical treatment of Si-O bond breakage in silica under time dependent dielectric breakdown testing. IEEE International Reliability Physics Symposium (IRPS), pp. 209–216.
40. Bersuker, G., Korkin, A., Fonseca, L., Safonov, A., Bagatur'yants, A., and Huff, H.R. (2003) The role of localized states in the degradation of thin gate oxides. *Microelectron. Eng.*, **69** (2-4), 118–129.
41. Vandelli, L., Padovani, A., Larcher, L., Bersuker, G., Jung, Y., and Pavan, P. (2011) A physics-based model of the dielectric breakdown in HfO₂ for statistical reliability prediction. IEEE International Reliability Physics Symposium (IRPS), pp. 807–810.
42. Raghavan, N., Pey, K.L., Shubhakar, K., Wu, X., Liu, W.H., and Bosman, M. (2012) Role of grain boundary percolative defects and localized trap generation on the reliability statistics of high- κ gate dielectric stacks. IEEE International Reliability Physics Symposium (IRPS), Anaheim, Orange, CA, pp. 6A.1.1–6A.1.11.
43. Li, X., Tung, C.H., Pey, K.L., and Lo, V.L. (2008) The chemistry of gate dielectric breakdown. IEEE International Electron Devices Meeting (IEDM), pp. 1–4.
44. Prasad, C., Agostinelli, M., Auth, C., Brazier, M., Chau, R., Dewey, G., Ghani, T., Hattendorf, M., Hicks, J., Jopling, J., Kavalieros, J., Kotlyar, R., Kuhn, M., Kuhn, K., Maiz, J., McIntyre, B., Metz, M., Mistry, K., Pae, S., Rachmady, W., Ramey, S., Roskowski, A., Sandford, J., Thomas, C., Wiegand, C., and Wiedemer, J. (2008) Dielectric breakdown in a 45 nm high- κ /metal gate process technology. IEEE International Reliability Physics Symposium, (IRPS), pp. 667–668.
45. Blöchl, P.E. and Stathis, J.H. (1999) Hydrogen electrochemistry and stress-induced leakage current in silica. *Phys. Rev. Lett.*, **83**, 372–375.
46. Wu, E.Y. and Suñé, J. (2013) Generalized hydrogen release-reaction model for the breakdown of modern gate dielectrics. *J. Appl. Phys.*, **114**, 014103.
47. Shen, T.-C., Wang, C., Abeln, G.C., Tucker, J.R., Lyding, J.W., Avouris, P., and Walkup, R.E. (1995) Atomic-scale desorption through electronic and vibrational excitation mechanisms. *Science*, **268**, 1590–1592.
48. Ribes, G., Bruyere, S., Denais, M., Roy, D., and Ghibaudo, G. (2004) Modeling charge-to-breakdown using hydrogen multivibrational excitation (thin SiO₂ and high- γ dielectrics). IEEE International Integrated Reliability Workshop (IIRW), Final Report, pp. 1–3.
49. Salam, G.P., Persson, M., and Palmer, R.E. (1994) Possibility of coherent multiple excitation in atom transfer with a scanning tunneling microscope. *Phys. Rev. B*, **49**, 10655–10662.
50. Chang, W.Y., Huang, H.W., Wang, W.T., Hou, C.H., Chueh, Y.L., and He, J.H. (2012) High uniformity of resistive switching characteristics in a Cr/ZnO/Pt device. *J. Electrochem. Soc.*, **159** (3), G29–G32.
51. Degraeve, R., Groeseneken, G., Bellens, R., Ogier, J.L., Depas, M., Roussel, P.J., and Maes, H.E. (1998) New insights in the relation between electron trap generation and the statistical properties of oxide breakdown. *IEEE Trans. Electron Devices*, **45**, 904–911.
52. Suñé, J. (2001) New physics-based analytic approach to the thin oxide breakdown statistics. *IEEE Electron Device Lett.*, **22**, 296–298.
53. Krishnan, A.T. and Nicollian, P.E. (2007) Analytic extension of the cell-based oxide breakdown model to full percolation and its implications. IEEE International Reliability Physics Symposium (IRPS), pp. 232–239.

54. Suñé, J., Tous, S., and Wu, E.Y. (2009) Analytical cell-based model for the breakdown statistics of multilayer insulator stacks. *IEEE Electron Device Lett.*, **30** (12), 1359–1361.
55. Blonkowski, S. (2010) Filamentary model of dielectric breakdown. *J. Appl. Phys.*, **107**, 084109.
56. Raghavan, N., Pey, K.L., Liu, W.H., and Li, X. (2010) New statistical model to decode the reliability and Weibull Slope of high- κ and interfacial layer in a dual layer dielectric stack. IEEE International Reliability Physics Symposium (IRPS), Anaheim, CA, pp. 778–786.
57. Bersuker, G., Yum, J., Vandelli, L., Padovani, A., Larcher, L., Iglesias, V., Porti, M., Nafría, M., McKenna, K., Shluger, A., Kirsch, P., and Jammy, R. (2011) Grain boundary-driven leakage path formation in HfO₂ dielectrics. *Solid-State Electron.*, **65–66**, 146–150.
58. Shubhakar, K., Pey, K.L., Kushvaha, S.S., Bosman, M., O’Shea, S.J., Raghavan, N., Kouda, M., Kakushima, K., Wang, Z.R., Yu, H.Y., and Iwai, H. (2011) Nanoscale electrical and physical study of polycrystalline high- κ dielectrics and proposed reliability enhancement techniques. 49th IEEE International Reliability Physics Symposium (IRPS), pp. GD.1.1GD.1.6.
59. McKenna, K. and Shluger, A. (2009) The interaction of oxygen vacancies with grain boundaries in monoclinic HfO₂. *Appl. Phys. Lett.*, **95**, 222111.
60. Raghavan, N., Pey, K.L., Shubhakar, K., and Bosman, M. (2011) Modified percolation model for polycrystalline high- κ gate dielectric stack with grain boundary defects. *IEEE Electron Device Lett.*, **32** (1), 78–80.
61. Shubhakar, K., Pey, K.L., Raghavan, N., Bosman, M., Kushvaha, S.S., Wang, Z.R., and O’Shea, S.J. (2013) Study of preferential localized degradation and breakdown of HfO₂/SiO_x dielectric stacks at grain boundary sites of polycrystalline HfO₂ dielectrics. *Microelectron. Eng.*, **109** (9), 364–369.
62. Long, S., Cagli, C., Ielmini, D., Liu, M., and Suñé, J. (2012) Analysis and modeling of resistive switching statistics. *J. Appl. Phys.*, **111**, 074508.
63. Young-Fisher, K.G., Bersuker, G., Butcher, B., Padovani, A., Larcher, L., Veksler, D., and Gilmer, D.C. (2013) Leakage current-forming voltage relaxation and oxygen gettering in HfO_x RRAM devices. *IEEE Electron Device Lett.*, **34** (6), 750–752.
64. Raghavan, N., Pey, K.L., Li, X., Liu, W.H., Wu, X., Bosman, M., and Kauerauf, T. (2011) Very low reset current in RRAM achieved in the oxygen vacancy controlled regime. *IEEE Electron Device Lett.*, **32** (6), 716–718.
65. Raghavan, N., Padovani, A., Wu, X., Shubhakar, K., Bosman, M., Larcher, L., and Pey, K.L. (2013) The ‘buffering’ role of high- κ in post breakdown degradation immunity of advanced dual layer dielectric gate stacks. 51st IEEE International Reliability Physics Symposium (IRPS), Monterey, CA, pp. 5A.3.1–5A.3.8.
66. Degraeve, R., Kaczer, B., De Keersgieter, A., and Groeseneken, G. (2001) Relation between breakdown mode and location in short-channel nMOSFETs and its impact on reliability specifications. *IEEE Trans. Device Mater. Reliab.*, **1**, 163–169.
67. Li, X., Tung, C.H., and Pey, K.L. (2008) The nature of dielectric breakdown. *Appl. Phys. Lett.*, **93**, 072903.
68. Li, X., Tung, C.H., Pey, K.L., and Lo, V.L. (2009) The physical origin of random telegraph noise after dielectric breakdown. *Appl. Phys. Lett.*, **94**, 132904.
69. Tung, C.H., Pey, K.L., Tan, L.J., Radhakrishnan, M.K., Wen, L.H., Palumbo, F., and Lombardo, S. (2003) Percolation path and dielectric-breakdown-induced-epitaxy evolution during ultrathin gate dielectric breakdown transient. *Appl. Phys. Lett.*, **83**, 2223–2225.
70. Pey, K.L., Ranjan, R., Tung, C.H., Tang, L.J., Lin, W.H., and Radhakrishnan, M.K. (2004) Gate dielectric degradation mechanism associated with DBIE evolution. IEEE International Reliability Physics Symposium (IRPS), pp. 117–121.
71. Li, X., Liu, W.H., Raghavan, N., Bosman, M., and Pey, K.L. (2010) Resistive switching in NiSi gate metal-oxide-semiconductor transistors. *Appl. Phys. Lett.*, **97**, 202904.

72. Ranjan, R., Pey, K.L., Tung, C.H., Ang, D.S., Tang, L.J., Kauerauf, T., Degraeve, R., Groeseneken, G., De Gendt, S., and Bera, L.K. (2006) Failure defects observed in post-breakdown high- γ /metal gate stack MOSFET. 44th Annual IEEE International Reliability Physics Symposium (IRPS), pp. 590–594.
73. Li, X., Pey, K.L., Bosman, M., Liu, W.H., and Kauerauf, T. (2010) Direct visualization and in-depth physical study of metal filament formation in percolated high- γ dielectrics. *Appl. Phys. Lett.*, **96**, 022903.
74. Dumin, D.J., Maddux, R.J., Scott, R.S., and Subramaniam, R. (1994) A model relating wearout to breakdown in thin oxides. *IEEE Trans. Electron Devices*, **41** (9), 1570–1580.
75. Zhang, L. and Mitani, Y. (2006) Structural and electrical evolution of gate dielectric breakdown observed by conductive atomic force microscopy. *Appl. Phys. Lett.*, **88**, 032906.
76. Faivre, E., Fares, L., Putero, M., Blonkowski, S., and Muller, C. (2012) Nanostructural defects evidenced in EEPROM memories by advanced electron microscopy, L2-5. E-MRS Strasbourg, Spring.
77. Pan, L., Lin, X., Miner, B., and Johnson, K. (2010) Challenges and opportunities in characterizing modern nano-devices. *Microsc. Microanal.*, **16**, 1906–1907.
78. Kil-Soo Koo, C., Chen, P., Yue, D., Lavangkul, S., Mogul, H., Siddiqui, S., and Bonifield, T. (2005) Characterization of defect formation during Ni silicidation for CMOS device application. *Microsc. Microanal.*, **11**, 2096–2097.
79. Yamaguchi, T., Kashihara, K., Kudo, S., Okudaira, T., Tsutsumi, T., Maekawa, K., Asai, K., and Kojima, M. (2009) Anomalous nickel silicide encroachment in n-channel metal-oxide-semiconductor field-effect transistors on Si(110) substrates and its suppression by Si+ ion-implantation technique. *Jpn. J. Appl. Phys.*, **48**, 066513.
80. Wu, X., Pey, K.L., Zhang, G., Bai, P., Li, X., Liu, W.H., and Raghavan, N. (2010) Electrode material dependent breakdown and recovery in advanced high- κ gate stacks. *Appl. Phys. Lett.*, **96**, 202903.
81. Lombardo, S., Crupi, F., La Magna, A., Spinella, C., Terrasi, A., La Mantia, A., and Neri, B. (1998) Electrical and thermal transient during dielectric breakdown of thin oxides in metal-SiO₂-silicon capacitors. *J. Appl. Phys.*, **84**, 472–479.

9**Physics and Chemistry of Nanoionic Cells**

Ilia Valov and Rainer Waser

Redox-based resistive switching memories (ReRAMs) can be regarded as nano-scaled electrochemical cells, which are characterized by thermodynamics and kinetics of the constituting materials and interfaces. Correspondingly, we introduce the formation of phases by nucleation and growth, phase boundary properties, and boundary layers, the origin of electromotive forces (emfs), solid-state reactions such as oxidation of metals and phase segregation, electrochemical interface reactions, as well as kinetic demixing processes. Emphasis is placed on the effect showing up at nanometer dimensions because of their relevance to ReRAM cells.

9.1**Introduction**

In this chapter, we introduce concepts of the physics and chemistry of nanoionic systems, which are beyond the electronic structure of oxides (Chapter 3), point defects and lattice disorder in crystals (Chapter 4), transport of atoms and ions in ionic materials (Chapter 5), and electronic transport properties (Chapter 6). We cover the thermodynamics and kinetics of heterogeneous systems, that is, equilibria at phase boundaries including space-charge layers, phase transition processes including the nucleation and growth of new phases, and, in particular, solid-state reactions such as oxidation of metals. Furthermore, we provide an introduction to the thermodynamic driving forces in ReRAM cells and to the kinetics of electrochemical processes in such cells. Due to limited length of the chapter, all these topics can only be covered briefly. For gathering a more thorough background in this field, for a deeper understanding, and for many other details, the reader is referred to excellent textbooks by some authors, for example, Schmalzried [1, 2], Maier [3], Bard and Faulkner [4], Rieger [5], and Rickert [6].

Because of the size of typical redox-based memristive cells in the nanometer range, the scale needs to be taken into account in discussing the topics just mentioned. In other words, we have to deal with states and processes on the nanometer scale, which leads to interesting differences when compared to the

phenomena occurring in the macroscopic world. In addition to well-known size effects such as energy quantization, Coulomb blockade, and electronic conductance quantization, additional phenomena arise from the different *thermodynamics and kinetics of nanosized systems*. For example, the extension of space-charge regions at the phase boundary in ionic or mixed conducting systems may formally be larger than the cell dimensions so that fully depleted or fully enriched regions occur; additional driving forces may balance the large impact of the surface energy in the formation or dissolution of nanosized phases, such as nanoparticles and atomic filaments.

For all the topics, we will first provide a classical (i.e., macroworld) introduction and then mention the aspects that are encountered when shrinking the system sizes to nanoscale.

9.2

Basic Thermodynamics and Heterogeneous Equilibria

Equilibrium thermodynamics describes the conditions for phase formation, phase stability, and phase transitions as a function of the state variables, for example, the temperature T , the pressure p , and the number of molecules n (in chemistry, often expressed in moles, i.e., units of $6 \cdot 10^{23}$) of one or more components by state functions. A state function is independent of the path the system has taken to reach its particular function value and is unambiguously determined by the state variables. State functions are, for example, the internal energy W and the Gibbs free energy G . The equilibrium thermodynamics allows for the description/prediction of phase transitions such as evaporation and condensation, melting and crystallization, as well as solid–solid reactions, phase equilibria, mixing and demixing reactions, and so on. Thermodynamics has three basic laws:

- 1) The first law postulates the conservation of energy. Energy changes are generally possible by exchange of heat or work. The electrical work is given, for example, by $-\varphi Q$, where φ is the electric potential and Q is the charge.
- 2) The second law says that the entropy S of an isolated system can only increase (or stay constant) but never decrease. It is defined as $S = k_B \ln \Omega$, where Ω denotes the number of possible arrangements within a system, or more precisely, the number of microstates consistent with a given macrostate.
- 3) The third law says that, as the temperature tends to $T = 0\text{ K}$, the entropy of a perfect crystal approaches a constant value S_0 . It is not possible to cool down a system to the absolute zero.

The classical thermodynamics helps in describing systems in equilibrium and predict their properties. For this reason, it is not always easy to experimentally determine the thermodynamic equilibrium and the corresponding properties of solid–solid systems, especially at low temperatures. For constructing the phase diagrams, the most convenient state function is the Gibbs free energy, defined as:

$$G = W + pV - TS, \quad (9.1)$$

where W denotes the internal energy and pV the product of pressure and volume, that is, the volume work.

The chemical potential η_{chem} represents the change in the Gibbs free energy by the addition of one particle of the substance i under consideration. For a system with more chemical components, the chemical potential of the species i is:

$$\eta_{\text{chem},i} = \left(\frac{\partial G}{\partial n_i} \right)_{T,p,n_j (j \neq i)} . \quad (9.2)$$

The condition for equilibrium can be formulated as:

$$\Delta G = \sum_i v_i \eta_{\text{chem},i} = 0, \quad (9.3)$$

where v denotes the stoichiometric coefficient.

For example, the reaction of formation of metal oxide (MeO) is given by the chemical equation $\text{Me} + 1/2\text{O}_2 \rightleftharpoons \text{MeO}$. The condition of Eq. (9.3) is then:

$$\eta_{\text{chem,MeO}} - \eta_{\text{chem,Me}} - \frac{1}{2}\eta_{\text{chem,O}_2} = 0. \quad (9.4)$$

If any of the terms changes, the other terms try to compensate, for example, if in a system of Ni , O_2 , and NiO , one atom oxygen is added to the system, then one atom Ni will "catch" it to maintain the equilibrium. If we further add more and more oxygen molecules, the process will continue until the total amount of Ni is consumed.

One of the important rules in the classical thermodynamics is the *Gibbs phase rule*. It gives the number of phases P in one (multicomponent) system that can coexist and relates them to the number of degrees of freedom F , that is, the number of independent intensive properties (e.g., temperature and pressure).

$$F = C - P + 2, \quad (9.5)$$

where C is the number of components. The number 2 in (Eq. (9.5)) accounts for the fact that besides the concentration of the components, the temperature and pressure can also be varied. Equation (9.5) helps to determine the regions of phase stability, the equilibrium boundary lines, and the triple points, where equilibrium lines cross. Let us consider as an example the system "Ti in contact with TiO_2 ": $C=2$ – the number of components is 3, that is, Ti , TiO_2 , and O_2 but minus one chemical equation relating them; the number of the phases is $P=3$ (Ti , TiO_2 , and O_2), so $F=1$. This means that if we fix the temperature, the oxygen partial pressure (very low) at the Ti/TiO_2 boundary will be thermodynamically fixed.

In a similar manner, one can calculate the phase transitions of water, for example, at its triple point: $C=1$ (one component) and $P=3$ (three phases). Thus, the degree of freedom at the triple point is $F=0$ (the system is invariant).

Following the Gibbs phase rule, the phase diagrams can be constructed and presented as different plots, for example, pressure–temperature plots, temperature–mole fraction (x_i) plots, and Gibbs energy change versus temperature plots, depending on the information one wants to extract.

As an example, T - x diagram for the binary system $\text{Zr}-\text{O}$ is shown in Figure 9.1.

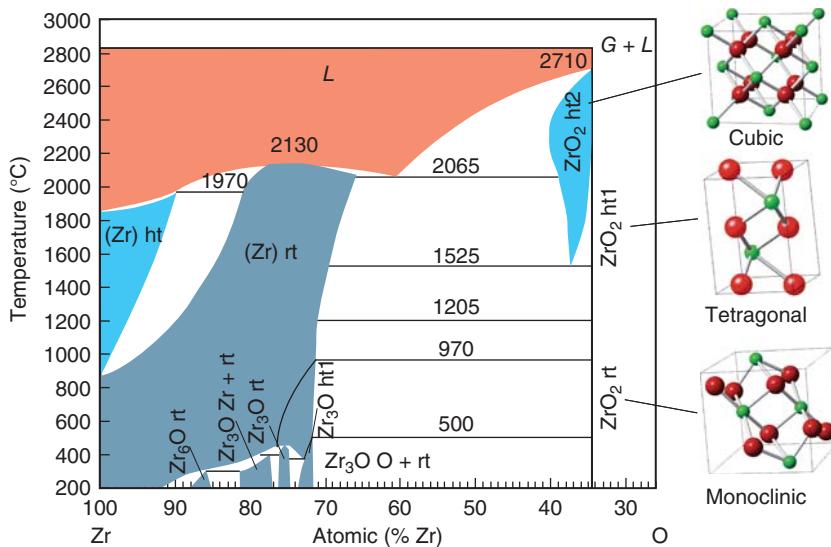


Figure 9.1 Equilibrium phase diagram of the Zr–O system as a function of the temperature. Subphases can be observed at higher metal content, corresponding to lower oxidation state of the Zr ion. rt and

ht in the diagram denote room temperature and high temperature, respectively. L denotes liquid phase. The diagram is modified from Ref. [7]. The phase diagram of Hf–O is also quite similar.

It can be seen that the phase stabilities depend on both temperature and mole fraction, provided that the other variables are kept constant. The principle construction of such a diagram is explained in the example illustrated in Figure 9.2a.

The equilibrium condition is given by:

$$\eta_{\text{chem},i}^{\alpha} = \eta_{\text{chem},i'}^{\beta} \quad (9.6)$$

where α and β are the phase indices. Mostly, convenient conditions are provided when both liquid and solid states of the two components are fully miscible. For the derivation of equations mathematically describing the two lines, it is assumed that the melting enthalpies are approximately independent of the temperature in the range from T_1 to T_2 . For more detailed information, the change in the Gibbs free energy at temperature T is also shown in Figure 9.2b as a function of the molar fraction.

Different information can be extracted from the Gibbs energy–temperature– $p\text{O}_2$ diagrams (Ellingham diagrams, see Chapter 3) from which the stability of typically oxide phases as a function of the temperature and equilibrium oxygen partial pressure can be determined. For more detailed discussion on the Ellingham diagrams, we refer the reader to Chapter 3.

The construction of the phase diagrams is usually much more complicated than in the model condition shown in Figure 9.2. There are, for example, binary Me–O systems (Me is a transition metal oxide) where the metal component can take different oxidation states and several discrete oxide phases exist. Examples are

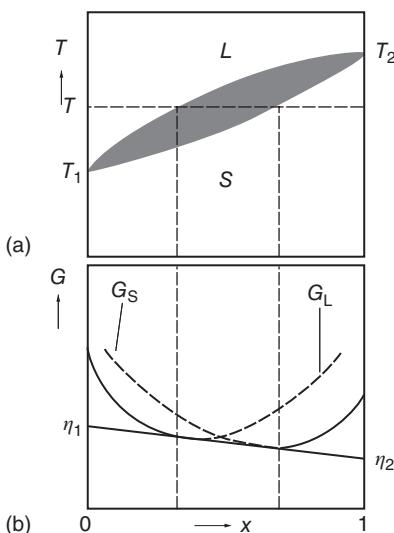


Figure 9.2 (a) T - x phase diagram for a simple binary system assuming ideal solubility of the phases. L and S denote the liquid and solid phase, respectively. (b) The corresponding change in the Gibbs energy G . The figure is adapted from Ref. [2].

$\text{Ti}-\text{O}$, $\text{W}-\text{O}$, $\text{V}-\text{O}$, and $\text{Mo}-\text{O}$. They form homologous series of stoichiometric oxides and along with the stoichiometries such as MeO or MeO_2 or Me_2O_3 , some nontrivial compounds were reported, for example, Mo_9O_{26} . In systems such as $\text{MeO}_2-\text{MeO}_3$ and/or $\text{Me}_2\text{O}_3-\text{MeO}_2$, a series of oxides of compositions, for example, $\text{Me}_x\text{O}_{3x-1}$ or $\text{Me}_x\text{O}_{2x-1}$ can be observed. As possible structure of these phases is suggested a periodically repeated dislocated structural unit which can appear in 1D, 2D, or 3D space. Not all of the series compositions must be formed, and most often, the Magnéli phases were reported.

The most prominent example is the $\text{Ti}-\text{O}$ system. TiO_2 shows several types of structural and chemical defects. The point defects include oxygen vacancies and Ti interstitials. Apart from this, it also shows particular accumulation of point defects in random shear planes (Wadsley defects). At a higher densities, these Wadsley defects (see Figure 9.3) order themselves into Magnéli phases [9].

In Figure 9.4, the thermodynamic data for high temperature (over 1300 K) equilibrium are shown.

Several important details can be extracted from these diagrams. (i) The solubility of oxygen into the metallic phase is substantial, and an additional phase can be observed for temperatures above ~ 1200 K. (ii) The $\text{TiO}_{1\pm x}$ phase (NaCl structure) forms structures with different numbers of cation and anion vacancies, different orders and defects, demonstrating a large range of homogeneities. (iii) In the range between Ti_2O_3 and TiO_2 (range of higher oxygen chemical potentials), a series of Magnéli phases were observed with a general formula $\text{Ti}_x\text{O}_{2x-1}$. As seen from Figure 9.4b, the difference in the Gibbs free energy of formation of these phases is quite low, that is, about 9 meV (0.8 kJ mol^{-1}). Therefore, it is not surprising that it is difficult to define and set precisely equilibrium conditions for the different Magnéli phases.

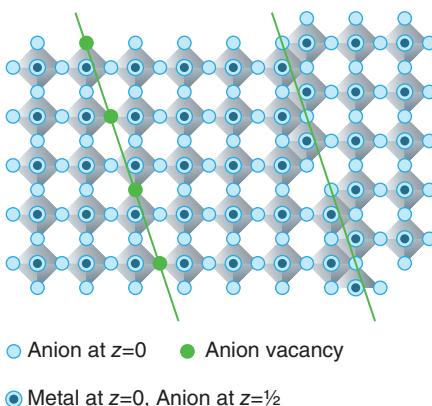


Figure 9.3 Ordered arrays of oxygen vacancies (left) are eliminated by the formation of shear planes (right), such as Wadsley defects, in which the cations move into interstitial positions [8].

9.3

Phase Boundaries and Boundary Layers

9.3.1

Driving Force for the Formation of Space-Charge Layers

Establishment of a contact between phases proceeds with exchange and/or rearrangement of matter and/or charge(s) at the interface and in the adjacent phase boundary region. The general condition for equilibrium is a more general form of Eq. (9.3):

$$\sum_i v_i \eta_i = 0, \quad (9.7)$$

with η_i denoting the *electrochemical potential* of species i , and v_i is the stoichiometric coefficient. The electrochemical potential is defined as:

$$\eta_i = \eta_{\text{chem},i} + z_i e \varphi, \quad (9.8)$$

where z_i denotes the number of charges of species i , e is the electron charge, and φ is the electrostatic potential, also called *Galvani potential*. In the case of neutral species ($z=0$), the chemical and electrochemical potentials are, of course, equivalent, $\eta_i = \eta_{\text{chem},i}$.

Equation (9.7) says that in the case of an electrode in contact with an electrolyte, the different energetics of both phases result in an electrical potential difference. Here, we define an electrolyte as any solid, gas or liquid phase that is able to transport ions. In resistive switching memories, a variety of oxides, higher chalcogenides, and sometimes halides are used as electrolytes, for example, in electrochemical metallization (ECM) cells [10–12], valence change memory (VCM) cells [13], and thermochemical memory (TCM) cells [14]. They are termed *solid electrolytes* even when they belong to insulators in a macroscopic sense, as long as they

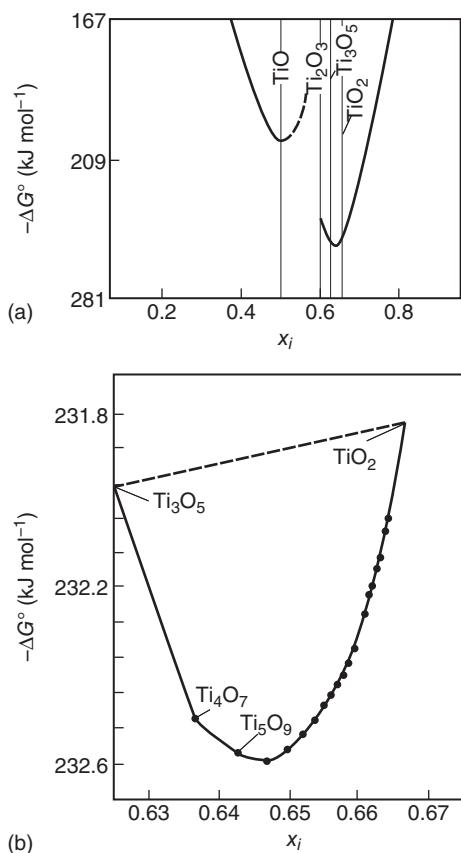


Figure 9.4 Equilibrium phase diagram for TiO_x system. The formation of a number of Magnéti phases separated by small Gibbs formation energies can be observed at

lower oxygen potentials. a) in the entire range of Ti/O ratio and b) in a selected region zoomed around $x_i = 0.63$. The figure is adapted from Ref. [2].

conduct ions in the nanometer size cells, supported by local fields and temperatures, as in the resistive switching processes. Moreover, many of those materials (mainly oxides) transit from one type to another as shown in Figure 9.5, depending on the particular experimental conditions and the electrode material.

The most typical example of this phenomenon is TiO_2 : It was reported to conduct cations in ECM by using, for example, Cu as active electrode [15, 16], or anions in VCM material [17–19], and TCM material [14, 19]. Despite the insulating properties at macroscopic level, this and many other oxides, for example, Ta_2O_5 , Al_2O_3 , SiO_2 , ZrO_2 , and HfO_2 are reported to show sufficient ionic mobility to be considered as “electrolytes at nano-dimensions.” Cyclic voltammetry and other typical electrochemical techniques have been used in studying redox processes on materials such as SiO_2 , and Ta_2O_5 used in ECM cells [20–22].

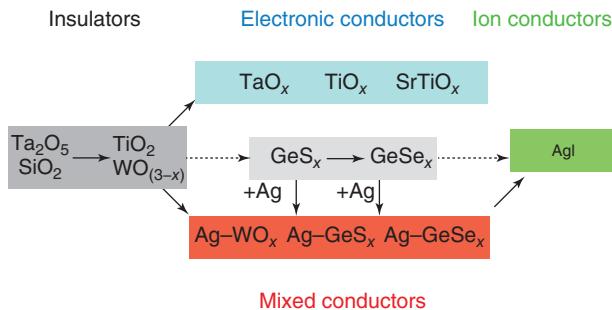


Figure 9.5 Sketch of different material classes used for ReRAM and the transition of their transport properties, depending on the thermodynamic conditions and operation (pre)history.

The combination of Eqs. (9.7) and (9.8) for one species in two different phases leads to the expression:

$$(\varphi^\alpha - \varphi^\beta) = \frac{\eta_{\text{chem},\beta} - \eta_{\text{chem},\alpha}}{ze}. \quad (9.9)$$

For achieving equilibrium at the interface for the case where exchange of matter is not possible, the difference in the chemical potential of the species is compensated by the electrical potential. Thus, a *space-charge layer* is formed. In the field of electrochemistry, this space-charge layer is called an electrical double layer (EDL).

9.3.2

Enrichment and Weak Depletion Layers

The concentration of charged species i in the vicinity of the electrode/electrolyte interface as a function of the local electric potential φ is expressed by the Boltzmann distribution function:

$$\frac{N_i}{N_i^\circ} = \exp\left(-\frac{z_i e \varphi}{k_B T}\right) \quad (9.10)$$

where N_i° is the number of charges (per unit volume) in the bulk and N_i is this number at potential φ . The space-charge density is given by the sum over all charged species:

$$\rho = \sum z_i e N_i^\circ \exp\left(-\frac{z_i e \varphi}{k_B T}\right) \quad (9.11)$$

In combination with the Poisson equation $\epsilon_0 \epsilon \nabla^2 \varphi = \rho$, this leads to the Poisson–Boltzmann equation:

$$\nabla^2 \varphi = \frac{e}{\epsilon_0 \epsilon} \sum_i z_i N_i^\circ \exp\left(-\frac{z_i e \varphi}{k_B T}\right) \quad (9.12)$$

where ϵ_0 and ϵ denote the permittivity of free space and the dielectric constant of the medium, respectively. This equation is used to describe enrichment and weak

depletion space-charge layers. The concentration profiles that can be described by Eq. (9.12) are called *Gouy–Chapman distribution*.

For conditions fulfilling $z_i e \varphi \ll k_B T/e$ ($k_B T/e = 25$ mV at room temperature) and by defining the term *ionic strength* $I = 1/2 \sum_i z_i^2 N_i^\circ$, Eq. (9.12) can be rewritten as:

$$\nabla^2 \varphi = \frac{2e^2 I \varphi}{\epsilon_0 \epsilon k_B T} \quad (9.13)$$

By combining all constants from Eq. (9.13), we obtain an important parameter, that is, the Debye (charge screening) length:

$$\lambda_D = \sqrt{\frac{\epsilon_0 \epsilon k_B T}{2e^2 I}} \quad (9.14)$$

Equations (9.7)–(9.14) apply for solid–liquid as well as for solid–solid interfaces in the case of *enrichment* space-charge layers and weak-to-moderate *depletion* space charge layers. The charge screening length varies for different materials and conditions taking values in the range between few angstroms and few nanometers. The higher the concentration of mobile charged species, the shorter the Debye length. For example, AgBr and AgCl are Ag^+ -ion solid electrolytes at 573 K with cation-type Frenkel disorder. Because the concentration of interstitials ions and vacancies in AgBr is almost 10 times higher than in AgCl, the charge screening length for AgBr is also much smaller, that is, ~ 0.6 nm, whereas under the same conditions for AgCl, the Debye length is ~ 2 nm.

We need to define the application range of Gouy–Chapman model (Figure 9.6). The charge distribution according to Eq. (9.12) is applicable when either two (oppositely charged) mobile species are present in a phase at the interface or if only one charged species is mobile and its depletion at the interface is only weak. There are two cases in which the model fails:

If there is a strong enrichment of ions with a Debye length smaller than the radius of the (solvated) ions, typically in concentrated electrolytes, the space-charge layer shrinks to a layer of ions adsorbed at the interface. This is called the *Helmholtz double layer*. In electrochemistry, the combination between a Helmholtz double layer and a Gouy–Chapman space-charge layer is called a *Stern layer*.

If the electrolyte contains only one mobile charge carrier while the compensating ions are immobile, then a strong depletion of the mobile carriers at the interface may occur. This layer is called the *Mott–Schottky*-type space-charge layer and is described in the next section.

9.3.3

Strong Depletion Layers

When only one type of mobile charge carrier is present and it is depleted so strongly that its concentration at the interface approaches zero, further depletion according to Eq. (9.14) is not possible. Instead, the layer of full depletion expands

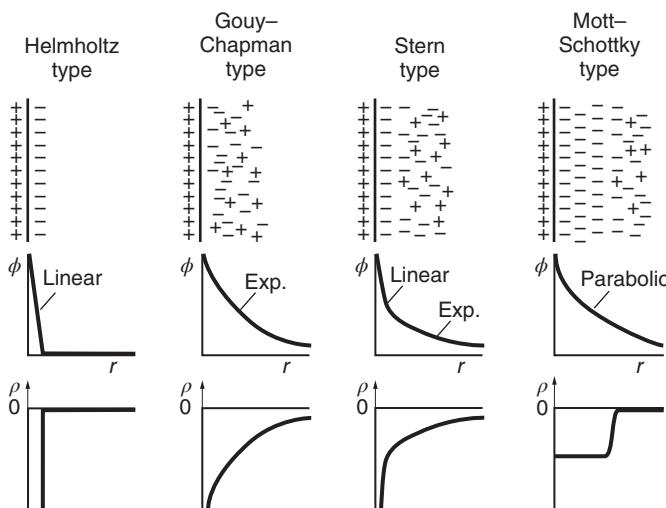


Figure 9.6 Different models for the formation of the space-charge layer (electrical double layer) described in the main text. The corresponding charge and electrical potential distributions are also presented. A positively charged metal is assumed on the left side.

in its thickness. This thickness is then not anymore influenced by the permittivity ϵ_r as in Eq. (9.14). Instead, the charge screening length is then given by:

$$\lambda_{M-S} = \frac{Q^*}{N_i} \quad (9.15)$$

where Q^* is the charge per unit area to be compensated by the space-charge layer. The charge screening length may be up to some tens or even hundreds of nanometers in the case of low N_i .

This Mott–Schottky-type space-charge layer forms in doped semiconductors, for example, at p-n junctions where the electronic carriers, electrons on the donor-doped side and holes on the acceptor-doped side, become strongly depleted. Another example is slightly acceptor-doped SrTiO_3 , which shows a (low) mixed electronic (holes) and ionic (oxygen vacancy) conduction. Grain boundaries exhibit positively charged, immobile interface states (presumably structurally fixed oxygen vacancies), which give rise to a Mott–Schottky-type space-charge layer in which holes and oxygen vacancies are (strongly) depleted [23, 24]. The situation may even lead to the formation of an inversion layer, that is, an electron space-charge layer, as it is known in the case of MOSFETs. In SrTiO_3 , this can give rise to a certain regime of enhanced n-conduction parallel to grain boundaries or dislocations – a situation that can be used to explain some effects encountered in VCM systems.

More detailed discussion on the space-charge layer formation is made in Chapter 4.

The structure (including defect chemical structure for the solid electrolytes) of the near-interface regions at the electrode/electrolyte interface is of highest

importance to understand the electrode kinetics and the electrocatalytic activity. The theory of the space-charge layers in the case of electrolyte–semiconductor and metal–semiconductor interface is extensively discussed in Ref. [25], and for ionic solids, this theory has been intensively developed by Maier [26]. However, direct experimental studies on the structure of the space-charge layers on the nanoscale are limited by the required high sensitivity of the equipment. The development of new spectroscopic and microscopic techniques has been advanced over the years, improving the prospects of detailed studies of space-charge effects in resistive switching cells (e.g., Ref. [27]).

It should be noted that, in addition to electrically charged interface states acting on ions, chemical energy differences between the bulk and the interface (e.g., surface energy, adsorption energy) and/or mechanical energy differences (e.g., interface stress) also lead to composition changes in the bulk when the interface is approached. This is called *segregation*, and it is frequently encountered in solids (and liquids) consisting of more than one component. The component providing conditions for minimum surface energy will segregate. When the segregation exceeds the solubility limit, *phase separation* occurs.

9.3.4

Nanosize Effects on Space-Charge Regions

Because of the nanometer dimensions of ReRAM cells, we face the fact that the space-charge layers of both electrodes may overlap. As the electrolyte thickness typically varies between few nanometers and 100 nm, ReRAM cells are expected to show significant deviations in their thermodynamic properties related to the overlapping of the space-charge layers at the metal electrode–solid electrolyte interfaces. The effective thickness below which these size effects occur, mainly depends on the (ionic and electronic) charge carrier concentrations and mobilities. It should be noted that the aspect of overlapping space-charge layers is crucial for solid films in ReRAM cells, which are conventionally considered as insulators such as SiO_2 , Ta_2O_5 , HfO_2 , and so on, due to their nominally low carrier concentrations.

As discussed, after a contact has been established at the interface, the accumulated excess surface charge in one of the phases is compensated by a counter charge in the other phase. A schematic representation of the classical macroscopic picture, the overlapping space-charge layers and the corresponding distribution of the (electro)chemical potentials is shown in Figure 9.7.

When the electrolyte thickness is smaller than the screening length, the charge neutrality condition (as defined for the bulk) may not necessarily be obeyed. Depending on the properties of the contacting phases, different situations of the charge distribution can be expected. Each particular situation should be considered individually by taking into account the number and oxidation states of the charges, and whether Gouy–Chapman, Mott–Schottky, or even mixed space-charge layers are favored. A good review of possible charge (re)distributions can be found in Ref. [28]. Overlapping of space-charge layers indeed leads to a

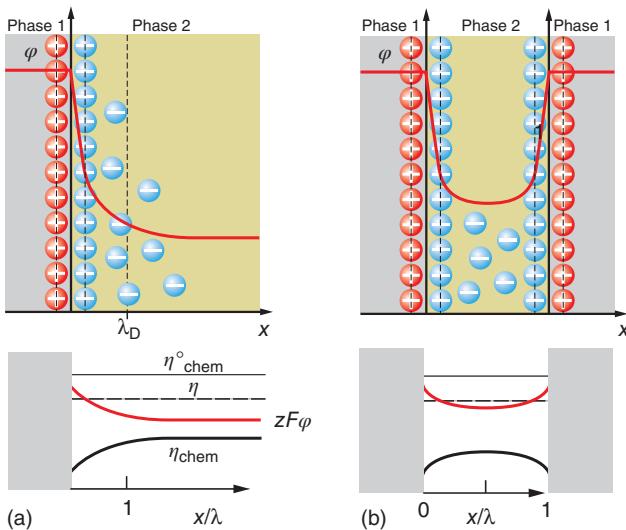


Figure 9.7 Electrostatic potential φ across an electric double layer (space charge layer) (a) at the electrode–electrolyte interface according to the Stern model and (b) for

the situation in the nanoworld where the electrolyte thickness is comparable to twice the Debye length ($d < 2\lambda_D$). The figure is adapted from [11] (top) and [3] (bottom).

significant change in the transport properties as demonstrated for $\text{CaF}_2/\text{BaF}_2$ system [29] and discussed for depletion layers in SrTiO_3 [30].

Additionally, in crystallites smaller than approximately 50 nm, a de-doping effect may be observed. In this case, the concentration of mobile ions and/or defects can be drastically reduced due to segregation processes, leaving the grain without mobile defects [31–33]. The difference in the properties may be so great that grain cores are considered to be in a self-consistent phase (abrupt spatial change of properties). Thus, the increase in the volume fraction of the grain cores for nanoparticles can result in a change in the transport properties of the system as a whole. These effects can be significant in a variety of polycrystalline oxides and binary superionic materials and also in phase-separated ternary electrolytes [34]. The boundary region may be depleted, respectively enriched, and the transport properties deviate. It has been demonstrated that the conductivity of the grain boundaries can decrease with decreasing the grain size [35], and it has also been shown that the conductivities of the grain boundaries in perpendicular and parallel directions are not necessarily equal [26]. If the grain size in some systems (e.g., yttria-doped zirconia, $\text{Zr}_{1-x}\text{Y}_x\text{O}_{2-x/2}$; short: YSZ) decreases below a certain dimension (approximately 50 nm in diameter), the conductivity σ_{gb}^{\parallel} parallel to the grain boundaries becomes negligible, and the transport proceeds mainly perpendicular to the grain boundary σ_{gb}^{\perp} [26]. For nanoscaled materials, both parallel and perpendicular paths in the grain boundaries are important, thus increasing the contribution of the grain boundaries to the total conductivity. A practical example for the space-charge layer effect on the properties of materials

is the *heterogeneous doping*. It has been demonstrated that the addition of insulating particles of, for example, SiO_2 or Al_2O_3 , or other electrolytes into solid ion conductors (LiI , AgBr , CaF_2 , etc.) can increase the ionic conductivity by orders of magnitude [36, 37]. The reason is the formation of Gouy–Chapman-type space-charge layers at the insulator/electrolyte interface, leading to an enrichment of particular charged species (vacancies or ions), thus ensuring a fast pathway. The influence of the insulating phase and the orientation has also been elaborated. It has been shown that on reaching a certain volume fraction of the insulating component in the solid electrolyte, a conductivity enhancement effect occurs as a percolation mechanism (the finely dispersed insulating particles are located mainly at the grain boundaries of the larger ionic grains) and further enrichment with the insulating component does not lead to further significant increase in the conductivity. A detailed overview on the effect of heterogeneous doping is provided in Ref. [26].

9.3.5

Nanosize Effects due to Surface Curvature

Apart from overlapping of space-charge layers, the properties of matter at the nanoscale can differ significantly from bulk properties due to effects related to excess surface free energy (higher chemical potential) and charge (re)distribution. This effect accounts for the increase in the electrochemical potential $\tilde{\eta}_i$ (or the chemical potential $\eta_{\text{chem},i}$ for neutral species) of the components due to enhanced surface energy contribution, given by the Gibbs–Thomson equation:

$$\left(\frac{\partial \Delta \eta_{\text{chem},i}}{\partial p} \right)_{T,n} = -k_B T \left(\frac{\partial \ln a_i}{\partial p} \right)_{T,n} = \Delta V \quad \text{or} \quad \Delta \eta_{\text{chem},i} = \Delta V \cdot \Delta p = \Delta V \cdot \frac{2\bar{\gamma}}{\bar{r}}, \quad (9.16)$$

with ΔV denoting the molar volume, $\bar{\gamma}$ is the average surface tension, p is the pressure, T the temperature, a_i is the activity of the i th component, n is the number of moles, and \bar{r} is the average radius of the nanoparticles. Here, we simplify the conditions by assuming that the surface tension remains constant with decreasing particle size.

9.3.6

Formation of New Phases at Phase Boundaries

The boundary layers often undergo phase transitions, as a result of segregation effects and subsequent exceeding of the solubility limit. For example, in the systems $\text{Ag}-\text{GeS}_x$ and $\text{Zn}-\text{GeS}_x$, the formation of a boundary compound (ternary metal– GeS_x) was reported [38, 39]. In some particular cases, even the deposition method plays a role to determine the boundary layer structure and composition. Thus, Ag deposited on GeS_x leads to a preferable formation of an $\text{Ag}-\text{Ge}-\text{S}$ surface layer, whereas deposition of GeS_x on Ag results in an $\text{Ag}-\text{S}$ boundary layer [38].

In other cases, the nanosize of the individual layers leads to deviation of their chemical activity and reactions can proceed even if they are not thermodynamically favored for macroscopic systems. For example, at the Cu/SiO₂ interface, an intermediate boundary layer of ~1 nm thick Cu_{1+x}O is formed even in vacuum conditions [40], despite the fact that this process is not expected for macroscopic systems.

9.4

Nucleation and Growth

Thus far, we have discussed the thermodynamics, that is, the energy aspects of static situations. We now include processes as well, that is, the kinetics of the change of states. We start with the nucleation and growth of new phases, which are essential for the operation of ReRAM cells. In ECM-type cells, metal phases are built and dissolved in every SET and RESET cycle. In VCM and TCM cells, new phases are typically generated in the initial electroforming cycle.

9.4.1

Macroscopic View

The process of formation of a new phase on a foreign substrate is related to overcoming of a certain thermodynamic barrier $\Delta G(N_c)$, for example, the nucleation process for forming a critical nucleus consisting of N_c atoms, and under certain conditions, becomes the rate limiting process [41]. The nucleation theory (for both classical and atomistic point of view) is based on the Becker–Döring approach [42] treating the nucleation as a sequence of bimolecular chemical reactions where each cluster of size N can be transformed either to a higher or to a lower class by adding or losing an atom, respectively, where at equilibrium, the rates of attachments and detachments are equal:



The change in the free enthalpy of formation for a cluster consisting of N atoms can be presented as:

$$\Delta G(N) = Ne\Delta\varphi + \Phi(N) + N\varepsilon \quad (9.18)$$

where $\Delta\varphi < 0$ is the cathodic electrical potential, $\Phi(N)$ is a specific surface energy term accounting for the formation/disappearance of surfaces, and ε is the strain energy per atom. At constant volume, the form of a nucleus of N atoms can change until the contribution part of $\Phi(N)$ reaches a minimum corresponding to the equilibrium form of the crystal (Gibbs–Curie condition).

For sufficiently large clusters, the number of atoms in the nucleus $\Phi(N)$ and $\Delta G(N)$ can be treated as continuous variables. In this case, the term $\Phi(N)$ is expressed by means of specific free surface energy. At a particular voltage $\Delta\varphi_c$, the

derivative $\partial\Delta G(N)/\partial N$ becomes 0 corresponding to the equilibrium condition ($N = N_c$), where the relation between the size of the critical nucleus N_c and the electrochemically induced supersaturation is given by:

$$\Delta\eta = \left[\frac{d\Phi(N)}{dN} \right]_{N=N_c}, \quad (9.19)$$

with $\Delta\eta$ representing the electrochemical supersaturation. Equation (9.19) represents a general form of the Gibbs–Thomson equation. Upon fulfilling the equilibrium condition, the nucleus consisting of N_c atoms can disarrange (the nucleus dissolves) or further grow with equal statistical probability.

A schematic representation of the forces influencing the process of nucleation in accordance to Eq. (9.18) (excluding, for the general case, the strain) is shown in Figure 9.8.

For a system where the number of atoms constituting the critical nucleus is sufficiently large (e.g., $N_c > 20$), the Gibbs energy for formation of the critical nucleus and the nucleation rate are given by the models for 2D and 3D nucleation. The criterion to distinguish these two different modes is the dependence of the nucleation rate (J) on the applied potential being proportional to $1/|\Delta\varphi|^2$ for 3D nucleation and to $1/|\Delta\varphi|$ for 2D nucleation. Detailed discussion on this topic can be found in Ref. [43].

9.4.2

Atomistic Theory

For clusters consisting of small number of atoms, the macroscopic quantities, for example, specific surface energy, lose their physical meaning. $\Delta G(N)$ and $\Phi(N)$ can take only discrete values, and the macroscopic theory cannot be applied anymore. In the atomistic model [41, 44], the surface energy term associated with the formation of a configuration of N atoms ($N < 20$) is estimated using the bond energies between the atoms and the substrate ψ_{a-s} and the atoms themselves ψ_{a-a} (the energy term accounting for a substrate-induced strain will be omitted for the sake of simplicity). These energies are considered to be restricted to the first nearest neighbors and to be equal to the bond energies in a vacuum.

$$\Phi(N) = N\psi_{\text{kink}} - \sum_N \psi_x, \quad (9.20)$$

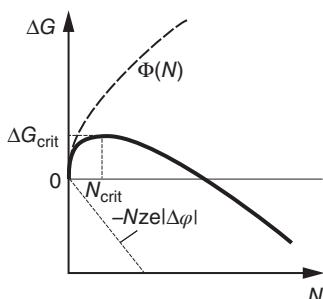


Figure 9.8 Schematic presentation of the Gibbs free energy for nucleation of a hemispherical critical nucleus N_c at an electrode surface. The influence of the surface free energy Φ and the applied cathodic potential $\Delta\varphi$ is also illustrated. The figure is adapted from Ref. [43].

where ψ_{kink} is the bond energy of an atom in a kink (half crystal) position and $\sum_N \psi_x$ is the sum of the bond energies of atoms in all other positions in the cluster. The calculation of the surface energy contribution as a function of the number of atoms is schematically represented in Figure 9.9.

Then, Eq. (9.18) becomes the form:

$$\Delta G(N) = Nze\Delta\varphi + N\psi_{\text{kink}} - \sum_N \psi_x, \quad (9.21)$$

where $\Delta G(N)$ calculated on the basis of this equation can take only discrete values as shown in Figure 9.10.

By variation of the applied potential, one can easily shift the equilibrium toward the conditions under which the critical nucleus contains one or only few atoms and is stable in a wide range of applied potentials.

The nucleation rate $J(s)$ is given in accordance to the atomistic model for electrocrystallization, for applied cathodic potentials ($\Delta\varphi < 0$) higher than $k_B T/e$, by:

$$J = N_{\text{Ag}^+} Z_0 \Gamma \underbrace{\frac{k_B T}{h} \exp\left(-\frac{\Delta G^\ddagger + \Phi(N_c)}{k_B T}\right)}_{J_0} \exp\left(-\frac{(N_c + \alpha) e \Delta\varphi}{k_B T}\right), \quad (9.22)$$

where N_{Ag^+} and Z_0 are the numbers of Ag^+ ions and active sites, respectively, which in the case of STM (scanning tunnelling microscope) can take a value of unity; Γ is the Zeldovich factor, which tends to unity for small clusters and high cathodic $\Delta\varphi$; \aleph is the transmission probability of an electron; h is the Planck's constant; ΔG^\ddagger is the free activation enthalpy of the charge-transfer reaction (at $\Delta\varphi = 0$); and α is the cathodic transfer coefficient.

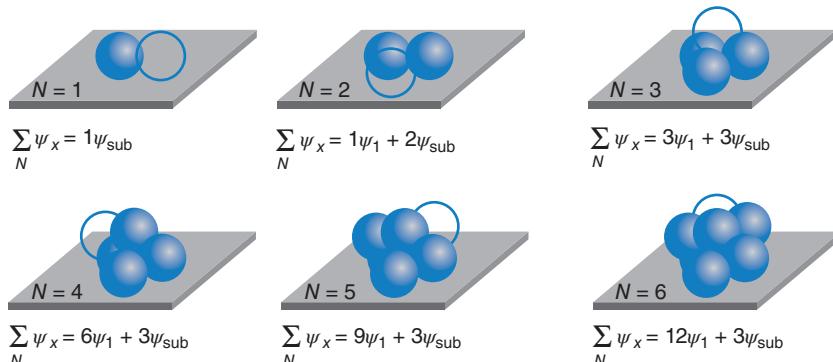


Figure 9.9 Atomistic theory of nucleation. Binding energy of atomic assemblies based on atomic bond energies in a vacuum. The bond energies between

neighboring Me atoms and the substrate are denoted as Ψ_1 and Ψ_{sub} , respectively. The figure is adapted from Ref. [44].

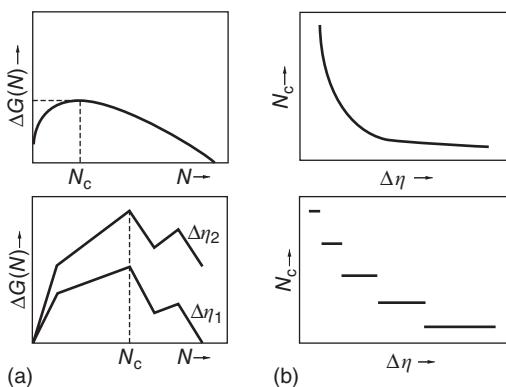


Figure 9.10 Comparison between the classical and atomistic treatment of the nucleation process. The free reaction enthalpy change (a) and the change in the elec-

trochemical potential (b) as a function of the number of constituent atoms N are schematically represented. The figure is modified from Ref. [45].

The nucleation rate–voltage dependence given in logarithmic coordinates $\ln J$ versus $\Delta\varphi$ has a similar form as the charge-transfer equation and is used to calculate the parameter N_c . In the lower limit boundary condition, that is, $N_c = 0$, the slope of this dependence transforms to the classical logarithmic slope derived from the Butler–Volmer equation.

It is important to note that, in the general case, the nucleation is a statistical process. Only under very specific conditions, for example, STM experiments, where the electron beam is focused, it can become invariant [46].

9.5

Electromotive Force

In this section, we discuss the electrical potential differences that are generated at the interfaces involving redox reactions.

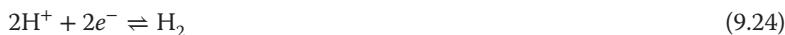
9.5.1

Electrochemical Cells of Different Half Cells

The electromotive force (emf) is a measure of the differences in the chemical potentials of two half-cell reactions of an electrochemical cell. One of the most convenient examples is the reaction of water with the general formula:



Equation (9.24) represents the total cell reaction, which includes two half-cell reactions:





In the case of electrolysis of water, reaction (9.24) will proceed to the right and reaction (9.25) to the left. In the case of a solid oxide fuel cell, the directions for both reactions are the opposite. If the two half-cell reactions are separated by, for example, a purely oxygen-ion-conducting solid electrolyte and a direct chemical reaction is avoided, the equilibrium (i.e., open-circuit conditions) can be described in accordance to Eq. (9.7):

$$\eta_{\text{chem},\text{H}_2} - 2\eta_{\text{H}^+} - 2\eta_{e^-} = \eta_{\text{O}^{2-}} - 2\eta_{e^-} - \frac{1}{2}\eta_{\text{chem},\text{O}_2} \quad (9.26)$$

The electrodes need to be chemically inert in this case (e.g., Pt), that is, chemical interaction between the electrodes and the electrolytes is not assumed.

By applying Eq. (9.8) and taking into account that

$$\eta_{\text{chem},i} = \eta_{\text{chem},i}^\circ + k_B T \ln a_i \quad (9.27)$$

($\eta_{\text{chem},i}^\circ$ denotes the standard chemical potential and a_i the activity), one obtains the expression:

$$\Delta\varphi = \varphi^\circ + \frac{k_B T}{2e} \ln \frac{a_{\text{H}^+}^2 \cdot a_{\text{O}^{2-}}}{a_{\text{H}_2} \cdot a_{\text{O}_2}^{1/2}}, \quad (9.28)$$

with φ° being the standard electrode potential ($\varphi^\circ = \sum \eta^\circ / 2e$). From Eq. (9.28), it is clear that changing the activities of the component(s) results in a change in the emf.

The electric work is given by $-ze\varphi$. Thus, the relation between the Gibbs free energy and the emf is:

$$\Delta G_{p,T} = -ze\varphi, \quad (9.29)$$

where z denotes the number of electrons exchanged during the reaction, for example, $z=2$ as in Eqs. (9.24) and (9.25).

The emf under the open-circuit conditions is a thermodynamic value. By using Eq. (9.29), one can easily calculate the emf of any electrochemical cell, having known the Gibbs formation energy of the product. For example, the Gibbs free energy of formation of water is $\Delta_f G_{\text{H}_2\text{O}} = -237 \text{ kJ mol}^{-1}$, which according to Eq. (9.29) corresponds to $\text{emf} = 1.23 \text{ V}$. Experimentally (in electrochemical cell), this value has not been measured because of the kinetic irreversibility of the oxygen half-cell reaction (Eq. (9.25)) [47].

9.5.2

Emf Caused by Surface Curvature Effects

Other examples include negative change in the Gibbs energy by transition between nano- and macrostructured phases. Thus, if we have nanosized grains of one phase at one of the electrodes and microsized grains of the same phase at the other electrode, due to the *Gibbs–Thomson equation* (see Eq. (9.16)), the

nanosized phase (higher chemical potential) will tend to dissolve and deposit at the side of the microsized phase (lower chemical potential). The corresponding emf is given by:

$$\Delta\varphi = -\frac{\eta_{\text{chem},i}^{\text{micro}} - \eta_{\text{chem},i}^{\text{nano}}}{ze}. \quad (9.30)$$

The derivation of the Nernst equation based on Eq. (9.7) (or Eqs. (9.29) and (9.30)) is generally applicable for any two-electrode electrochemical cell, irrespective of whether the electrolyte is in liquid or solid state. However, when working with solid electrolytes, we acknowledge some specific aspects that have to be taken into account. One important difference is that, whereas liquid electrolytes usually conduct more kinds of ions (both cations and anions), solid electrolytes typically conduct only one kind of ion. Looking into the literature on solid conductors, one can easily find solid electrolytes conducting either cations or anions, and the mobility of the counter ions is negligible. Typical cation-conducting solid electrolytes are, for example, RbAg_4I_5 , AgI , CuI , Ag_{2+x}S , and Cu_{2+x}S , and anion conductors are, for example, yttria-doped zirconia ($\text{Zr}_{1-x}\text{Y}_x\text{O}_{2-x/2}$; short: YSZ) and gadolinia-doped ceria (GDC). The only exception are the proton-conducting materials, for example, BZY (yttria-doped barium zirconate) where both protons and oxygen ions are mobile under certain conditions (however, these are not of the same order).

9.5.3

Emf Caused by Concentration Differences

In the solid electrolyte systems, symmetric cells of the type $(\text{O}_2)\text{Pt}/\text{YSZ}/\text{Pt}(\text{O}_2)$ find wide application as, for example, sensors or oxygen pumps. Because the same half-cell reaction (of course, in a different direction) takes place, Eq. (9.28) (Nernst emf) simplifies to:

$$\Delta\varphi = \frac{k_B T}{4e} \ln \frac{a'_{\text{O}_2}}{a''_{\text{O}_2}} \quad (9.31)$$

This means that the measured voltage is a function of only the ratio of the oxygen activities (in good approximation: oxygen partial pressures) at both inert electrodes. The essential requirements for Eq. (9.31) to be valid are (i) a constant spatial composition of the solid electrolyte, that is, the concentration of oxygen ions should be constant throughout the entire electrolyte thickness, and (ii) the conductivity should be purely ionic.

Another important difference between liquid and solid electrolytes is that, in liquids, typically no free electrons exist (with exceptions such as liquid ammonia), whereas solids also conduct electrons (or holes). This applies even for materials such as superionic conductors. Therefore, strictly speaking, each solid ion electrolyte possesses some electronic conductivity. Indeed, the difference in the ionic and electronic partial conductivities may be orders of magnitude, but under certain conditions (e.g., conditions of low oxygen partial pressure), they

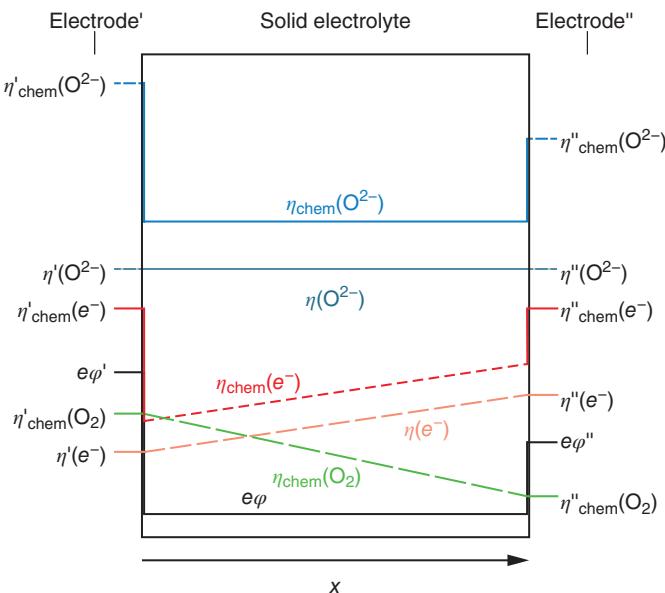


Figure 9.11 Schematic representation of the distribution of the potentials in a symmetric solid-state cell with predominant oxygen ion conductivity. The electrochemical potentials

of the individual components are split into chemical and electrical parts in accordance to Eq. (9.16).

may become comparable. A schematic representation of the distribution of the (electro)chemical potentials in a symmetric solid-state electrochemical cell is shown in Figure 9.11.

As it can be seen in the YSZ solid electrolyte, the chemical potential of oxygen and the chemical and electrochemical potentials of the electrons are not constant within the bulk phase, compensating each other to fulfill the general condition of Eq. (9.7) applicable for open-circuit conditions. For the case that the electronic partial conductivity is not negligible compared to the oxygen ion conductivity, Eq. (9.31) transforms into:

$$\Delta\varphi = \bar{t}_{O^{2-}} \frac{k_B T}{4e} \ln \frac{a'_{O_2}}{a''_{O_2}} \quad (9.32)$$

where $\bar{t}_{O^{2-}}$ denotes the average transference number of the oxygen ions, that is, the part of the conductivity due to oxygen ions. The transference number can also be defined as:

$$t_{O^{2-}} = \frac{\sigma_{O^{2-}}}{\sum_i \sigma_i} = \frac{\mu_{O^{2-}}}{\sum_i \mu_i} \quad (9.33)$$

with σ_i and μ_i representing the conductivity and the mobility of a species i , respectively. As a general condition, the sum of all transference numbers of all mobile

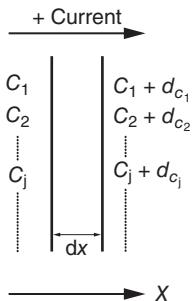


Figure 9.12 Formation of the diffusion potential. The figure is adapted from Ref. [48].

charged species must be 1:

$$\sum_i t_i = 1 \quad (9.34)$$

9.5.4

Diffusion Potentials

The transference number plays an important role in the definition of another possible source of emf, namely the diffusion potential. In contrast to the Nernst potential, the diffusion potential φ_d is a nonequilibrium potential. It appears at the interface between two contacting electrolytes each with different concentrations of mobile charge species or in the same electrolyte with a concentration gradient where one of the charges has much higher (or lower) mobility than the other(s). In such a case, the faster species move ahead and generate an electric field, which slows down their motion and accelerates the motion of the charges with lower mobility in the diffusion process (resulting in a single *ambipolar* diffusion coefficient of the two species). The formation of diffusion potential can be derived on the basis of Figure 9.12.

The flow of 1 faraday (96 485 C) will result in a movement of t_i/z_i mole of i th kind of ions from section 1 to section 2 if z_i is positive and from section 2 to section 1 if z_i is negative. Thus, we have either enrichment (dc_i) or attenuation ($-dc_i$) of the two sections, and as a consequence of this process, the Gibbs energy also changes:

$$dG = \sum \frac{t_i}{z_i} d\eta_{\text{chem},i} \quad (9.35)$$

By substituting Eqs. (9.8) and (9.27) in Eq. (9.35), we obtain the general expression for the diffusion potential:

$$\nabla \varphi_d = \frac{k_B T}{e} \sum \int_{a_{i,1}}^{a_{i,2}} \frac{t_i}{z_i} d \ln a_i \quad (9.36)$$

If one assumes that the only difference between the electrolytes is the concentration of the mobile ions and the transference number is independent of the concentration, Eq. (9.36) can be easily differentiated and leads to the expression:

$$\Delta \varphi_d = \sum \frac{t_i}{z_i} \frac{k_B T}{e} \ln \frac{a_2}{a_1} \quad (9.37)$$

or by taking into account Eq. (9.33) for z,z -electrolytes ($z_{\text{cation}} = -z_{\text{anion}}$), we obtain a simple form of the Henderson equation:

$$\Delta\varphi_d = \frac{\mu_+ - \mu_-}{\mu_+ + \mu_-} \frac{k_B T}{e} \ln \frac{a_2}{a_1} \quad (9.38)$$

Example for different types of diffusion potentials are presented in Figure 9.13.

The diffusion potentials in liquids can be eliminated by using electrolyte bridges connecting the anodic and the cathodic compartments. The electrolyte in the bridge must be composed of an electrolyte with equal ion mobilities, for example, KCl or NH_4NO_3 at high concentrations.

For solid electrolytes, the situation is more complicated because electrolytic bridges cannot be used. Thus, diffusion potentials can only be eliminated with difficulty (if at all). In addition, because some of the solid materials used to transport ions do not contain mobile ions initially, situation as shown in Figure 9.12d can be observed where anions and cations diffuse in different directions [20] and equations of the type of Eqs. (9.28) and (9.37) have to be considered together, and the total emf of the cell of the type Me/insulator/Pt(H_2O) is given by [20]:

$$\Delta\phi = \Delta\varphi^\circ + C + \bar{t}_{\text{OH}^-} \frac{k_B T}{e} \ln a_{\text{Me}^+} + \bar{t}_{\text{Me}^+} \frac{k_B T}{e} \ln a_{\text{OH}^-} \quad (9.39)$$

with C being a constant.

The emf can have different origins and the nanosize effects, and the space-charge layers make the description of these processes in nanoionic systems quite complicated.

9.6

General Transport Processes and Chemical Reactions

The classical thermodynamics describes systems in equilibrium. However, often, we have to deal with systems that are metastable or far from equilibrium. In other words, the thermodynamics provides information about the magnitude of the driving forces and the possible direction but does not bear any information about the path or the rate. There are two examples that make this difference more clear:

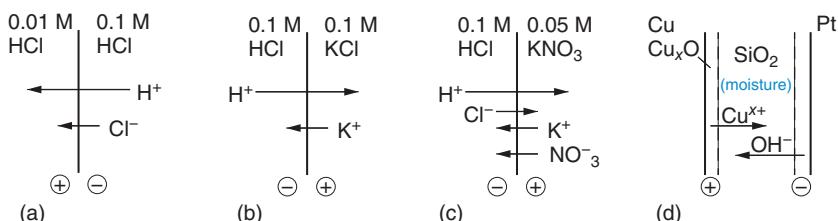


Figure 9.13 (a–d) Situation in insulating materials where the opposite charges are attracted. Situation (d) represents a model ECM cell based on SiO_2 solid electrolyte and Cu active electrode.

TiN is a material widely used as a functional film and a protective layer. Theoretically, when in contact with air, it must completely transform to TiO_2 , according to the reaction: $2\text{TiN} + 2\text{O}_2 = 2\text{TiO}_2 + \text{N}_2$, which has a negative Gibbs free energy (should spontaneous proceed to the right). However, at room temperature, TiN remains stable for many years without being oxidized. The reason is that the kinetics of the process is very slow because of the low mobility of the nitrogen ions. Thus, only the first TiN layer partially oxidizes, forming TiON, whereas the underlying material remains unreacted. Answers to such types of questions are given by the *irreversible thermodynamics*. It applies thermodynamic principles to nonequilibrium systems and offers ways for them to reach equilibrium or at least steady state. The main aim of the irreversible thermodynamics is to relate the rate (kinetics) to the driving force (thermodynamics). A starting point is the dissipation of energy being a product of the temperature and the entropy production (σ_S). It is related to a generalized force X as follows:

$$T\sigma_S = \sum_i J_i X_i, \quad (9.40)$$

where the flux J and the force X are linearly related through the phenomenological coefficient L .

$$J_i = \sum_i \sum_j L_{ij} X_i X_j \quad (9.41)$$

In fact, the flux of J_i is related to not only the conjugated force but also to other present forces. The phenomenological coefficients L_{ii} and L_{jj} are the direct coefficients, and L_{ij} and L_{ji} are denoted as coupling coefficients. It has been postulated by Onsager that $L_{ij} = L_{ji}$. In a simple case, that is, $J_i = L_i X_i$, several flux–force relations can be defined:

$$\vec{J}_Q = -\kappa \nabla T; \quad \vec{J}_m = -\frac{D_m c_m}{k_B T} \nabla \eta_m; \quad \vec{J}_q = -\frac{\sigma}{ze} \nabla \varphi,$$

representing the heat flux (Fourier's law), the mass flux (Fick's first law), and the charge flux (Ohm's law), respectively.

By using Eq. (9.41), one can describe the transport processes, chemical reaction, and so on. More detailed description of the flux equations and transport processes is presented in Chapter 5.

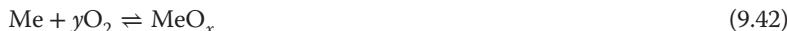
9.7

Solid-State Reactions

The chemical reactions are driven by the differences in the (electro)chemical potentials of the participating species and Eq. (9.7) (alternatively, Eq. (9.3) also applies in the same way). As a special example for a reaction driven by a general force $X = \nabla \eta$ is the *oxidation of metals*. Here, both the electric potential and the chemical potential gradients are important to be considered and different boundary cases can be distinguished. Moreover, the oxidation of metals was in

fact one of the main reasons for the development of the field of defect chemistry and solid-state ionics. Since this is essential and educative for the subject of ReRAM elements, in particular the operation of VCM and TCM cells, we will discuss the topic of metal oxide film formation in more detail.

The formation of oxide films on metal surfaces is a complex process involving redox reactions, nucleation and growth, and a transport of mass and charge across interfaces. In the case of chemical oxidation of (non-noble) metals, the driving force for oxide formation is the highly negative Gibbs free energy of the reaction:



After the formation of the first few monolayers, we can distinguish two interfaces at which the two half-cell reactions proceed. At the Me/MeO_x interface, the reaction



proceeds to the left, because most of the metals (in fact, Me^{z+}/Me redox couples) are characterized by an equilibrium redox potential more negative than that of the oxygen (O_2/O^{2-}). The oxygen redox reaction (see Eq. (9.25)) proceeds to the right and takes place at the metal oxide/gas (e.g., oxygen) interface. As shown in Figure 9.14, different situations are possible, depending on the properties of the formed initial oxide scale.

Very large lattice misfit and specific volume between the metal and oxide phases result in the formation of porous oxide films without strong adhesion on the metallic substrate, for example, Fe_3O_4 on Fe (Figure 9.14a). In this case, called *rusting*, the oxidation process proceeds fast until the Fe metal is fully consumed.

In the case of dense films (Figure 9.14b), two main aspects have to be discussed: the active ions, that is, the type of the most mobile ions, and the growth kinetics, that is, the time evolution of the thickness of the oxide film. The first aspect determines the position of a marker layer (e.g., by isotope tracers) in the course of the film growth. The two boundary cases for dense films (Figure 9.14b) are schematically shown in Figure 9.14c. If the oxygen ions (typically through a vacancy mechanism) are more mobile than the cations, the film grows toward the metal substrate and the markers remain at the surface. If the metal cations are more mobile, the opposite situation holds.

The second aspect, the growth kinetics of dense films, depends on the origin of the ion mobility, that is, (i) if the mobility is due to high (internal) electric fields or (ii) predominantly by temperature activation, and on the rate-determining step of the entire oxidation process, that is, if the ion transport in the oxide film is rate-limiting or (iii) if the oxygen incorporation reaction at the oxide surface is rate-limiting. These three cases are discussed in somewhat more detail:

It depends on the transport properties of the oxide layer. If it is predominantly insulating (this depends on the thermodynamic conditions such as defect concentrations and temperature), then its growth continues until the electric field generated by the driving force (see Eq. (9.28)) is compensated by the IR drop in the oxide. The film thickness increases and, therefore, the field decreases, and a

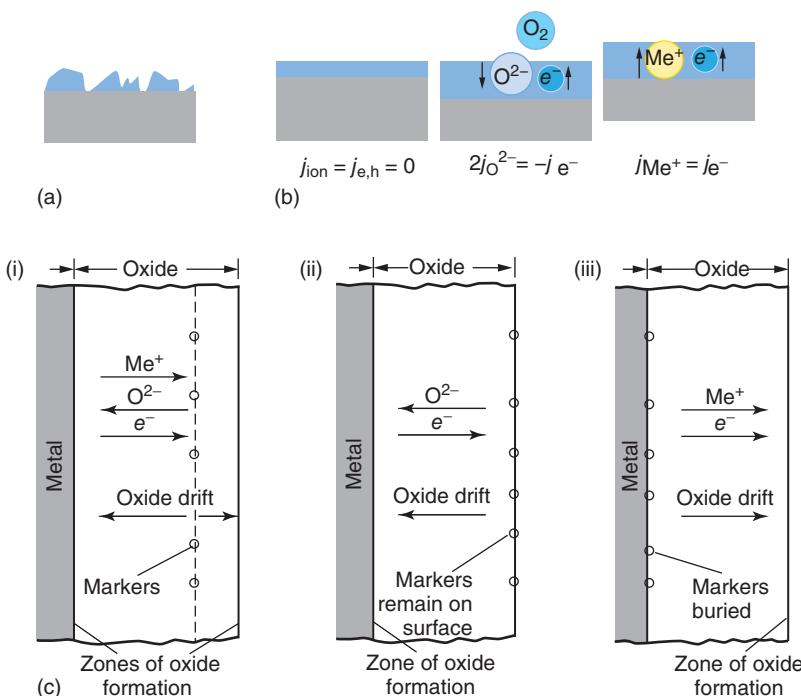


Figure 9.14 Different mechanisms of formation of oxide thin films on metal surfaces. (a) Formation of porous films, typically with poor adhesion to the metal substrate. (b) Formation of different types of dense films. (c) Growth mechanism accounting for

growth; (i) general processes; (ii) growth toward the metal substrate (anion dominated); and (iii) outward growth (cation dominated). The markers are used to determine the direction of film growth.

passive state of the system is achieved (important for the corrosion protection, e.g., of Al or Cr metal). If the ions and electrons are mobile within the film, further growth can proceed with a speed determined by the transport rate of the mobile charges. The boundary cases for dense films (Figure 9.14b) have been distinguished as schematically shown in Figure 9.14c:

- Formation of thin oxide films at low temperatures, that is, temperatures at which the thermal activation is not sufficient to ensure sufficient ionic mobility and the main driving force for ion movement (and, therefore, film growth) is field-accelerated ion mobility in the case of very high electric fields (approximately $>0.5 \text{ MV cm}^{-1}$). This is described by the Mott–Gurney law in Chapter 14. Because the interface potentials are fixed, sufficiently high fields occur as long as the film is still quite thin. Hence, the oxidation process is initially rapid but then the film growth rate decreases to negligible values. For that reason, Cr, Al, and Ti used as metals in everyday's articles show shiny metallic appearance although they are always covered

by a (very thin) oxide layer. The growth follows a logarithmic law:

$$d = k \ln(at + b) \quad (9.44)$$

where d is the film thickness, t is the time, a and b are constants, and k is the rate constant. Depending on the film thickness and the lattice parameters of the metal and the metal oxide, the films may appear pseudomorphic and compressed. Approximately 15% lattice misfit has been considered as the limit [49].

- (ii) Film growth at temperatures at which the thermally activated ion motion can ensure the oxide film growth, that is, the film can further grow due to the chemical potential gradient alone without the support by a high electric field as a driving force. The growth proceeds by a parabolic law:

$$d^2 = 2kt + d_0, \quad (9.45)$$

with k denoting the diffusion rate constant (proportional to the ionic mobility), and d_0 is a constant. This situation ensures the formation of thick oxide films. The parabolic law has been derived by Wagner [50] from constant diffusivities in decreasing concentration gradients because of increasing d . It is also known as *Wagner's parabolic law*.

Intermediate situations have been reported for, for example, Cu_xO but also for low temperatures and a variety of other oxide materials such as Ta_2O_5 or TiO_2 [51].

- (iii) A much simpler linear growth mode has also been reported for certain conditions of the completely covering oxide films (as in Figure 9.14b):

$$d = kt \quad (9.46)$$

This law applies to the formation of dense films when the diffusion in the film is fast compared to the incorporation reaction at the surface, so that the surface reaction becomes the rate-limiting step.

The time evolution of the thickness for the cases (i)–(iii) is sketched in Figure 9.15.

A detailed overview on the different oxidation modes can be found in Ref. [49].

The oxidation rate of metals can be significantly increased (forced) by applying positive voltage to the electrode to be passivated in an electrochemical cell

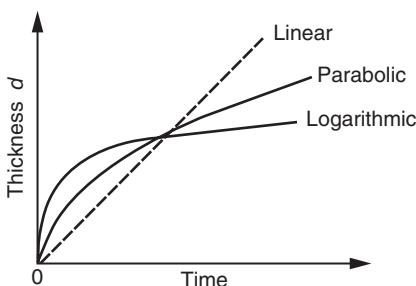


Figure 9.15 Different growth modes for oxide layers on metal substrates. The figure is adapted from Ref. [49].

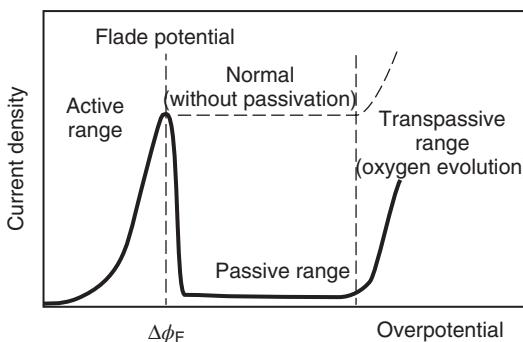
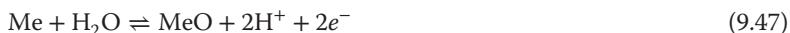


Figure 9.16 Current–voltage dependence during oxidation and passivation of metal electrodes. The metal is in contact with an aqueous electrolyte, and it is positively polarized.

within an aqueous electrolyte. In this way, oxide films can be grown on many transition or refractory metal electrodes. Depending on the electrolyte composition, these oxide films can be dense (for passivation and protection from corrosion) or can be porous and offer particular functionalities, for example, porous Al_2O_3 is used for hosting dyes for optical coloring or incorporation of functional lacks. Porous anodic TiO_2 is used as an electrode catalyst for photodissociation of water, and so on. In either of the cases, the kinetics of formation and the thickness of the oxide film are strongly influenced by the electric field. The process of electrolytic (anodic) oxide formation on metals can be better understood on the basis of Figure 9.16.

The current–voltage dependence shows three characteristic ranges representative of the particular electrochemical processes. The active range shows the active oxidation and dissolution of the metal electrode. On reaching the potential $\Delta\phi_F$, the reaction of formation of MeO takes place:



The reaction product is insoluble and leads to passivation of the metal surface and to rapid drop in the current (reaction rate). $\Delta\phi_F$ is termed as *passivation potential* or *Flade potential*.

The following range is the passive range and is characterized by very small currents. In this range, the oxide film is stable and grows further (depending on the applied voltage i.e. the electric field strength).

The third range shown in Figure 9.16 is called *transpassive*, and the increase of the current is due to the oxygen evolution reaction. If the potential is further increased, the oxygen evolution is usually accompanied by mechanical destruction of the oxide film.

The anodic oxidation of metals is an important part of industrial processes, and for many metals, for example, Ta, Ti, Al, W, Zr, and Hf, it is studied in details under different conditions to ensure proper film functionalities.

9.8

Electrochemical (Electrode) Reactions

Electrochemical electrode reactions can be considered as a more generalized case of chemical reactions. The difference between these two types of reactions is the driving force – the chemical potential difference, and the resulting negative Gibbs reaction energy (Eqs. (9.3) and (9.27)) leads to a “purely” chemical reaction, whereas by applying voltage to an electrochemical cell, one changes the electrochemical potential via the electrical component (Eqs. (9.7) and (9.8)). Therefore, the theoretical derivation of the reaction rate equations is in accordance to the classical transition-state theory. The kinetics of the reaction expressed by the reaction rate v is given by:

$$v = k \cdot c \exp\left(-\frac{\Delta G^\neq}{k_B T}\right) \quad (9.48)$$

where c is the concentration of the reactant(s), ΔG^\neq is the free energy of activation, and k is a constant.

A schematic of the charge-transfer process in the coordinates of potential energy versus distance from the electrode surface is shown in Figure 9.17.

The reaction rate is directly proportional to the current density j of the particular reaction in accordance to:

$$v = \frac{j}{ze} \quad (9.49)$$

or by combining Eqs. (9.48) and (9.49), one obtains:

$$j = z \cdot e \cdot k \cdot c \exp\left(-\frac{\Delta G^\neq}{k_B T}\right). \quad (9.50)$$

9.8.1

Charge-Transfer Process Limitations

Equation (9.50) is a basic equation for the derivation of the equation for the charge-transfer kinetics, also denoted as Butler–Volmer equation [52, 53].

$$j = j_0 \left[\exp\left(\frac{(1-\alpha)z_i e}{k_B T} \Delta\varphi_{\text{anodic}}\right) - \exp\left(-\frac{\alpha z_i e}{k_B T} \Delta\varphi_{\text{cathodic}}\right) \right], \quad (9.51)$$

where j is the current density, j_0 is the exchange current density, and α is the transfer coefficient ($0 < \alpha < 1$). The transfer coefficient can be understood in terms of geometrical position of the activation energy maximum in the Helmholtz double layer, or alternatively, it can be formulated as the effective change in the activation energy per unit increase in the applied potential. As it can be seen, the sum of the transfer coefficients for the anodic and the cathodic reactions must be 1.

The exchange current density j_0 is defined at zero net current and zero overpotential as:

$$j_0 = |j_{\text{cathodic}}| = j_{\text{anodic}}, \quad (9.52)$$

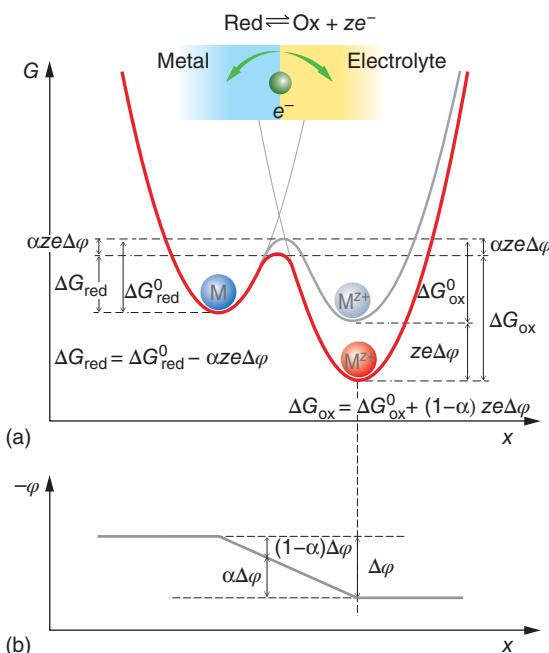


Figure 9.17 (a) Potential energy profile of an electron-transfer reaction at the electrode–electrolyte interface. The gray line represents equilibrium conditions. The

red line shows the situation with applied potential. (b) Electrical potential distribution within the inner dense part of the Helmholtz double layer. From Ref. [13].

where j_0 represents the natural exchange of matter and charge between the electrodes and the electrolyte under equilibrium conditions. Therefore, the fluxes in both directions are equal so that the net current remains zero. Under this condition, Eq. (9.51) transforms to the Nernst equation (e.g., Eq. (9.28)).

A detailed derivation of the Butler–Volmer equation can be found in general textbooks on physical chemistry/electrochemistry, for example [5].

Equation (9.51) has a general validity and applies to electrochemical reactions, where the effect of the double-layer structure can be neglected. In the cases of multielectron transfer steps or multistep reaction, the equation requires additional amendments.

For low overvoltages, for example, $\Delta\varphi \ll k_B T/e$ close to equilibrium, both terms in Eq. (9.51) should be considered. The equation can be then approximated to the linear dependence:

$$j = \frac{j_0 ze}{kT} \Delta\varphi = \frac{1}{R_p} \Delta\varphi, \quad (9.53)$$

where R_p is often called *polarization resistance*.

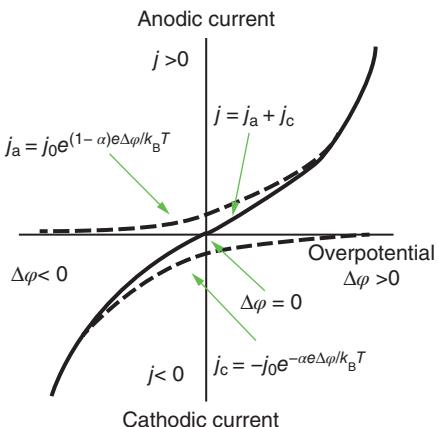


Figure 9.18 Current density versus overpotential in accordance to Butler–Volmer equation. The individual partial currents are represented by dashed lines. The total current is the sum of the partial currents. $j_{\text{total}} = j_a + j_c$.

At large overpotentials, $|\Delta\varphi| > k_B T/e$, one of the exponential terms becomes negligible for the particular half-cell reaction. It should be noted that Eq. (9.51) is directly applicable to single-electron, single-step charge transfer. Additional processes such as nucleation phenomena, multielectron and/or multistep charge transfers, or electrode processes preceded or followed by chemical reactions require additional modification of the equation [4].

The current density–overpotential ($j - \Delta\varphi$) dependence for both the regions and also the individual partial current densities are shown in Figure 9.18.

9.8.2

Diffusion-Limited Electrochemical Processes

If diffusion of reactants and/or products supplied to the electrode/electrolyte interface or away from it limits the overall reaction rate, the overvoltage is classified as diffusion-limited electrode reaction and the corresponding voltage required to be applied to overcome this limitation is the diffusion overvoltage ($\Delta\varphi_{\text{diff}}$). The current–voltage dependence under steady-state conditions is given by:

$$\Delta\varphi_{\text{diff}} = \frac{k_B T}{ze} \ln \left(1 - \frac{j}{j_d} \right) \quad (9.54)$$

In this equation, j_d is the diffusion-limited current. It remains constant and independent of the increasing voltage. The diffusion current depends on the concentration of reactive species and their diffusion coefficient:

$$j_d = z_i e D_i \frac{c_i - c_i^s}{\delta}, \quad (9.55)$$

where c_i^s is the surface concentration and δ is the thickness of the diffusion layer. More precisely, the condition for the constant diffusion-limited current is reached when the surface concentration c_i^s drops to 0.

Diffusion-limited processes will be of importance in systems using materials with low ion conductivity or larger thicknesses. Because a countercharge is required, often the oxidized metal ions cannot enter the electrolyte and then typical diffusion limitation occurs. Detailed description on diffusion processes in solids is given in Chapter 5.

9.9

Stoichiometry Polarization

A ReRAM cell is typically composed of an electrolyte with more than one mobile carrier. Often, these are mixed ionic–electronic conducting electrolytes (with one kind of ions and electrons) in VCM-, TCM-, and some ECM-type cells or, more rarely, purely ionic-conducting ECM cells with redox-active cations (e.g., Ag^+) and corresponding counter ions. In general, the metal electrodes will show different transparencies for the two types of carriers. In the case of mixed ionic–electronic conducting oxides, the electrodes will hardly block the electronic current while they may partially or fully block the ionic current. If a DC voltage is applied to such a cell, a *kinetic demixing* (also called a *stoichiometry polarization* or *concentration polarization*) will take place, which leads to the formation of gradients of the carriers between the electrodes. The magnitude of the carrier concentration difference and the time required to reach its steady state depend on the voltage and the material parameters.

For small voltages, this can easily be calculated analytically. The current density j in a mixed-conducting solid electrolyte is composed of an ionic and an electronic contribution:

$$j = j_{\text{ion}} + j_{\text{el}}. \quad (9.56)$$

Each partial current density obeys the general transport law, that is, it has a drift and a diffusion term,

$$j_k = j_{k,\text{drift}} + j_{k,\text{diff}} \quad (9.1)$$

$$= \sigma_k E + z_k e D \frac{\partial N_k}{\partial x} \quad (9.2)$$

for $k = \text{ion}$, or electron (Precisely speaking, here, the mobile species k are to be used in the sense of conservative ensembles as defined in Ref. [54].) In Eq. (9.2), D denotes the ambipolar diffusion. In the case of sufficiently small voltages V applied, the material properties σ_k and D will remain approximately constant, that is, they will not change with time or position. This condition is satisfied when the final change in the concentration is small, that is, $\Delta N_k \ll N_k$. When the electrodes are assumed to be blocking the ionic current, this partial current will vanish in the

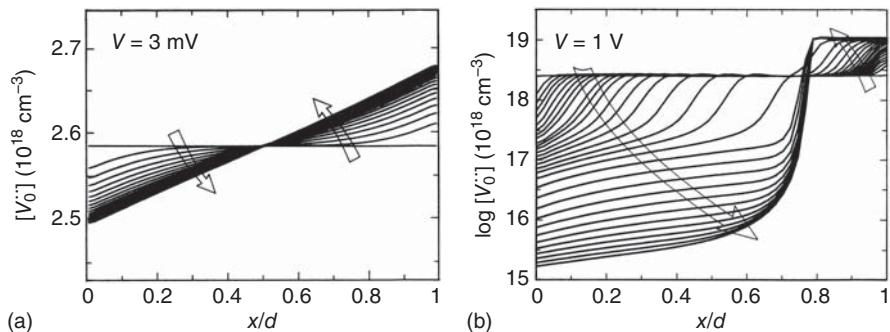


Figure 9.19 Calculated time evolution of the oxygen vacancy profiles after applying a DC voltage V to a 0.1 at % acceptor-doped SrTiO_3 crystal of thickness $d = 0.1$ cm at $T = 500$ K for two different voltages. The electrodes are assumed to block the ion transfer. (a) $V = 3$ mV, within the validity range of the linear transport theory and

(b) $V = 1$ V, leading to highly nonlinear concentration profiles. Please note the very different scales of the ordinate: (a) a small interval on the linear scale and (b) several orders of magnitude on the logarithmic scale. For the concentration symbol, the bracket [] notation is used. (From Ref. [55]; details are provided in Ref. [56].)

steady state, that is, $j_{\text{ion}} = 0$ for $t \rightarrow \infty$. From Eq. (9.2), the steady-state concentration gradient can be calculated:

$$\frac{\partial N_{\text{ion}}}{\partial x} = \frac{\sigma_{\text{ion}} E}{z_{\text{ion}} e D} \quad (9.59)$$

that is, a linear profile of the concentration $N_{\text{ion}}(x)$. Integration of Eq. (9.59) shows that the concentration difference between the electrodes ΔN_{ion} is proportional to the voltage difference $\Delta\varphi$:

$$\Delta\varphi = \frac{z_{\text{ion}} e D}{\sigma_{\text{ion}}} \Delta N_{\text{ion}}. \quad (9.60)$$

Because of the conservation of the total amount of mobile ions, the profile of $N_{\text{ion}}(x)$ is symmetrical with respect to the initial concentration $N_{\text{ion}}(t = 0)$. A numerically calculated time evolution of the profile of the example of the oxygen vacancy concentration is shown in Figure 9.19a for slightly acceptor-doped SrTiO_3 in a macroscopic cell for $\Delta\varphi = 3$ mV. Corresponding profiles of the electronic carrier concentrations are established in order to maintain local electroneutrality, except for the space-charge regions at the electrodes.

If higher voltages are applied, the concentration changes can be enormous, and highly nonlinear concentration profiles are typically obtained (Figure 9.19b). In the case of large changes of the concentration of mobile ionic defects, the formation of new phases may occur when the solubility limit of the point defects is exceeded and, as a consequence, the range of the original stoichiometry is exceeded. An example is the formation of Magnéli phases in TiO_2 cells [57].

The case of ionic conductors with two mobile ionic carriers of different mobilities is comprehensively described by Martin [58]. As an example, Figure 9.20

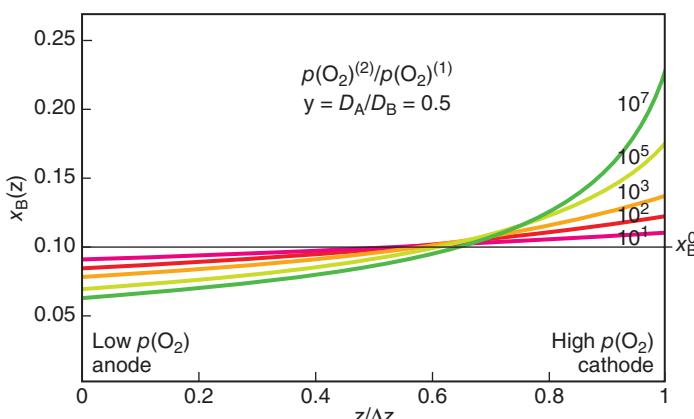


Figure 9.20 Calculated steady-state demixing profiles, $x_B(z)$, of the dopant B in an oxygen ion conductor AO_2 ($+\text{B}_2\text{O}_3$) for a ratio of the diffusion coefficients $\gamma = D_A/D_B = 0.5$ and different values of ratio

of oxygen partial pressures on both sides of the cell $p\text{O}_2(2)/p\text{O}_2(1)$. The initial dopant fraction is $x_B^0 = 0.1$. Details are provided in Ref. [58].

shows the concentration profiles for the fraction B in a composition AO_2 ($+\text{B}_2\text{O}_3$) for different values of the oxygen partial pressure $p\text{O}_2$ on both sides. Please note that in Figure 9.20, the set of profiles represent the steady-state profiles for various $p\text{O}_2$ ratios, while in Figure 9.19, the set of profiles show the time evolution of the profiles.

Summary

In addition to the previous chapters in this book, we have introduced thermodynamics and kinetics fundamentals that are relevant for understanding the microscopic mechanisms in ReRAM cells. We have discussed the formation of new phases and phase stability, the formation of space-charge layers and segregation, including kinetic demixing. Furthermore, details are given on solid-state reactions, electrode redox kinetics and possible rate limitations, and – as an example – the different routes of formation of thin oxide layers at metal/oxide interfaces. The nanosize effects leading to deviation from the classical macroscopic form are highlighted and related to the physics and chemistry of the ReRAM cells, their performance, and ways for control and improvement. It has been emphasized that due to these deviations, materials that are insulators on the macroscopic scale, for example, Al_2O_3 , Ta_2O_5 , and SiO_2 , allow for ion transport, a fact that is utilized in ReRAMs. It has also been shown that, at the nanoscale, the properties of matter are often quantized, showing discrete values, for example, in a small number of atoms that constitute the critical size of a nucleus during electrocrystallization.

All these issues should be considered for the microscopic physical description of ReRAMs and for the design of materials systems and devices with improved performance.

Acknowledgments

The authors are indebted to Susanne Hoffmann-Eifert (FZ Jülich, Germany) and Dirk Wouters (RWTH Aachen, Germany) for carefully reading the manuscript and for their helpful discussions and Michael Lübben for the support by the proof reading. We would also like to acknowledge the support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

- Schmalzried, H. and Navrotsky, A. (2004) *Festkörperthermodynamik*, Verlag Chemie, p. 122.
- Schmalzried, H. (1995) *Chemical Kinetics of Solids*, Wiley-VCH Verlag GmbH.
- Maier, J. (2004) *Physical Chemistry of Ionic Materials*, John Wiley & Sons, Ltd.
- Bard, A. and Faulkner, L. (2001) *Electrochemical Methods: Fundamentals and Applications*, John Wiley & Sons, Inc., New York.
- Rieger, P.H. (1994) *Electrochemistry*, Chapman & Hall, Inc., New York.
- Rickert, H. (1982) *Electrochemistry of Solids*, Springer.
- Abriata, J.P., Garces, J., and Versaci, R. (1986) *Bull. Alloy Phase Diagrams*, **7** (2), 116–124.
- Van Landuyt, J. and Amelinckx, S. (1973) *J. Solid State Chem.*, **6**, 222.
- Magneli, A. (1978) *Pure Appl. Chem.*, **50** (11-12), 1261–1271.
- Valov, I. and Koziicki, M.N. (2013) *J. Phys. D Appl. Phys.*, **46**, 074005.
- Valov, I., Waser, R., Jameson, J.R., and Koziicki, M.N. (2011) *Nanotechnology*, **22** (25), 254003/1–254003/22.
- Hasegawa, T., Terabe, K., Tsuruoka, T., and Aono, M. (2012) *Adv. Mater.*, **24** (2), 252–267.
- Waser, R., Bruchhaus, R., and Menzel, S. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 683–710.
- Ielmini, D., Bruchhaus, R., and Waser, R. (2011) *Phase Transitions*, **84** (7), 570–602.
- Yang, L., Kuegeler, C., Szot, K., Ruediger, A., and Waser, R. (2009) *Appl. Phys. Lett.*, **95** (1), 013109.
- Huang, Y., Lin, H., and Cheng, H. (2013) Proceedings of the 2013 IEEE 5th International Nanoelectronics Conference (INEC).
- Jeong, D.S., Schroeder, H., and Waser, R. (2007) *Electrochem. Solid-State Lett.*, **10** (8), G51–G53.
- Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) *Nature*, **453** (7191), 80–83.
- Kim, S., Kim, K., Jeong, D., Jeon, W., Yoon, K., and Hwang, C. (2013) *J. Mater. Res.*, **28** (3), 313–325.
- Valov, I. et al. (2013) *Nat. Commun.*, **4**, 1771.
- Tappertzhofen, S., Waser, R., and Valov, I. (2014) *ChemElectroChem*, **1**, 1287–1292.
- Tappertzhofen, S., Menzel, S., Valov, I., and Waser, R. (2011) *Appl. Phys. Lett.*, **99** (20), 203103/1–203103/3.
- Vollman, M. and Waser, R. (1994) *J. Am. Ceram. Soc.*, **77** (1), 235–243.
- Waser, R. and Hagenbeck, R. (2000) *Acta Mater.*, **48** (4), 797–825.
- Morrison, S.R. (1980) *Electrochemistry at Semiconductor and Oxidized Metal Electrodes*, Plenum Press, New York.
- Maier, J. (1995) *Prog. Solid State Chem.*, **23** (3), 171–263.
- Marchewka, A. et al. (2014) *Sci. Rep.*, **4**, 6975.
- Kim, S., Fleig, J., and Maier, J. (2003) *Phys. Chem. Chem. Phys.*, **5**, 2268–2273.

29. Sata, N., Eberman, K., Eberl, K., and Maier, J. (2000) *Lett. Nat.*, **408**, 946–949.
30. Maier, J. (2004) *Solid State Ion.*, **175**, 7–12.
31. Chiang, Y., Lavik, E., and Blom, D. (1997) *Nanostruct. Mater.*, **9**, 633–642.
32. Terwilligert, C.D. and Chiang, Y. (1995) *Acta Metall. Mater.*, **43**, 319–328.
33. Guo, X., Vasco, E., Mi, S., Szot, K., Wachsman, E., and Waser, R. (2005) *Acta Mater.*, **53** (19), 5161–5166.
34. Mitkova, M., Kozicki, M.N., Kim, H.C., and Alford, T.L. (2006 21st International Conference on Amorphous and Nanocrystalline Semiconductors, Lisbon, PORTUGAL) *J. Non Cryst. Solids*, **352**, 1986–1990.
35. Verkerk, M., Middelhuis, B., and Burggraaf, A. (1982) *Solid State Ion.*, **6**, 159–170.
36. Liang, C.C. (1973) *J. Electrochem. Soc.*, **120** (10), 1289–1292.
37. Shahi, K. and Wagner, J.B. Jr. (1980) *Appl. Phys. Lett.*, **37** (8), 757–759.
38. Horton, J.H., Peat, K.L., and Lambert, R.M. (1993) *J. Phys. Condens. Matter*, **5**, 9037.
39. Horton, J.H., Hardacre, C., Baddeley, C.J., Moggridge, G.D., Ormerod, R.M., and Lambert, R.M. (1995) *Phys. Rev. B*, **52**, 2054–2062.
40. Cho, D.Y., Tappertzhofen, S., Waser, R., and Valov, I. (2013) *Nanoscale*, **5** (5), 1781–1784.
41. Milchev, A., Stoyanov, S., and Kaischev, R. (1974) *Thin Solid Films*, **22**, 255–265.
42. Becker, R. and Doering, W. (1935) *Ann. Phys.*, **416**, 719–752.
43. Staikov, G. (2007) *Electrocrystallization in Nanotechnology*, Wiley-VCH Verlag GmbH.
44. Budevski, E., Staikov, G., and Lorenz, W.J. (1996) *Electrochemical Phase Formation and Growth*, VCH Publishers.
45. Valov, I. et al (2012) *Nat. Mater.*, **11** (6), 530–535.
46. Valov, I. and Staikov, G. (2013) *J. Solid State Electrochem.*, **17** (2), 365–371.
47. Pletcher, D. and Schiffrian, D.J. (1983) *The Electrochemistry of Oxygen*, Royal Society of Chemistry, pp. 126–170.
48. Vetter, K.J. (1961) *Electrochemical Kinetics*, Springer-Verlag.
49. Cabrera, N. and Mott, N.F. (1949) *Rep. Prog. Phys.*, **12**, 163–184.
50. Wagner, C. (1933) *Z. Phys. Chem.*, **B21**, 25–41.
51. Waber, J.T. (1952) *J. Chem. Phys.*, **20** (4), 734–735.
52. Butler, J. (1924) *Trans. Faraday Soc. (London)*, **19**, 734–739.
53. Erdey-Gruz, T. and Volmer, M. (1930) *Z. Phys. Chem. A (Leipzig)*, **150A**, 203–213.
54. Maier, J. and Schwitzgebel, G. (1982) *Phys. Status Solidi B, East Germany*, **113** (2), 535–547.
55. Waser, R. (1993) *Processing of Dielectric Titanates: Aspects of Degradation and Reliability, Ferroelectric Ceramics*, Birkhaeuser-Verlag, pp. 273–298.
56. Baiatu, T., Waser, R., and Hardtl, K.H. (1990) *J. Am. Ceram. Soc.*, **73** (6), 1663–1673.
57. Szot, K., Rogala, M., Speier, W., Klusek, Z., Besmehn, A., and Waser, R. (2011) *Nanotechnology*, **22** (25), 254001/1–254001/21.
58. Martin, M. (2000) *Solid State Ion.*, **136–137**, 331–337.

10

Electroforming Processes in Metal Oxide Resistive-Switching Cells

Doo Seok Jeong, Byung Joon Choi, and Cheol Seong Hwang

10.1

Introduction

Nonvolatile resistive-switching behaviors have been observed in a large number of materials systems, such as binary transition metal oxides (TMOs), perovskite-type complex TMOs, large-band-gap high- k oxides, and higher chalcogenides [1]. To achieve reversible resistive-switching behaviors, so-called electroforming (forming for short) process is required in advance. Among the diverse resistive-switching systems, however, this chapter covers only anion-based resistive-switching systems [2, 3] and their forming process. The forming process in other systems will be covered in the other chapters. A number of resistive-switching materials based on cation migration also have been reported [4, 5], and the working principles of the cation-based memories will be covered in the other chapters. Anion-based memories are popularly referred to as *valence change memories* (VCMs) [3], which denote nonvolatile resistive-switching results from valence change in cations. This valence change is triggered by the drift of anions (or their defects) due to the internal electric field. This is generally related with bipolar switching, but unipolar switching is also often observed in anion-migration-based switching systems, which are popularly referred to as *thermochemical memories* (TCMs). In this chapter, only a capacitor-like two-terminal metal–insulator–metal (MIM) structure is dealt with, and the more complicated three-terminal active structure (e.g., a transistor) will not be covered. In addition, an attempt to draw conclusions on the general forming features from case studies on particular switching materials is made. Thus, it should be kept in mind that there may be several exceptions.

VCM materials are most likely to be quite insulating in their pristine (i.e., as-fabricated) states. There are some exceptional switching materials, however, that are more or less leaky (i.e., semiconducting) in their pristine states. Thus, unless otherwise stated, in this chapter, VCM materials are regarded as initially insulating. It is known that initial treatments by means of bias application are required to assign a resistive-switching property to initially insulating VCM materials. This initial treatment is referred to as *forming*. Although the meaning of forming in

resistive switching differs from its traditional meaning, the electrochemical process producing a “macroscopic” metallic structure on a mandrel, they have several aspects in common (e.g., electrochemical process). Forming can be described in a phenomenological way, such as a process bringing about “nonvolatile and controllable” resistance degradation in insulating VCM materials in their as-fabricated states by imposing a high voltage and/or current stress. It should be noted that forming differs from “hard dielectric breakdown,” implying “permanent resistance degradation” that barely allows the high resistance to be recovered in any manner.

10.1.1

Forming Methods

One of the most common forming methods is applying a voltage/current sweep (i.e., the potentiodynamic method) to a VCM cell. Figure 10.1a,d illustrate the schematics of current–voltage (I - V) behaviors undergoing forming, realized by applying voltage and current sweeps (Figure 10.1b,e), respectively. Triangular voltage sweeps in a time domain (illustrated in Figure 10.1b), whose maximum is higher than a particular forming voltage (V_f), lead the VCM cell to a formed state. A first prominent signature of the forming process is a drastic change in the measured current (Figure 10.1c), as illustrated in Figure 10.1a. To avoid permanent resistance degradation (i.e., hard dielectric breakdown), a compliance current is

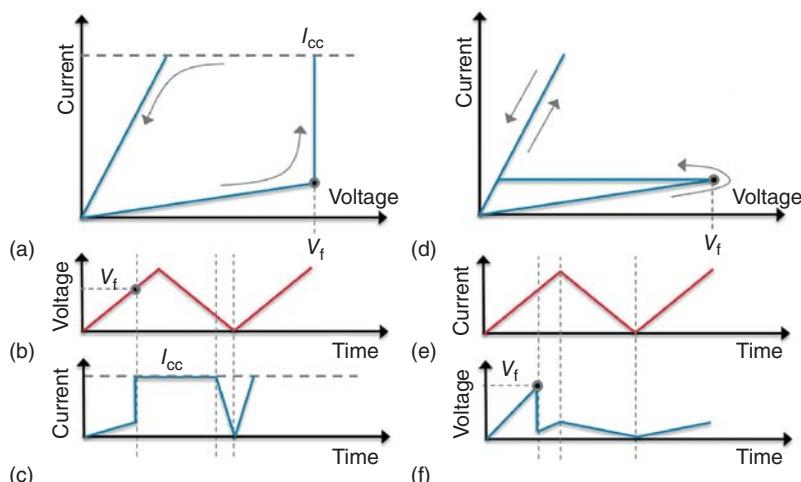


Figure 10.1 Schematics of the current–voltage behaviors undergoing electroforming with applied (a) voltage sweeps and (d) current sweeps. The applied input sweeps in a time domain for each case are

illustrated in (b) and (e), respectively. The corresponding output current and voltage in the time domains are also illustrated in (c) and (f), respectively.

usually required to be employed, which limits the current flow through the cell up to the set value. Gradual rather than abrupt changes in the current, however, are noticed during forming in some cases. A second signature is the large degradation of resistance immediately after the abrupt current change, which can be kept unless subsequent resistance manipulation processes, such as resistive-switching operation, follow.

In the case of current-sweep-driven electroforming, the *I-V* behavior differs from that of the voltage-sweep-driven case, as illustrated in Figure 10.1d. The most significant difference is the nonuse of a compliance current. In this method, the current flowing through the cell is always under control as no compliance functions are used. In fact, the use of a compliance current leads to the uncertainty that the actual amount of current flowing through the cell at the moment of the abrupt current jump shown in Figure 10.1a cannot be precisely evaluated due to the current overshoot occurring when the current reaches the compliance current [6–8]. Several technical issues resulting from this current overshoot problem are addressed in detail in Section 10.3. Other than the aforementioned potentiodynamic methods, potentiostatic methods (e.g., constant voltage or current application) can also be employed to realize the forming process. Inasmuch as potentiostatic methods enable easier quantitative analysis of forming, potentiostatic rather than potentiodynamic methods are suitable for forming mechanism studies. The difficulty of the potentiodynamic methods mostly arises from the time-dependent forming voltage variation (Figure 10.1f).

As will be addressed in detail, forming is known to lead to largely scattered as-electroformed resistance states in various TMOs, which perhaps arise from its stochastic nature [9]. Especially, the effect is well known in TiO_2 , one of the prototypical resistive-switching materials, sandwiched between symmetric and asymmetric electrodes [9, 10]. Among the aforementioned forming methods, it appears to result in the least scattered yield for the application of a negative current sweep to the top electrode of the TiO_2 -based cell at a rate of several hundred nanoamperes per second [10]. This is an example addressing the significant importance of not only a forming method (either potentiodynamic or potentiostatic) but also particular forming conditions, such as bias polarity and sweep rate, in the subsequent resistive-switching behavior. Given that the switching behavior strongly relies on the forming process, particular emphasis on forming is required when one deals with resistive-switching materials, unless they are forming-free.

10.1.2

Dependence of the Bipolar Switching Behavior on the Forming Conditions

Forming bias polarity perhaps determines the bipolar switching polarity (i.e., set and reset occur in positive and negative bias voltage, respectively, or vice versa), as observed in several switching cells (i.e., $\text{Pt}/\text{TiO}_2/\text{Pt}$ [9, 11], $\text{Pt}/\text{SrTiO}_3/\text{Au}$ [12]). An example is shown in Figure 10.2 Pt/ TiO_2 /Pt cells, formed by means of a potentiostatic method (constant current application) in vacuum, exhibit different bipolar switching polarities depending on the bias polarity. Figure 10.2a shows a

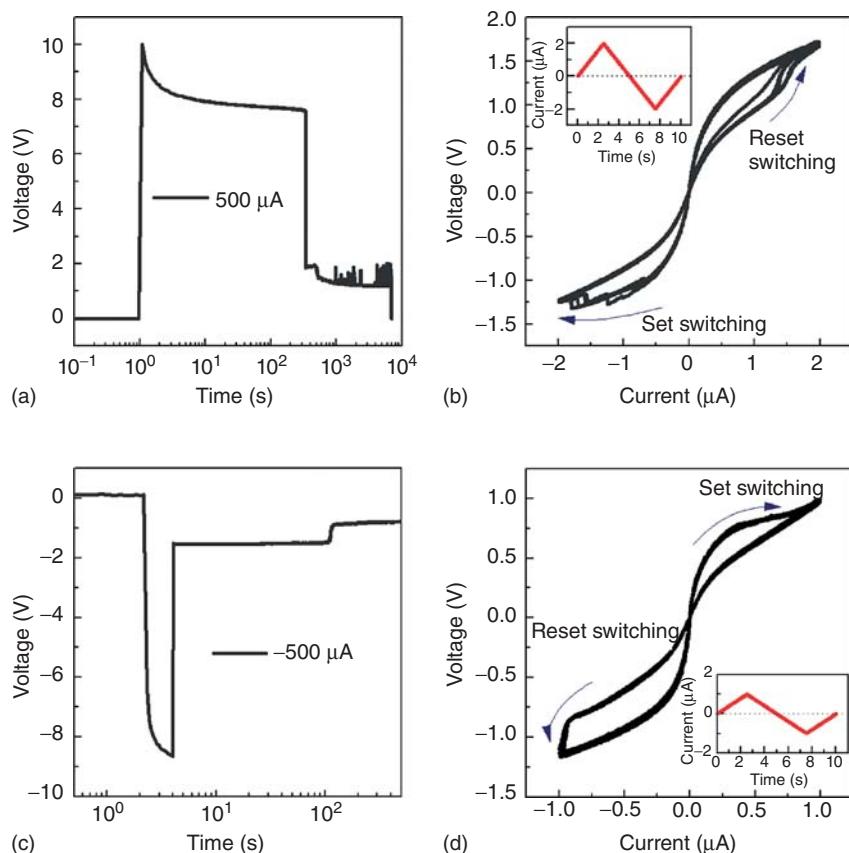


Figure 10.2 Bipolar switching behaviors depending on the forming bias polarity. (a) The positive constant current ($500 \mu\text{A}$) application to a Pt/TiO₂/Ti/Pt cell in vacuum showed the forming V - t profile and (b) consequently led to the bipolar switching I - V behavior: reset and set switching under a positive and a negative current, respectively.

(c) The negative constant current ($-500 \mu\text{A}$) application in vacuum, however, showed the forming V - t profile and (d) consequently led to the opposite bipolar switching behavior: reset and set switching under a negative and a positive current, respectively. (Redrawn from [9].)

voltage–time (V - t) relation at an applied current of $500 \mu\text{A}$, in which forming is completed at approximately 400 s . The subsequent voltage–current (V - I) curves are plotted in Figure 10.2b. It can be seen that reset and set switching takes place under a positive and a negative current, respectively. The cell, undergoing forming at an applied current of $-500 \mu\text{A}$, represents the V - t relation shown in Figure 10.2c. The forming was completed at approximately 100 s . The subsequent hysteretic V - I curves are shown in Figure 10.2d, whose switching polarity is opposite. That is, reset and set switching occurs under a negative and a positive current, respectively.

Unlike the forming behaviors shown in Figure 10.2a,c a10.310.3a, multiple forming steps with alternating forming bias polarity are sometimes required to complete the forming process. In the case of multiple forming processes employing a potentiodynamic method, the *I-V* curve, corresponding to the last forming step, may not be easily distinguished from that of set switching, as can be seen in Figure 10.3b. Nevertheless, the difference between the two similar-looking phenomena lies in the threshold voltage. The threshold voltage for the last forming process (i.e., “1st ON” shown in Figure 10.3b) is obviously higher than that for the reversible set switching. Thus, it may be reasonable to define all processes required to realize reversible switching behavior as forming

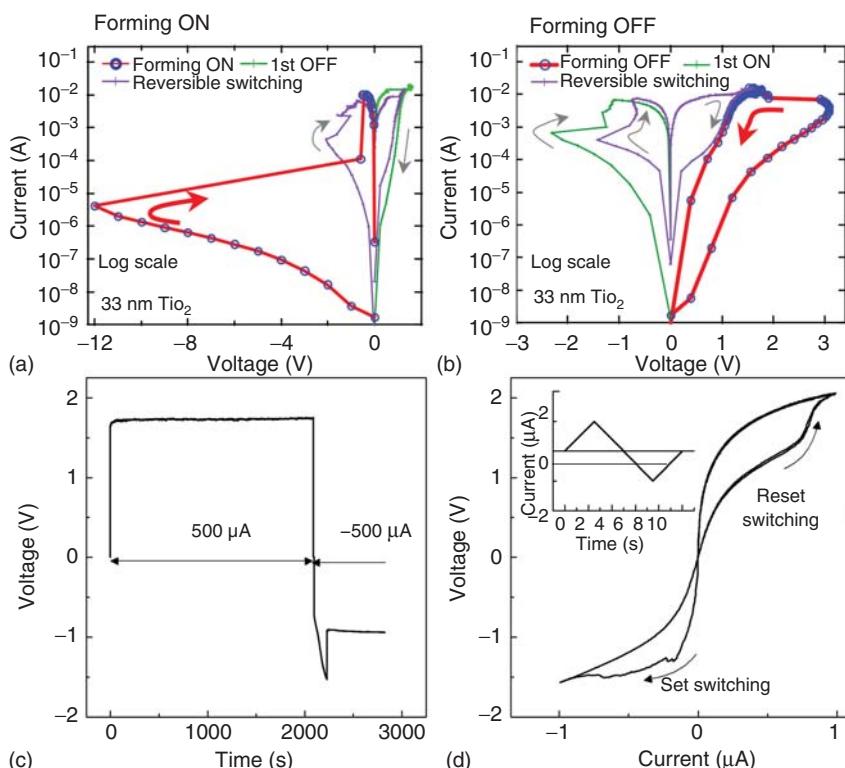


Figure 10.3 Forming processes required for Pt/TiO₂/Pt switching cells. Potentiodynamic methods (i.e., (a) a negative voltage sweep and (b) a positive voltage sweep) and a potentiostatic method (i.e., application of constant currents alternating the polarity at 2000 s) were employed to form the cells. (a) *I-V* curve representing a single forming process indicated by “Forming ON.” (b) *I-V* curves representing a multiple forming

process (i.e., first, a positive, and second, a negative voltage sweep, denoted by “Forming OFF” and “1st ON,” respectively). (c) *V-t* profile exhibiting a multiple forming process. (d) Bipolar switching *V-I* curves of the cell that underwent the forming process shown in (c). (Reprinted with permission from [9, 13]. © 2008, American Institute of Physics, and 2009, IOP Publishing, respectively.)

processes. Figure 10.3c shows a $V-t$ profile of the forming process performed by applying consequent constant currents altering the polarity at approximately 2000 s. The first forming step was finished immediately after the 500 μA current application, for which reason the $V-t$ profile does not show the immediate response. The consequent bipolar switching behavior is plotted in Figure 10.3d.

10.1.3

Factors Influencing Forming Behavior

The comparison of Figures 10.2a and 10.3c identifies the effect of the atmosphere on the forming behavior, where forming was performed in vacuum and air, respectively. In fact, atmosphere-dependent forming characteristics were already reported in the 1960s and 1970s, and their details will be addressed later. This dependence perhaps points to the fact that the switching cell reacts with ambient gas molecules so that it should be regarded as an open rather than a closed system. Moreover, it is determined that forming is not a purely electronic process; instead, coupling between ionic and electronic phenomena is likely of considerable importance. Among the different kinds of gases in the air, it is probable that oxygen gas molecules play a key role in influencing the forming behavior, as discussed in the mechanism section. Indirect evidence for this interaction was found in Fe-doped SrTiO_3 , in which the influence of oxygen partial pressure on the resistance of the Fe-doped SrTiO_3 -based cell formed through a potentiostatic method was observed (see Figure 10.4) [14]. An increase in oxygen partial pressure led to a large increase in resistance, by approximately 2 orders of magnitude. As expected, the formed cell underwent oxidation in high oxygen partial pressure, which resulted in increased resistance. Thus, it is considered that the forming process may be equivalent to the reduction of the switching cell in the sense that it drives oxygen ions out of the cell.

Ambient temperature also determines forming behavior [15]. In $\text{Pt/TiO}_2/\text{Pt}$ switching cells, the ambient temperature has been found to notably influence the forming kinetics (i.e., forming time) observed under a constant voltage, as shown in Figure 10.5. The forming time tends to decrease with increasing ambient temperature, implying that the forming process is activated by the temperature. Figure 10.5 shows that the thermal activation becomes obvious at low voltages ($<\sim 4 \text{ V}$); the forming time is largely different at different temperatures. Increasing voltage, however, results in a weak dependence of the forming time on the ambient temperature. This is particularly obvious in the high-voltage region as long as the voltages do not lead to permanent dielectric breakdown rather than forming. This weak dependence may be attributed to the fact that the high power dissipation under the high voltages causes a temperature increase that overwhelms the relatively small difference in the ambient temperature [15].

Microstructure of switching materials, for example, grain size, appears to affect forming voltage [16]. Grain boundaries are two-dimensional microstructural defects so that they influence forming behavior to some extent regarding the role of defects in forming, which is dealt with in Section 10.2. The smaller the grain

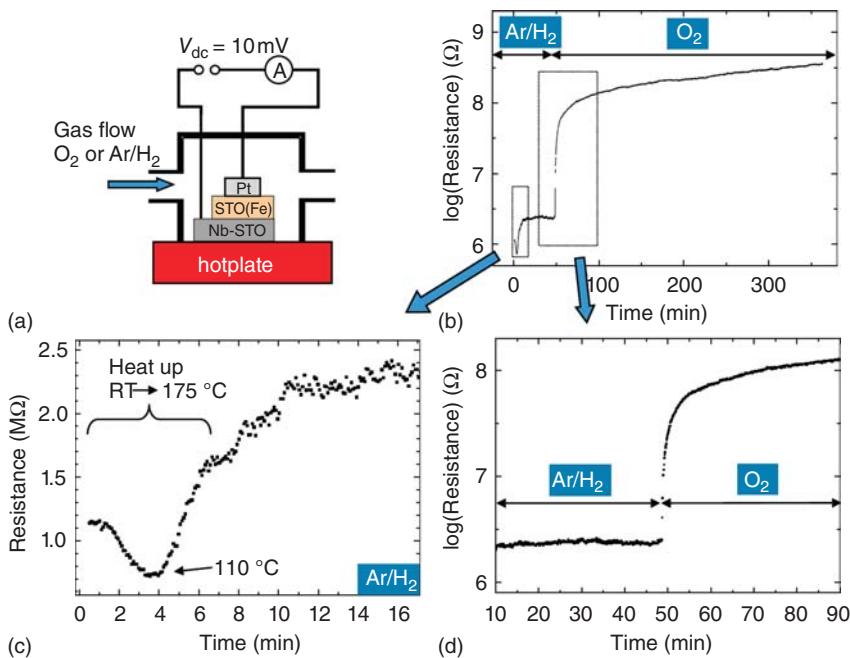


Figure 10.4 (a) Illustration of the setup employed in the investigation of the oxygen partial pressure dependence of the resistance of the formed Fe-doped $SrTiO_3$ -based switching cell. (b) The measured resistance change with respect to time ($R-t$) upon the change of atmosphere (first

Ar/H_2 and then O_2 at 50 min). The initial and transitional behaviors upon the atmosphere change are enlarged in (c) and (d), respectively. (Reprinted with permission from [14]. © 2009, American Institute of Physics.)

size is, the larger the number of grain boundaries underneath the top electrode. Thus, by means of grain size control, the correlation between grain size and forming behavior can be identified. An alternative way for this identification is to vary the electrode size and, thus, the grain boundary volume underneath. However, when it comes to extremely small electrode size (< grain size), the grain size effect can be ruled out, which was shown in HfO_x -based switching cells confined in $10 \times 10\text{ nm}^2$ crossbar structure [17]. In this extreme case, the thickness of the switching layer is a forming-voltage-determining factor as the voltage scales with the thickness as shown in Figure 10.6. This case is also of a main concern from technical perspectives, which is addressed in Section 10.3.

10.1.4

Forming in Bipolar and Unipolar Switching

Forming is required for the activation of not only bipolar switching but also unipolar switching that represents the TCM effect [1, 3]. An overview of a

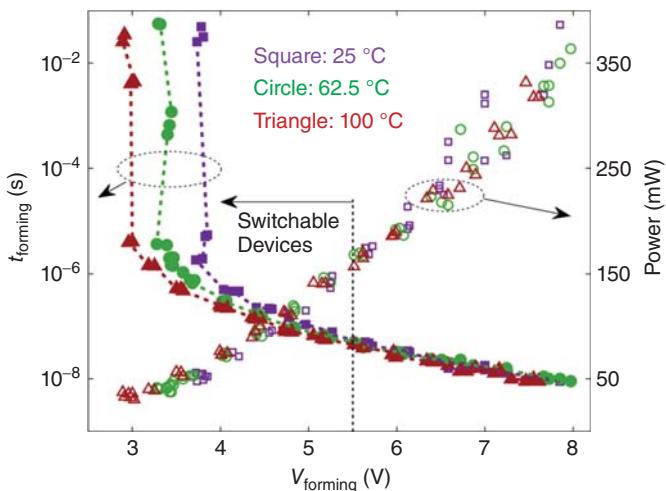


Figure 10.5 Voltage-pulse-induced forming in Pt/TiO₂/Pt switching cells at three different temperatures (25, 62.5, and 100 °C). For each temperature, the relationship between forming time and voltage pulse height is plotted. The power was evaluated from the measured voltage responses to the applied

voltage pulses. The short-dashed vertical line at approximately 5.5 V demarcates the dielectric breakdown (right side) and bipolar-switching-stable (left side) regions. (Reprinted with permission from [15]. © 2013, American Institute of Physics.)

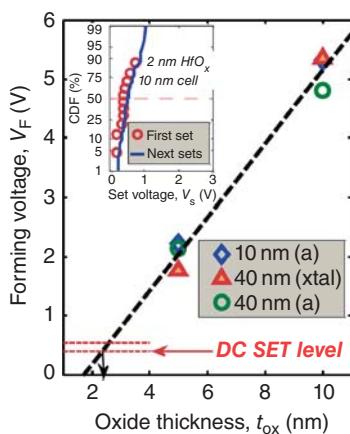


Figure 10.6 Forming voltage scaling with the thickness of the HfO_x films in the 10 × 10 and 40 × 40 nm² sized TiN/HfO_x/Hf/TiN switching cells. The inset shows a comparison between the forming and the subsequent set switching voltages for the 10 × 10 nm² sized cell employing a 2 nm thin HfO_x film. (Reprinted with permission from [17]. © 2011, IEEE.)

number of publications on unipolar switching in various TMOs gives the idea that the forming procedure for the TCM effect barely differs from that for the VCM, and thus, the aforementioned forming methods can be employed. There are several differences in the detailed forming conditions, however, particularly in the compliance current in the case of voltage-application-driven forming. The evidence can be found in the TiO₂ case, where the VCM and TCM effects coexist [18]. Either memory effect could be activated depending on the compliance

current; low (about <0.1 mA) and high compliance currents (about >1 mA), respectively, led the Pt/TiO₂/Pt switching cell to the VCM and TCM modes [18]. Interestingly, the switching cell that was once led to the VCM mode could undergo a transition to the TCM mode by allowing higher currents during the VCM operation. The previous VCM mode in TiO₂ could be recovered from the TCM mode from the reset state of TCM [19]. The VCM mode operation could be achieved from the conducting filament (CF) ruptured region of the reset state TCM, where the conductivity of that region could be modulated rather mildly by the applied voltage (or current) without completely recovering the ruptured filament. Given these features of TCM activation, the current and the consequent power dissipation during either forming or switching operation may be a switching-mode-determining factor.

10.1.5

Phenomenological Understanding of Forming

Phenomenologically, the bipolar switching behavior in the VCM case may be understood to arise from the asymmetry of the VCM cell. Here, “asymmetry” appears to be abstract as a mechanism for bipolar switching and has yet to be completely clarified. Despite this ambiguity, phenomenological understanding based on experiment results helps in achieving a more detailed understanding. Resistive-switching cells are most likely to include inherent asymmetries in their as-fabricated states (even before forming) due to asymmetric electrodes and asymmetric interfaces even when using symmetric electrodes due to the fabrication procedure. Apart from these inherent asymmetries, forming is likely to give rise to asymmetry of the cell, which appears to be the most significant factor determining the consequent bipolar switching. A significant evidence is the forming-polarity-dependent bipolar switching polarity (i.e., set and reset occur in positive and negative bias voltage, respectively, or vice versa) observed in the symmetric TiO₂ switching cells Pt/TiO₂/Pt [9], as shown in Figure 10.2.

10.2

Forming Mechanisms

The history of the research on forming in several MIM capacitors dates back to the 1960s. As the resistive random-access memory (RRAM) concept was brought up at the beginning of this millennium [20, 21], the focus of the early study on forming was on fundamental understanding rather than memory application. A review of the old publications on forming revealed that the negative differential resistance (NDR) effect was an important keyword in such research and that such effect can be classified as an S- or N-shaped NDR effect depending on the *I-V* curves exhibiting such effect. Forming exhibits the S-shaped NDR effect, as can be seen in Figure 10.1d. Many important concepts accounting for forming, which have been employed often, had already been introduced in the 1960s and 1970s

[22–28]. A lack of microstructural analysis techniques in the period hindered the justification of such concepts for the correlation between the microstructural change and the electrical-property change. Thus, it is important to overview the early forming mechanisms or the S-shaped NDR effect.

10.2.1

Early Suggested Forming Mechanisms

The several early mechanisms can perhaps be classified into two categories: purely electronic mechanisms and electrochemical mechanisms. Hickmott suggested a forming mechanism belonging to the former case [26, 27]. The mechanism explains forming as an electronic process leading to an impurity energy band in the forbidden gap, which serves as a low-energy electronic pathway. As the impurity band is placed in the vicinity of the Fermi level of the cathode, electrons can be injected into the impurity band without the thermal activation of the electrons in the cathode. Hickmott regarded the impact ionization of immobile impurities as the origin of the impurity band. Simmons and Verderber had a basically similar idea, with a difference only in the origin of the impurity band [28]. In their model, metal ion injection from the electrode was regarded as the origin of the impurity that forms the impurity band in the forbidden gap.

The electrochemical model suggested by Greene *et al.* is different from the aforementioned electronic mechanism [24]. In their model, it was presumed that the insulating matrix “MX” is phase-separated electrochemically, with the oxidation of anion $X^- (X^- \rightarrow X + e^-)$ and the reduction of cation $M^+ (M^+ + e^- \rightarrow M)$ at the anode and cathode, respectively. Particularly, as a result of the anion oxidation, gas evolution was predicted, but the model assumes the introduction of anion vacancies at the cathode, which was inconsistent with the current understanding of the point defect theory; anion vacancy V_X^- formation appears to take place at the anode through the following reaction in the Kröger–Vink notation [29]: $X_X^{\times} \rightarrow V_X^- + e' + 1/2X_{2(g)}$. Electron hopping through ordered $M - V_X^- - M -$ chains was estimated as the dominant conduction mechanism in the formed state.

Dearnaley *et al.* first suggested the filament mechanism where CFs are introduced in the insulating matrix during forming, which consequently results in inhomogeneous conductivity in the switching material [23]. This model indeed paved the way for the current understanding of the forming process. In this model, CFs are assumed to grow from several points at the insulator/anode interface, where the local electric fields are concentrated [30]. It was pointed out that electrons flow through the insulating layer by hopping along one-dimensional chains.

10.2.2

Conducting Filament Formation

One of the most remarkable features caused by forming is CF formation. An obvious signature of CF formation after forming can be found in the switching cell’s

resistance, which hardly scales with the cell's electrode size (or area). In fact, most switching materials are known to exhibit this signature to a certain degree after forming. It should be noted, however, that CF formation does not necessarily lead to the electrode-size-independent resistance of the switching cell because it concerns the size and distribution of the CFs in the insulating matrix. The uniform distribution of CFs with a very small cross section over the large electrode area appears to enable the cell to represent electrode-size-dependent resistance. Thus, the electrode-size-independent resistance of the switching cell indicates the formation of CFs whose distribution is nonuniform. Unless otherwise stated, the term *CF formation* basically implies the nonuniform distribution of CFs, in this chapter. Another complication regarding the electrode-size-dependent resistances can be induced when the CF size becomes comparable to the memory cell size, which may be the case when the memory cell dimension becomes $< \sim 10$ nm. This aspect is less known, however, and thus, in this chapter, the CF size (cross section) is always assumed to be much smaller than the electrode size.

Indirectly, the evolution of CFs in switching materials can be identified by monitoring the current with respect to time at a forming bias, as shown in Figure 10.7 for TiO_2 [10]. The initial resistance of the square-shaped switching cells (five different electrode sizes; square of the line width) appeared to scale with the electrode size, implying a homogeneous current flow over the electrode area. Increasing the applied current, however, triggered an inhomogeneous current flow. For instance, the current at the electroforming moment (i.e., the current) measured prior to the abrupt voltage drop (I_f) no longer scales with the electrode size. Instead, I_f is proportional to the line width, perhaps implying enhanced current conduction along the perimeter of the cell due to the fringing effect. Immediately after the abrupt voltage drop, the formed state was achieved, which shows resistance scaling with neither the electrode size nor the line width. Thus, it is noticed that the distribution of the CFs resulting from the forming is randomly inhomogeneous rather than uniform.

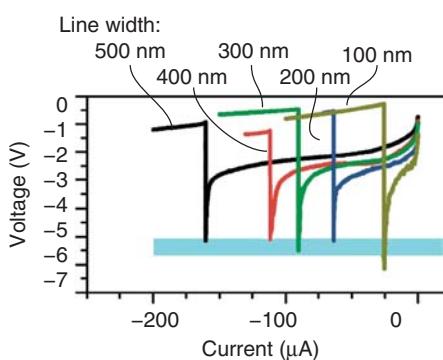


Figure 10.7 Current-sweep-driven forming for vertical asymmetric $\text{Pt}/\text{TiO}_2/\text{Ti}/\text{Pt}$ cells (from the bottom). The square-shaped cells have five different pad sizes:

100×100 , 200×200 , 300×300 , 400×400 , and $500 \times 500 \text{ nm}^2$. (Reprinted with permission from [10]. © 2010, American Institute of Physics.)

More details on the states before the filament formation are provided by Sharma *et al.* [31]. Beyond a critical voltage, the initially homogeneous electronic current flow encounters filamentation, which is accompanied by a significant increase in the local temperature and an NDR regime. While the electronic stages of the forming process are still reversible (volatile), the local high temperature in combination with the electric field triggers the motion ions, which then lead to irreversible changes, often accompanied by a filamentary phase transformation.

10.2.3

Redox Reactions and Ion or Ionic Defect Migration during Forming

Oxygen gas formation leading to structural deformation is a representative feature of forming, which has been observed in various switching oxides, such as SiO_x [23], TiO_2 [9, 13, 32, 33], SrTiO_3 [34], and Fe-doped SrTiO_3 [35]. Concerning common elements that can form gas at room temperature in these oxides, the observed evolved gaseous species can be estimated to be oxygen gas. If this is the case, oxygen gas is supposed to be formed by the following electrolytic reaction: $2\text{O}^{2-} \rightarrow 4e^- + \text{O}_{2(\text{g})}$. This reaction means the oxidation of the oxygen ions, thus taking place at the anode of the switching cell. That is, the location of oxygen gas formation depends on the forming bias polarity; positive (negative) forming bias application to the top electrode of a vertical switching cell leads to gas formation at the top (bottom) electrode. Bubble formation was found in an as-formed $\text{Si}/\text{SiO}_x/\text{Au}$ switching cell, at its anode (see Figure 10.8) [23]. In fact, this observation was a significant evidence for the aforementioned forming models suggested earlier [22–24]. A similar bubble evolution was also observed in the $\text{Pt}/\text{TiO}_2/\text{Pt}$ cells, where the structural deformation was attributed to

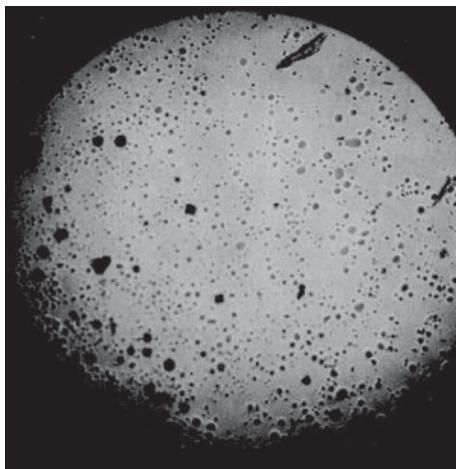


Figure 10.8 Optical microscope image of an as-formed $\text{Si}/\text{SiO}_x/\text{Au}$ switching cell. A number of bubbles are shown and placed at the anode. (Reprinted with permission from [23]. © 1970, North-Holland Publishing.)

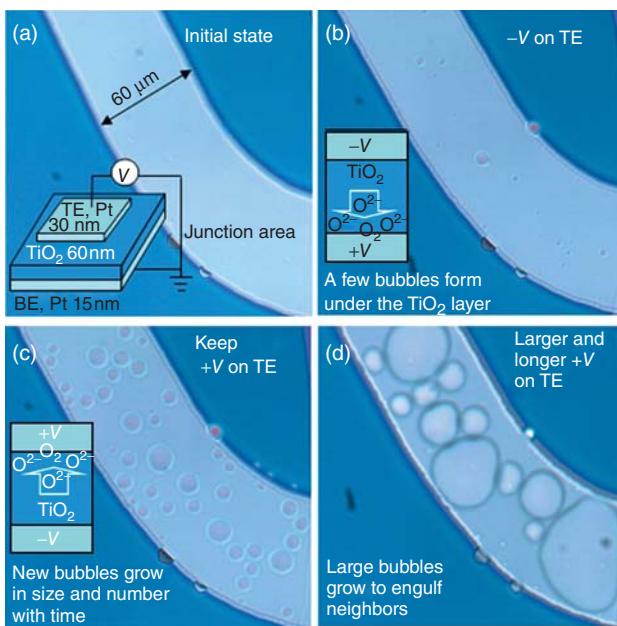


Figure 10.9 Optical microscope images of (a) an as-fabricated Pt/TiO₂/Pt switching cell (its schematic is shown in the inset), the same cell with (b) a negative voltage, (c) a positive voltage, and (d) a higher and longer voltage applied to the top electrode.

The schematic mechanisms for the bubble formation under the negative and positive voltages are illustrated in the insets of (b) and (c), respectively. (Reprinted with permission from [13]. © 2009, IOP Publishing.)

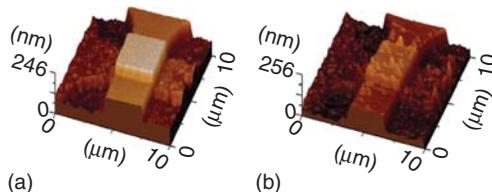


Figure 10.10 Atomic-force-microscope topographic images of (a) a pristine Pt/TiO₂/Pt switching cell and (b) the same cell after a forming process. (Reprinted with permission from [32]. © 2013, Wiley-VCH.)

oxygen gas formation depending on the forming bias polarity [9, 13]. Figures 10.9 and 10.10 show the optical-microscope and atomic-force-microscope images, respectively, of the TiO₂-based switching cells, including oxygen bubbles, as a result of the forming processes. It should be noted that the oxygen gas evolution reaction introduces four electrons per oxygen gas molecule, and they will most likely stay in the switching oxide, so that this reaction corresponds to a donor-doping process. This issue will shortly be addressed in detail.

It is quite probable that some oxygen gas evolved through the anodic reaction is incorporated into the anode. In particular, if noble metals such as Pt and Au are used as electrode materials, oxygen gas is likely to stay at the grain boundaries by forming a strong and dense chemisorption bond with noble metal atoms instead of forming metal oxide phases [36]. The forming-induced incorporation of oxygen into the Pt top electrode of a Pt/TiO₂/Pt cell was identified via time-of-flight secondary ion mass spectrometry analysis [9].

Given that oxygen gas evolution is a remarkable feature of forming, forming is likely dependent on the atmosphere. With regard to the reaction $2\text{O}^{2-} \rightarrow 4e^- + \text{O}_{2(\text{g})}$, this depends on oxygen partial pressure p_{O_2} ; that is, the reverse of this reaction simultaneously occurs with the oxygen gas evolution reaction, where the oxygen partial pressure determines the reverse reaction rate. Therefore, in several early forming mechanisms, the switching cells were regarded as open systems allowing oxygen gas exchange with the atmosphere that serves as an oxygen gas reservoir [22–24]. Greene *et al.* considered the incorporation of the oxygen from atmosphere into the switching material as a result of the cathodic reaction $1/2\text{O}_{2(\text{g})} + 2e^- \rightarrow \text{O}^{2-}$, which is the reverse of the oxygen gas evolution reaction [24]. Moreover, these pioneers predicted the electrode dependence of forming such that when an oxygen-reactive electrode is in use, the evolved oxygen gas reacts with the electrode, consequently forming second oxide phases that hinder forming [23, 24].

10.2.4

Point Defect Introduction

The aforementioned evolution of oxygen gas during forming introduces oxygen vacancies (V_{O}^{\cdot} in the Kröger–Vink notation [29]) in the matrix as the oxygen atoms are most likely from the lattice. The oxygen vacancy formation reaction is expressed in the Kröger–Vink notation as



where $\text{O}_{\text{O}}^{\times}$ denotes an oxygen ion on an oxygen site. Oxygen vacancies are assumed to have a charge of +2. This reaction actually corresponds to the oxygen gas formation reaction $2\text{O}^{2-} \rightarrow 4e^- + \text{O}_{2(\text{g})}$, in the regular nomenclature. It is noticed in Eq. (10.1) that the reaction introduces two electrons that are delocalized (i.e., free electrons) in the matrix. This reaction is referred to as *self-doping* as oxygen vacancies effectively serve as donors. Therefore, the larger the number of oxygen vacancies is, the higher is the conductivity. Experimental evidence for the oxygen vacancy introduction by forming has been found in Cr-doped SrZrO₃ by means of X-ray absorption near-edge spectroscopy (XANES) [37]. The XANES analysis of the gap between the anode and the cathode of a lateral Cr-doped SrZrO₃-based switching cell revealed that oxygen vacancies formed a channel connecting the two lateral electrodes. In addition, more oxygen vacancies were observed near the anode than near the cathode, and thus, the anode is estimated to be an oxygen

vacancy source. This finding is consistent with the aforementioned forming theory [37].

Some TMOs are known to inherently (i.e., even before forming) include a number of oxygen vacancies, which are classified as “hypostoichiometric” TMOs [38]. Hypostoichiometric TMOs are often expressed as $\text{MO}_{x-\delta}$, $\delta > 0$, where δ denotes a deviation from the stoichiometric MO_x . Thus, hypostoichiometry leads to oxygen deficiency. Of course, oxygen vacancy formation is not the only cause of hypostoichiometry; the cation interstitial $\text{M}_i^{\cdot\cdot}$ can also lead to hypostoichiometry via the following equation in the Kröger–Vink notation:



where M_M^X denotes a cation on a cation site. In a hypostoichiometric TMO, one of the two defect types (i.e., oxygen vacancy and cation interstitial) is dominant under the given thermodynamic conditions because it involves the energy required for each defect formation under particular thermodynamic conditions. For instance, the dominant defect type in TiO_2 depends on the temperature; the oxygen vacancy and Ti interstitial are low- and high-temperature-stable defect types, respectively [39]. Thus, forming in TiO_2 at a low temperature probably introduces the same type of point defect electrochemically as its inherent (i.e., thermodynamically assigned) point defect type. Regardless of the dominant defect type, however, hypostoichiometry results in additional free electrons in the matrix, as can be seen on the right-hand side of Eqs. (10.1) and (10.2). As such, hypostoichiometric TMOs are known as *n-type semiconductors*.

The other type of nonstoichiometry is hyperstoichiometry, which leads to $\text{MO}_{x+\delta}$, $\delta > 0$ (i.e., cation deficiency). Hyperstoichiometry arises from the formation of either oxygen interstitials (O_i'' in the Kröger–Vink notation) or cation vacancies (V_M'' in the Kröger–Vink notation). The formation reactions of an oxygen interstitial and a cation vacancy are expressed as



and



respectively. As can be seen on the right-hand side of Eqs. (10.3) and (10.4), either reaction increases the population of holes in the matrix. As such, hyperstoichiometric TMOs are p-type semiconductors. Table 10.1 lists the major electronic-carrier types of several binary TMOs of various valence states [38, 40–47]. For hyperstoichiometric TMOs, forming introduces a different type of point defect and electronic carrier from the thermodynamically induced types. Given this difference, the forming kinetics in hypo- and hyperstoichiometric TMOs are estimated to be different. The forming kinetics in each nonstoichiometric case will be addressed in the next section.

Table 10.1 Major electronic-carrier types of several TMOs of various valence states.

	Ti–O	V–O	Cr–O	Mn–O	Fe–O	Co–O	Ni–O
+1 (M_2O)							
+2 (MO)				p [40]	p [40]	p [40]	p [40]
+2.66 (M_3O_4)					p [38]		
+3 (M_2O_3)			n or p [41, 42]			n [42]	
+4 (MO_2)	n [42]	n [43]			n [42]		
+5 (M_2O_5)		n [42]					
	Cu–O		Nb–O		Ta–O		W–O
+1 (M_2O)		p [42]					
+2 (MO)		n or p [42]					
+2.66 (M_3O_4)							
+3 (M_2O_3)							
+4 (MO_2)			n [44]				
+5 (M_2O_5)			n [45]		n or p [46]		
+6 (MO_3)						n [47]	

The blanks in the table mean either that the corresponding TMOs are metallic or that they do not exist as stable phases.

Reprinted with permission from [1]. Copyright 2012, IOP Publishing.

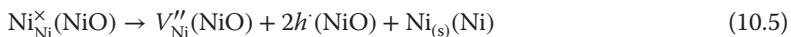
10.2.5

Point Defect Dynamics during the Forming Process

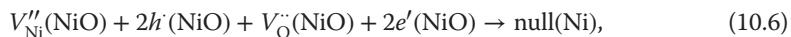
The aforementioned point defects move in the matrix due to two major driving forces, electrostatic and chemical potential gradients, which lead to drift and diffusion, respectively. Especially, the drift of point defects is an important mechanism for defect migration under the internal electric field resulting from the forming bias. A switching cell while forming is considered to consist of three serial parts: a defect (oxygen vacancy) source, a spacer where the defects migrate, and a defect sink (if it exists) [9, 34, 48]. On the assumption that an oxygen vacancy is the only defect type introduced during forming, the anode serves as the source. They drift through the TMO layer (i.e., the spacer) toward the cathode and reach the opposite electrode (cathode). In case no sink exists in the cell, oxygen vacancies are piled up at the cathode, resulting in the evolution of chemical potential [49]. This kind of contact is referred to as *blocking contact*. The direction of diffusion arising from the chemical potential evolution is opposite that of the drift, so that the diffusion contributes to depolarization [49]. In case a sink exists, oxygen vacancies are able to escape from the TMO layer and are thus accumulated at the anode, but not as much as in the blocking contact case. This kind of contact is referred to as *nonblocking* or *partially blocking contact* depending on how fast the oxygen vacancies are able to move out of the TMO layer through the contact between the electrode and the TMO layer.

The escape of oxygen vacancies from the TMO layer through the contact means incorporation of oxygen ions into the TMO from the electrode. Thus, the available oxygen ions in the electrode are required for either nonblocking or partially blocking contact. If oxide electrodes are used in the switching cell, they may serve as oxygen ion sources enabling the oxygen vacancies in the TMO layer to move out of the layer. Otherwise, it may not be possible to form nonblocking or partially blocking contact, particularly in the case of using noble metal electrodes (e.g., Au and Pt). At this point, it is worthwhile to look into the possible forming mechanisms in terms of point defect dynamics. Figure 10.11 illustrates the forming mechanisms in due course for NiO and TiO₂, the prototypical switching TMOs. NiO and TiO₂ are hyper- and hypostoichiometric TMOs, respectively. As mentioned earlier, TiO₂ at room temperature is regarded to include a number of oxygen vacancies, and thus free electrons, leading to an n-type semiconductor. The dominant defect type in NiO is thought to be the Ni vacancy (V''_{Ni} in the Kröger–Vink notation) rather than the oxygen interstitial [50, 51]. Note that in this model, noble metal electrodes are assumed to be used, forming blocking contact for oxygen and Ni vacancies.

An examination of the NiO case is in order. In a pristine NiO-based switching cell, there are a number of inherent Ni vacancies that have an effective charge of –2 (see Figure 10.11a). While a forming bias is applied to the cell (in Figure 10.11, the top electrode is the anode), oxygen vacancies are introduced at the top electrode (i.e., the oxygen vacancy source) and drift toward the bottom electrode (cathode). At the same time, Ni vacancies drift toward the top electrode as they have the opposite charge, and they eventually polarize the NiO layer. Insomuch as neither oxide electrodes nor Ni electrodes are assumed to be used in the switching cell, the oxygen and Ni vacancies are likely to be piled up at each electrode (see Figure 10.11a). Although TMOs can include a number of point defects, there is a particular limit of defect concentration, in which the phase stability is satisfied [52]. Therefore, the higher the oxygen vacancy concentration at the bottom electrode, the higher is the instability of the NiO phase. With regard to the hyperstoichiometric nature of NiO, it is noted that thermodynamics does not allow oxygen-deficient NiO. Thus, phase separation into Ni and NiO may occur at the bottom electrode, as illustrated in Figure 10.11a. This phase separation can be described in the Kröger–Vink notation by the following sequential reactions:



and



where Ni and NiO within the parentheses are the locations where the given defects or electronic carriers are present. Equation (10.5) accounts for a Ni vacancy formation reaction by Ni phase evolution [$\text{Ni}_{(\text{s})}(\text{Ni})$]. The formed Ni vacancy can react with an oxygen vacancy, two holes, and two electrons in the vicinity of the bottom electrode, resulting in the disappearance of the lattice sites of the Ni and oxygen ions, expressed in Eq. (10.6).

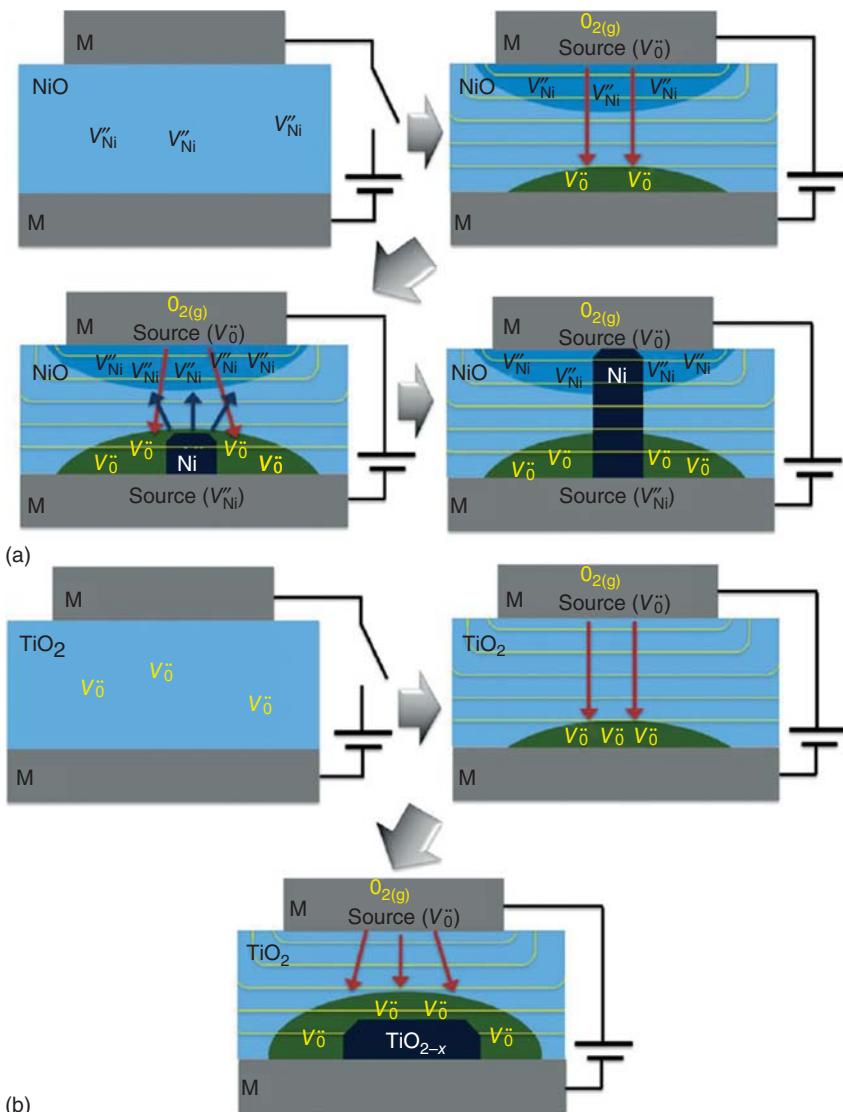


Figure 10.11 Schematics of the point-dynamics-triggering CF formation sequence for (a) hyperstoichiometric NiO and (b) hypostoichiometric TiO_2 . (Reprinted with permission from [1]. © 2012, IOP Publishing.)

The formed Ni phase at the cathode works as a moving cathode as well as a Ni vacancy source. Ni vacancies are perhaps generated at the front end of the Ni phase by Eq. (10.5), implying that the phase grows toward the anode. The growth of the Ni phase may terminate when it reaches the anode, completing the Ni CF formation.

There is another way of interpreting the formation of Ni CF during the forming process in NiO. When an oxygen vacancy is formed near the anode and stays there, the hyperstoichiometric property of NiO triggers Ni precipitation at locations near the anode. In this case, the Ni CF may start to form from the anode interface and grow toward the cathode interface. This is in contrast to the earlier discussions, where it was mentioned that the ionized oxygen vacancies drift toward the cathode interface region and their accumulation near the cathode induces the Ni CF initiation from the cathode, which will later grow toward the anode. It is believed that the dominant mechanism is determined by the drift speed of the oxygen vacancies in NiO; if it is faster than the speed of Ni precipitation, the latter will be the dominant mechanism, whereas if it is slower than that, the former will be the working mechanism. As the electric field for forming is usually very high in the MIM-type structure with an insulating-layer thickness of several nanometers to several tens nanometers, the drift speed usually must be quite high. When the oxygen vacancies are not ionized (i.e., the oxygen vacancies trap the electrons), however, the driving force for the vacancy migration is only the concentration gradient, which may not be sufficiently high to drive them fast. The degree of ionization of the oxygen vacancies in TMO depends on many factors, especially in the thin-film system, making its determination generally complicated. The energy level in band gap certainly influences it; it needs to be shallow to be easily ionized. When the film thickness is less than the depletion thickness of the insulator–metal junction, the Fermi level of the contacting metal has a profound effect on the ionization of the oxygen vacancies. NiO is somewhat peculiar in its energy band structure compared with other TMOs, typically titanates, in that its electron affinity is quite small (~2 eV, while that of most titanates is 3–4 eV) and its band gap is rather large (~4 eV), which is closely related to its hyperstoichiometric property. This makes the NiO–metal junction properties vary considerably depending on the types of contacting metal, whereas the titanate(TiO_2)–metal junction almost always has a lower Schottky barrier for electrons compared with that for holes. When the oxygen vacancies are not efficiently ionized for any reason, the Ni CF may initiate from the anode to the cathode, whereas growth may occur in the other direction when the oxygen vacancies are efficiently ionized. As there have not been many reports on this aspect of NiO, it is not easy to assess which is the more dominant case in this RS system. Kim *et al.*, however, reported works supporting the mechanism where the Ni CF was initiated at the anode and extended to the cathode [53–55]. By connecting two MIM (Pt/NiO/Pt) cells serially and switching them on and off, they found that the cell near the cathode was disconnected while the cell near the anode was still connected even after the reset operation in the serially connected configuration. This strongly suggests that the Ni CF was initiated from the anode and was thus stronger, causing the CF to remain intact even after the reset, while the Ni CF near the cathode was weaker because it was formed at the later stage of forming. It should be noted, however, that this may not be the case when the oxygen vacancies are well ionized for any reason.

The point defect dynamics and the consequent CF formation in TiO_2 appear to be less complicated than those in NiO because the inherent point defect type in TiO_2 is perhaps the same as a result of the oxygen gas formation reaction during forming. As shown in Figure 10.11b, the oxygen vacancies introduced at the anode interface drift toward the cathode due to the electrostatic potential gradient and are piled up at the cathode. Similar to the NiO case, the high concentration of oxygen vacancies at the cathode leads to the instability of the TiO_2 mother phase. As a result, some lower oxide phases (e.g., the Magnéli phases $\text{Ti}_n\text{O}_{2n-1}$, often $n=4, 5$) are evolved and work as CFs as they have much lower resistivities than the TiO_2 phase [56–59]. It should be noted that the lower oxide phase most likely serves as an oxygen vacancy sink as oxygen vacancies are annihilated on its surface. When the oxygen vacancies are accumulated on the (121) crystallographic plane of rutile-phase TiO_2 and the structure is collapsed along the direction normal to the vacancy-accumulated plane, the Magnéli phases are formed. Therefore, the Magnéli phases are the sinks of the oxygen vacancies. The growth rate of the oxide phase(s) toward the anode becomes lower as the anode keeps producing oxygen vacancies, which drift toward the lower oxide phase and disappear on the surface, allowing the phase to enlarge its volume in the mother matrix. Forming may be completed when the low oxide phase reaches the anode, forming a complete CF(s). Experimental evidences for such forming behavior could be directly found through the microscopic identification of the Magnéli CF, which is discussed in the next section. Meanwhile, the set–reset experiments with two serially connected MIM ($\text{Pt}/\text{TiO}_2/\text{Pt}$) cells corroborate the suggested growth direction of Magnéli-based CFs in TiO_2 during forming [54, 60]. When the serially connected memory cell was reset, the cell near the anode was almost always disconnected while that near the cathode remained connected, which is in stark contrast to the NiO case discussed. This suggests that the Magnéli-based CF was initiated from the cathode and was thus stronger near the cathode interface and later on grew toward the anode and was thus weaker near the anode interface.

Given these two examples, where the CF formation was triggered by the instability of the mother phases, it is considered that phase transition into a lower oxide or even into a metal phase with much lower resistivities than the mother oxide is an essential feature of CF formation in the forming process. In the next section, microscopic evidence for the phase transition taking place during forming is presented.

10.2.6

Microscopic Evidence for CF Formation during Forming

For many decades, it was “guessed” that forming induces CFs in the insulating matrix as the phenomenon has been supported by many indirect experimental data, such as the aforementioned forming-induced electrode-size independence of resistance, but there was no direct evidence of CF formation. In fact, it took almost 40 years to directly see some CFs embedded in the matrix after the first prediction by Dearnaley *et al.* In 2010, by means of transmission electron

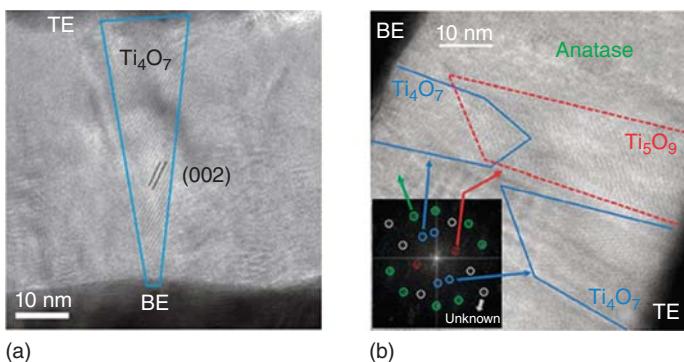


Figure 10.12 High-resolution TEM images of the CFs of (a) a single Magnéli phase ($n=4$; Ti_4O_7) and (b) two different Magnéli phases ($n=4, 5$; Ti_4O_7 and Ti_5O_9 , respectively).

(Reprinted with permission from [56, 61]. © 2010, Nature Publishing Group, and 2011, American Institute of Physics, respectively.)

microscopy (TEM), CFs in a Pt/TiO₂/Pt cell were directly observed, and it was revealed that they are made of the Magnéli phases ($\text{Ti}_n\text{O}_{2n-1}$, often $n=4, 5$), as shown in Figure 10.12 [56]. It was clearly determined that the Magnéli-phase-based CFs in TiO₂ were initiated from the cathode interface and grew toward the anode interface from the configurations of the connected CF and not the completed CFs. A similar observation was made in TiO₂ not only via TEM but also via scanning transmission X-ray microscopy [57]. This observation confirmed not only the formation of CFs by forming but also the phase transition involved in forming. This is a remarkable finding because it showed that CFs should be explained in terms of phase transition rather than point defects. Of course, as explained in the previous section, point defects may trigger CF formation, but they cannot be the major ingredients of CFs. The thermodynamics of phase transition and the consequent resistivity change in TiO₂ are well overviewed in a review paper by Szot *et al.* [59]. Forming-induced phase transition in several binary TMOs (e.g., TiO₂, V₂O₅, Fe₂O₃, and Nb₂O₅) was theoretically predicted by Chudnovskii *et al.*, indicating that forming leads to phase transition from the most resistive highest TMOs to the less resistive lower TMOs [62].

Forming-induced phase transitions have also been observed in different switching TMOs, such as Ta₂O₅ [63] and CuO [64]. Thus, nanoscale local phase transition may be an essential feature of forming. Indeed, phase transition seems to be the only way of guaranteeing the long-term stability of CFs in a formed cell as the evolved phase is of metastability [1]. It should be noted that point defects (or even their slight aggregation) cannot generally meet this stability criterion because they are supposed to undergo fast relaxation as soon as the external force disappears, and the relaxation is determined by the diffusion coefficient of the defects [1, 49]. Therefore, this relaxation characteristic of point defects is barely consistent with the experimental observations of CFs with long-term stability.

10.3

Technical Issues Related to Forming

Several features of forming in some prototypical switching TMOs have been overviewed. Although the forming mechanisms are not yet very clear, several general characteristics of forming could be deduced from these case studies. The forming phenomenon seems to be an interesting research topic because it has yet to be fully understood due to its complicated and stochastic nature. When it comes to memory application, however, forming can lead to a number of technical issues, such as current overshoot and nonuniform forming bias distribution in passive RRAMs, such as crossbar-array-based RRAMs. These issues are actually quite critical and challenging in terms of memory operation. The ultimate solution to these problems is most likely to adopt switching materials and processes that can avoid forming processes in resistive-switching operation (i.e., to realize forming-free RRAMs). The attempts to realize forming-free RRAMs is addressed in the last section.

10.3.1

Problems of Current Overshoot Forming

Current overshoot, occurring during voltage-application-driven forming when a compliance current is adopted, is believed to critically affect the subsequent switching behavior [6–8]. In particular, a high overshoot current has often been found to lead the switching cell to a nonswitchable low-resistance state or even to an irreversible dielectric breakdown. Several parameters of HfO₂-based resistive switches have been examined in relation to the current overshoot, and several experimental rules of thumb have been found [8]. First, the higher the initial resistance is, the larger the current overshoot. Gilmer *et al.* measured the current overshoot for HfO_x-based switching cells to which various initial resistances were assigned by varying the stoichiometry and/or the thickness of HfO_x and inserting a highly resistive Al₂O₃ layer(s) [8]. The stoichiometric and thick HfO₂ had an initial high resistance, and its current overshoot was larger than that those of the nonstoichiometric and thinner films. An increase in current overshoot results in an increase in reset current and high power consumption. In extreme cases, the formed resistance state was stuck to a very-low-resistance state, and no further reset operation worked out. The insertion of a highly resistive Al₂O₃ layer(s) caused an even higher initial resistance; accordingly, it resulted in a further increase in current overshoot.

Second, the higher the forming voltage is, the larger is the overshoot. This second rule of thumb is strongly related to the first one in the sense that higher initial resistances consequently lead to higher forming voltages in the cell. Therefore, it is not easy to decide whether the initial resistance or the high forming voltage caused the current overshoot from the aforementioned experimental data. These features of current overshoot can provide an important guideline for switching-cell design. With regard to the fact, however, that the high-resistance state of a switching cell

is in general less resistive than the as-fabricated state, attempts to decrease the cell's initial resistance may lead to a leaky high-resistance state, consequently narrowing down the resistance gap between the high- and low-resistance states. Thus, in switching-cell design, many different aspects of switching operation should be taken into account.

10.3.2

Nonuniform Forming Voltage Distribution

In crossbar-array-based passive RRAMs, the line resistance becomes a critical issue with increasing integration density. Scaling down the feature size inevitably brings up the issue of line resistance increase, which leads to a high voltage drop along the electrode line. When it comes to passive RRAMs employing forming-necessary switching cells, the situation is even worse. Unless each cell in the array is individually formed (which is impossible in reality, given the enormous number of switching cells in the array), each cell is supposed to be formed by applying a forming bias to the word and bit lines. If this is the case, a voltage drop along the word and bit lines arising from the line resistance hardly allows the same forming bias to be applied to each cell sharing the lines. This will most likely result in nonuniform switching operation over the cells in the array, which should be avoided. As mentioned earlier, forming under the same conditions hardly guarantees uniform switching operation due to its stochastic nature. Therefore, the complication induced by these uncertainties critically hinders the realization of passive RRAMs utilizing forming-necessary switching cells. A recent report by Yoon *et al.* suggested that limiting the area where the electric field is applied to a smaller portion of the electrode could considerably improve the forming and the subsequent switching uniformity [32].

10.3.3

Forming-Free Resistive Switching

The term *forming-free* is a phenomenological definition; the initial (i.e., set-like) switching behavior is more or less the same as the set switching behavior in the normal switching operation, in terms of a threshold voltage for the resistance decrease and the resistance before the forming process. When the forming voltage and the initial resistance are not particularly high, the aforementioned overshoot problem and the nonuniform forming bias problem may be solved. In addition, by avoiding current overshoot, low reset current and the consequent low operational power can be realized. There have been several attempts to achieve forming-free switching cells by involving either nonstoichiometric TMO phases or suboxide phases in highly insulating TMO mother matrices such as HfO_2 [8, 65], NiO [6, 66], WO_x [67], Cu_xO [68], and TiO_2 [13, 69]. The incorporation of these phases reduces the initial resistance, and thus, no high voltage is required for the first set-like switching.

Alternatively, forming voltage control by thickness of the switching materials can be a solution to the aforementioned problems. As mentioned in Section 10.1.3, the forming voltage of the $10 \times 10 \text{ nm}^2$ sized HfO_x -based cells was found to scale well with the thickness since the grain boundary effect on the forming behavior in such small cells was ruled out [17]. Voltages below 1 V successfully formed a 2 nm-thick HfO_x film, which were comparable with the subsequent set voltages as shown in Figure 10.6. Thus, forming-free switching was effectively realized in this cell.

Another type of forming-free switching cell is the resistive tunnel junction, in which the electron tunneling current undergoes changes upon changes in the electron supply function at the junction [70]. The electron supply function perhaps changes upon the redox reactions of the reactive electrode, and thus, nonvolatile resistive switching can be realized without forming.

10.4

Summary and Outlook

In this chapter, forming in unipolar (or nonpolar) and bipolar switching in mostly TMOs was addressed from the phenomenological and defect chemistry points of view. In most cases, forming is a local phase transition in both the vertical and lateral dimensions to form the CFs, and the repeated rupture and rejuvenation of the filaments during the subsequent reset and set operations are responsible for the observed resistive switching. The CFs in hypostoichiometric oxides are generally percolated aggregations of oxygen vacancies in the resistive-switching (RS) systems where distinctive conducting second phases are not present, such as HfO_2 and Ta_2O_5 , nor conducting second phases such as the Magnéli phases in TiO_2 and WO_3 , whereas the CFs in hyperstoichiometric oxides are percolated metallic paths such as Ni in NiO . Therefore, the kinetics of the forming process can be regarded as a nanoscale local phase transition. The sparse density and highly localized nature of the CFs in most RS systems hinder the accurate understanding of the kinetic process during CF formation, which is quite different from the situation for other functional memory elements, such as ferroelectric domain switching. The recent advancement in the understanding of the evolution of the switching current during forming or set switching proved, however, that the nanoscale local phase transition could still be understood from the Johnson–Mehl–Avrami-type kinetic model, where one-dimensional nucleation and two-dimensional growth can successfully explain the evolution of CFs with time or voltage [71]. The direction of CF growth during forming is another source of confusion for the different types of resistive-switching oxides. While the growth direction for the CFs in n-type oxides such as TiO_2 was confirmed to be from the cathode to the anode via the electrical and structural means, the growth direction in p-type oxides such as NiO is ambiguous. Such directionality has important implications on the subsequent switching operations because not the whole CF but only a certain portion of the weakest CF is most probably

responsible for the repeated rupture and rejuvenation in both unipolar and bipolar switching. As CF formation is a phenomenon kinetically driven by the applied bias voltage, its thermodynamic stability is a potential problem. While higher power dissipation generally induces a stronger CF, making the stability issue less severe, high power consumption and possibly its accompanying uncontrollable factors adversely interfere with stable and uniform switching performances. Eventually, the forming-free operation mode is required for the real application of switching materials into RRAMs with high densities.

There remain many unknown factors for forming in switching materials. One of the critical problems is the relative size of the CF in comparison with the memory cell area. In HfO_2 -based RRAMs, quantum point contact behaviors have been reported, suggesting that the CF size could be scaled down to the atomic diameter [72]. With this dimension, however, the thermal stability of such tiny CFs is a serious concern. Even when the CF is much larger than the atomic dimension, there is no guarantee that the average size is invariant with varying electrode size. The location of the CFs within a given electrode area could also be severely varied. In several works using conductive atomic force microscopy, the CFs are usually found at the edge region of a relatively large electrode (at least $>100 \mu\text{m}^2$) probably due to the local field concentration effect. As can be seen from the relatively uniform distribution of protrusions, however, which are believed to be related with the oxygen gas evolution and thus with CF formation, as observed in Figure 10.10b, the smaller electrode area ($4 \times 4 \mu\text{m}^2$) can result in a different situation.

In any case, the highly distinctive and abrupt formation of CFs in oxide-based resistive-switching memories during the forming step is highly undesirable in terms of the electrical and structural reliabilities of the devices. Memory devices with forming-free or only slight-forming properties are required, which can be achieved through the careful control of the stoichiometry, film thickness, cell topography, and operation conditions.

Acknowledgments

CSH acknowledges the support given by the Global Research Laboratory Program (2012040157) through the National Research Foundation (NRF) of Korea. DSJ also acknowledges the Korea Institute of Science and Technology grant (grant no. 2Z04510).

References

1. Jeong, D.S., Thomas, R., Katiyar, R.S., Scott, J.F., Kohlstedt, H., Petraru, A., and Hwang, C.S. (2012) *Rep. Prog. Phys.*, **75**, 076502.
2. Waser, R. and Aono, M. (2007) *Nat. Mater.*, **6**, 833.
3. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21**, 2632.
4. Valov, I., Waser, R., Jameson, J.R., and Kozicki, M.N. (2011) *Nanotechnology*, **22**, 254003.
5. Lu, W., Jeong, D.S., Kozicki, M.N., and Waser, R. (2012) *MRS Bull.*, **37**, 124.
6. Goux, L., Lisoni, J.G., Xin Peng, W., Jurczak, M., and Wouters, D.J. (2009) *IEEE Trans. Electron Devices*, **56**, 2363.

7. Wan, H.J., Zhou, P., Ye, L., Lin, Y.Y., Tang, T.A., Wu, H.M., and Chi, M.H. (2010) *IEEE Electron Device Lett.*, **31**, 246.
8. Gilmer, D.C., Bersuker, G., Park, H.Y., Park, C., Butcher, B., Wang, W., Kirsch, P.D., and Jammy, R. (2011) Effects of RRAM stack configuration on forming voltage and current overshoot. Presented at *3rd IEEE International Memory Workshop (IMW)*, 2011, May 22–25, 2011.
9. Jeong, D.S., Schroeder, H., Breuer, U., and Waser, R. (2008) *J. Appl. Phys.*, **104**, 123716.
10. Nauenheim, C., Kügeler, C., Ruediger, A., and Waser, R. (2010) *Appl. Phys. Lett.*, **96**, 122902.
11. Jiang, H. and Xia, Q. (2013) *Nanoscale*, **5**, 3257.
12. Yan, X.B., Xia, Y.D., Xu, H.N., Gao, X., Li, H.T., Li, R., Yin, J., and Liu, Z.G. (2010) *Appl. Phys. Lett.*, **97**, 112101.
13. Yang, J.J., Feng, M., Matthew, D.P., Douglas, A.A.O., Duncan, R.S., Chun Ning, L., and Williams, R.S. (2009) *Nanotechnology*, **20**, 215201.
14. Menke, T., Dittmann, R., Meuffels, P., Szot, K., and Waser, R. (2009) *J. Appl. Phys.*, **106**, 114507.
15. Nomar, M., Sharma, A.A., Lu, Y.M., Skowronski, M., Salvador, P.A., and Bain, J.A. (2013) *Appl. Phys. Lett.*, **102**, 023507.
16. Bersuker, G., Gilmer, D.C., Veksler, D., Yum, J., Park, H., Lian, S., Vandelli, L., Padovani, A., Larcher, L., McKenna, K., Shluger, A., Iglesias, V., Porti, M., Nafria, M., Taylor, W., Kirsch, P.D., and Jammy, R. (2010) Metal oxide RRAM switching mechanism based on conductive filament microscopic properties. Presented at *IEEE International Electron Devices Meeting (IEDM)*, 2010, December 6–8, 2010.
17. Govoreanu, B., Kar, G.S., Chen, Y., Paraschiv, V., Kubicek, S., Fantini, A., Radu, I.P., Goux, L., Clima, S., Degraeve, R., Jossart, N., Richard, O., Vandeweyer, T., Seo, K., Hendrickx, P., Pourtois, G., Bender, H., Altimime, L., Wouters, D.J., Kittl, J.A., and Jurczak, M. (2011) 10x10nm² Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. Presented at *IEEE International Electron Devices Meeting (IEDM)*, 2011, December 5–7, 2011.
18. Jeong, D.S., Schroeder, H., and Waser, R. (2007) *Electrochem. Solid-State Lett.*, **10**, G51.
19. Yoon, K.J., Lee, M.H., Kim, G.H., Song, S.J., Seok, J.Y., Han, S., Yoon, J.H., Kim, K.M., and Hwang, C.S. (2012) *Nanotechnology*, **23**, 185202.
20. Liu, S.Q., Wu, N.J., and Ignatiev, A. (2000) *Appl. Phys. Lett.*, **76**, 2749.
21. Beck, A., Bednorz, J.G., Gerber, C., Rossel, C., and Widmer, D. (2000) *Appl. Phys. Lett.*, **77**, 139.
22. Dearnaley, G., Stoneham, A.M., and Morgan, D.V. (1970) *Rep. Prog. Phys.*, **33**, 1129.
23. Dearnaley, G., Morgan, D.V., and Stoneham, A.M. (1970) *J. Non-Cryst. Solids*, **4**, 593.
24. Greene, P.D., Bush, E.L., and Rawlings, I.R. (1968) *Proceedings of the Symposium on Deposited Thin Film Dielectric Materials*, p. 167, Electro-Chemical Society, New York.
25. Barriac, C., Pinard, P., and Davoine, F. (1968) *C.R. Acad. Sci., Paris*, **266**, 423.
26. Hickmott, T.W. (1964) *J. Appl. Phys.*, **35**, 2118.
27. Hickmott, T.W. (1964) *J. Appl. Phys.*, **35**, 2679.
28. Simmons, J.G. and Verderber, R.R. (1967) *Proc. R. Soc. London, Ser. A*, **301**, 77.
29. Kröger, F.A. and Vink, H.J. (1956) *Solid State Phys.*, **3**, 307.
30. Dearnaley, G. (1967) *Phys. Lett.*, **25A**, 760.
31. Sharma, A.A., Norman, M., Abdelmoula, M., Skowronski, M., and Bain, J.A. (2014) *Adv. Funct. Mater.*, **24**, 5522.
32. Yoon, J.H., Han, J.H., Jung, J.S., Jeon, W., Kim, G.H., Song, S.J., Seok, J.Y., Yoon, K.J., Lee, M.H., and Hwang, C.S. (2013) *Adv. Mater.*, **25**, 1987.
33. Schroeder, H. and Jeong, D.S. (2007) *Microelectron. Eng.*, **84**, 1982.
34. Szot, K., Speier, W., Bihlmayer, G., and Waser, R. (2006) *Nat. Mater.*, **5**, 312.
35. Münstermann, R., Menke, T., Dittmann, R., and Waser, R. (2010) *Adv. Mater.*, **22**, 4819.

36. Li, W.X., Österlund, L., Vestergaard, E.K., Vang, R.T., Matthiesen, J., Pedersen, T.M., Lasgsgaard, E., Hammer, B., and Besenbacher, F. (2004) *Phys. Rev. Lett.*, **93**, 146104.
37. Janousch, M., Meijer, G.I., Staub, U., Delley, B., Karg, S.F., and Andreasson, B.P. (2007) *Adv. Mater.*, **19**, 2232.
38. Reed, T.B. (1970) *The Chemistry of Extended Defects in Non-Metallic Solids*, North-Holland, Amsterdam.
39. Marucco, J.-F., Gautron, J., and Lemasson, P. (1981) *J. Phys. Chem. Solids*, **42**, 363.
40. Chiang, Y.-M., Birnie, D.B. III., and Kingery, W.D. (1997) *Physical Ceramics*, John Wiley & Sons, Inc, New York.
41. Goodlet, G., Faty, S., Cardoso, S., Freitas, P.P., Simões, A.M.P., Ferreira, M.G.S., and Da Cunha Belo, M. (2004) *Corros. Sci.*, **46**, 1479.
42. Krylov, O.V. (1970) *Catalysis by Non-metals – Rules for Catalyst Selection*, Academic Press, London.
43. Surnev, S., Ramsey, M.G., and Netzer, F.P. (2003) *Prog. Surf. Sci.*, **73**, 117.
44. Roberson, J.A. and Rapp, R.A. (1969) *J. Phys. Chem. Solids*, **30**, 1119.
45. Chen, W.K. and Swalin, R.A. (1966) *J. Phys. Chem. Solids*, **27**, 57.
46. Kofstad, P. (1962) *J. Electrochem. Soc.*, **109**, 776.
47. Dixon, R.A., Williams, J.J., Morris, D., Rebane, J., Jones, F.H., Egddell, R.G., and Downes, S.W. (1998) *Surf. Sci.*, **399**, 199.
48. Jeong, D.S., Schroeder, H., and Waser, R. (2009) *Phys. Rev. B*, **79**, 195317.
49. Jeong, D.S., Kim, I., Lee, T.S., Lee, W.S., Lee, K.-S., and Lee, D.-K. (2012) *J. Korean Phys. Soc.*, **61**, 913.
50. Mitoff, S.P. (1961) *J. Chem. Phys.*, **35**, 882.
51. Karakasidis, T. and Meyer, M. (1997) *Phys. Rev. B*, **55**, 13853.
52. Barsoum, M. (2003) *Fundamentals of Ceramics*, Taylor & Francis, New York.
53. Kim, K.M., Choi, B.J., Song, S.J., Kim, G.H., and Hwang, C.S. (2009) *J. Electrochem. Soc.*, **156**, G213.
54. Kim, K.M., Jeong, D.S., and Hwang, C.S. (2011) *Nanotechnology*, **22**, 254002.
55. Kim, K.M., Song, S.J., Kim, G.H., Seok, J.Y., Lee, M.H., Yoon, J.H., Park, J., and Hwang, C.S. (2011) *Adv. Funct. Mater.*, **21**, 1587.
56. Kwon, D.-H., Kim, K.M., Jang, J.H., Jeon, J.M., Lee, M.H., Kim, G.H., Li, X.-S., Park, G.-S., Lee, B., Han, S., Kim, M., and Hwang, C.S. (2010) *Nat. Nanotechnol.*, **5**, 148.
57. Strachan, J.P., Pickett, M.D., Yang, J.J., Aloni, S., David Kilcoyne, A.L., Medeiros-Ribeiro, G., and Stanley Williams, R. (2010) *Adv. Mater.*, **22**, 3573.
58. Kim, K.M. and Hwang, C.S. (2009) *Appl. Phys. Lett.*, **94**, 122109.
59. Szot, K., Rogala, M., Speier, W., Klusek, Z., Beshmehn, A., and Waser, R. (2011) *Nanotechnology*, **22**, 254001.
60. Kim, K.M., Choi, B.J., Shin, Y.C., Choi, S., and Hwang, C.S. (2007) *Appl. Phys. Lett.*, **91**, 012907.
61. Kim, G.H., Lee, J.H., Seok, J.Y., Song, S.J., Yoon, J.H., Yoon, K.J., Lee, M.H., Kim, K.M., Lee, H.D., Ryu, S.W., Park, T.J., and Hwang, C.S. (2011) *Appl. Phys. Lett.*, **98**, 262901.
62. Chudnovskii, F.A., Odynets, L.L., Pergament, A.L., and Stefanovich, G.B. (1996) *J. Solid State Chem.*, **122**, 95.
63. Miao, F., Strachan, J.P., Yang, J.J., Zhang, M.-X., Goldfarb, I., Torrezan, A.C., Eschbach, P., Kelley, R.D., Medeiros-Ribeiro, G., and Williams, R.S. (2011) *Adv. Mater.*, **23**, 5633.
64. Yajima, T., Fujiwara, K., Nakao, A., Kobayashi, T., Tanaka, T., Sunouchi, K., Suzuki, Y., Takeda, M., Kojima, K., Nakamura, Y., Taniguchi, K., and Takagi, H. (2010) *Jpn. J. Appl. Phys.*, **49**, 060215.
65. Chen, Y.-S., Wu, T.-Y., Tzeng, P.-J., Chen, P.-S., Lee, H.-Y., Lin, C.-H., Chen, P.-S., and Tsai, M.-J. (2009) Forming-free HfO_2 bipolar RRAM device with improved endurance and high speed operation. Presented at *International Symposium on VLSI Technology, Systems, and Applications, 2009. VLSI-TSA '09*, April 27–29, 2009.
66. Kawai, M., Ito, K., Ichikawa, N., and Shimakawa, Y. (2010) *Appl. Phys. Lett.*, **96**, 072106-3.
67. Chien, W.C., Chen, Y.R., Chen, Y.C., Chuang, A.T.H., Lee, F.M., Lin, Y.Y., Lai, E.K., Shih, Y.H., Hsieh, K.Y., and Chih-Yuan, L. (2010) A forming-free

- WO_x resistive memory using a novel self-aligned field enhancement feature with excellent reliability and scalability. Presented at *IEEE International Electron Devices Meeting (IEDM)*, 2010, December 6–8, 2010.
- 68. Lv, H.B., Yin, M., Song, Y.L., Fu, X.F., Tang, L., Zhou, P., Zhao, C.H., Tang, T.A., Chen, B.A., and Lin, Y.Y. (2008) *IEEE Electron Device Lett.*, **29**, 47.
 - 69. Strachan, J.P., Yang, J.J., Montoro, L.A., Ospina, C.A., Ramirez, A.J., Kilcoyne, A.L.D., Medeiros-Ribeiro, G., and Williams, R.S. (2013) *Beilstein J. Nanotechnol.*, **4**, 467.
 - 70. Jeong, D.S., Cheong, B.-k., and Kohlstedt, H. (2011) *Solid-State Electron.*, **63**, 1.
 - 71. Song, S.J., Seok, J.Y., Yoon, J.H., Kim, K.M., Kim, G.H., Lee, M.H., and Hwang, C.S. (2013) *Sci. Rep.*, **3**, 3443.
 - 72. Degraeve, R., Fantini, A., Chen, Y.Y., Clima, S., Govoreanu, B., Goux, L., Wouters, D.J., Roussel, P., Kar, G.S., Pourtois, G., Cosemans, S., Groeseneken, G., Jurczak, M., and Altimime, L. (2012) Reliability of low current filamentary HfO₂ RRAM discussed in the framework of the hourglass SET/RESET model. Presented at *IEEE International Integrated Reliability Workshop Final Report (IRW)*, 2012, October 14–18, 2012.

11

Universal Switching Behavior

Daniele Ielmini and Stephan Menzel

11.1

General Properties of ReRAMs and Their Universal Behavior

A resistive-switching memory (ReRAM) is a device that can change its resistance by varying the local distribution of chemical species [1]. For instance, filamentary-switching ReRAM based on metal oxides features a conductive filament (CF), namely a localized channel with a lower concentration of oxygen and a higher electrical conductivity, which can be formed or retracted between the two electrodes, thus changing the ReRAM resistance [2]. In uniform-switching ReRAM (also called: *interface-type ReRAM*, see Chapter 16), instead, the oxygen concentration profile along the dielectric thickness can be changed, thus modulating the device resistance across the whole ReRAM area [3]. To improve and optimize ReRAM, therefore, the choice of the material and the engineering of the electrode/insulator stack are essential. From the empirical point of view, guidelines for the material selection must originate from a detailed comparison of the switching characteristics for different materials compositions and layer stacks. From this comparison, the switching behavior and parameters are generally similar for a broad spectrum of filamentary-type ReRAM materials, suggesting a *universal switching* behavior [4–7]. Understanding the microscopic origin of such universal switching is the key for a physically sound description of the ReRAM switching mechanisms and for a physics-based materials engineering of ReRAM devices.

To address universal switching in ReRAM devices, we will focus on filamentary-type ReRAM characteristics, such as the bipolar-switching HfO_x -based ReRAM [8]. Figure 11.1 shows the current–voltage (I – V) characteristics in a HfO_x -based ReRAM [5, 8]. Set transition occurs at positive voltage, while reset transition is operated at negative voltage. To avoid destructive breakdown during set process, the current is limited to the so-called compliance current I_C , either by connecting the ReRAM device with a transistor biased in its saturation regime [8] or through an external setup [9]. Three I – V curves are shown in the figure, corresponding to increasing I_C : The resistance R in the set state decreases with I_C , since the formed CF is increasingly more conductive. The reset current I_{reset} , marking the onset of the reset transition, is approximately equal to I_C . These empirical

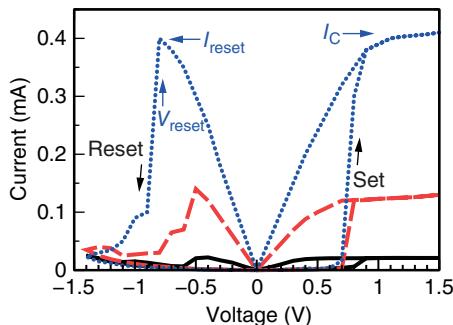


Figure 11.1 Measured I - V characteristics for a HfO_x -based ReRAM with bipolar switching. In positive-voltage set transition, the current is limited by a compliance current I_C , which controls the resistance in the set

state R_{ON} . In negative-voltage reset transition, the reset current I_{reset} is equal to I_C . (Reproduced with permission from Ref. [5]. © 2011, IEEE.)

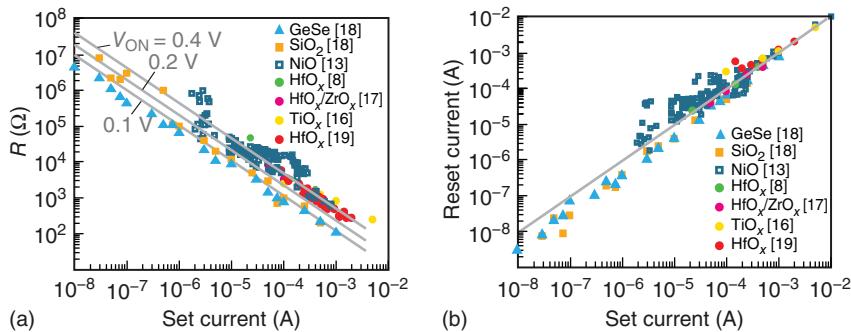


Figure 11.2 Measured R (a) and I_{reset} (b) as a function of I_C for different materials and switching concepts, including TCM cells based on NiO [13], VCM cells based on HfO_x [8, 19], $\text{HfO}_x/\text{ZrO}_x$ [17], and TiO_x

[16], and Cu ECM with various electrolytes, for example, $\text{Cu}:\text{SiO}_2$ [18] and $\text{Ag}:\text{GeSe}$ [18]. (Reproduced with permission from Ref. [20]. © 2012, Wiley-VCH.)

observations are summarized in Figure 11.2, showing the measured R (a) and I_{reset} (b) as a function of I_C for filamentary-type ReRAMs [4, 5]. The programmed set-state resistance obeys the empirical law $R = V_C/I_C$ (Figure 11.2a), where V_C is a material-dependent constant, while the reset current depends linearly on the set current, that is, $I_{\text{reset}} \propto I_C$ (Figure 11.2b). These universal characteristics were first observed for electrochemical metallization memory (ECM) cells [10, 11], then later demonstrated for oxide-based bipolar-switching ReRAM (or valence-change memory, VCM [1]) and unipolar-switching ReRAM (or thermochemical memory, TCM) cells [4, 5, 9, 12]. In both VCM and TCM devices, the reset current is approximately equal to I_C as also shown in Figure 11.1, whereas ECM devices exhibit asymmetric switching with $I_{\text{reset}} < I_C$. The universal switching characteristics have been demonstrated for up to 10 orders of magnitude in the

case of ECM cells [6] and for approximately 4 orders of magnitude for VCM and TCM cells, reflecting the lower resistance window and the lower reset-state resistance of VCM/TCM as compared to ECM. A large variety of materials have been considered in Figure 11.2, including unipolar-switching TCM based on NiO [13], bipolar-switching VCM based on HfO_2 [8, 14, 15, 19], TiO_2 [16], $\text{HfO}_x/\text{ZrO}_x$ [17], and bipolar-switching Cu-based ECM with various electrolytes, namely SiO_2 [18] and GeSe [18]. Besides demonstrating a striking universal relationship between the switching parameters, data in Figure 11.2 suggest the possibility of multilevel cell (MLC) operation of ReRAM, where a large number of resistance levels can be programmed by appropriately tuning the value of I_C . For instance, four programmed levels allow the representation of 2 bits thus reducing the effective bit size in the array and improving the area efficiency of the storage.

Another common behavior of resistive-switching devices is the highly nonlinear switching kinetics, where the set/reset times change with the voltage according to an approximately exponential dependence [21]. Figure 11.3 shows the nonlinear switching kinetics for various VCM-type devices (a), including SrTiO_x [30], TaO_x [22], HfO_x [15, 23, 27, 28, 29], and TiO_x [24] and for ECM-type device (b), namely AgI with Ag electrode [25], RbAg_4I_5 [31], Ag_2S [32], Cu_2S [33], and GeS_x with Ag electrodes [34–36]. Depending on the specific ReRAM material, different physical processes may be the origin of this nonlinearity, including ion transport, electron-transfer reactions at the boundaries, nucleation processes during phase transformations, or the growth of a phase. In general, the slowest process determines the switching kinetics. The switching kinetics for set and reset transitions in HfO_x , for example, was shown to be in agreement with an Arrhenius law with activation energy of about $E_A = 1.2 \text{ eV}$, where the local temperature enhanced by Joule heating accelerates ion migration at the origin of the set and reset processes [5, 15, 26]. In addition to the temperature acceleration, the activation energy can be lowered by the application of an electric field, as in the

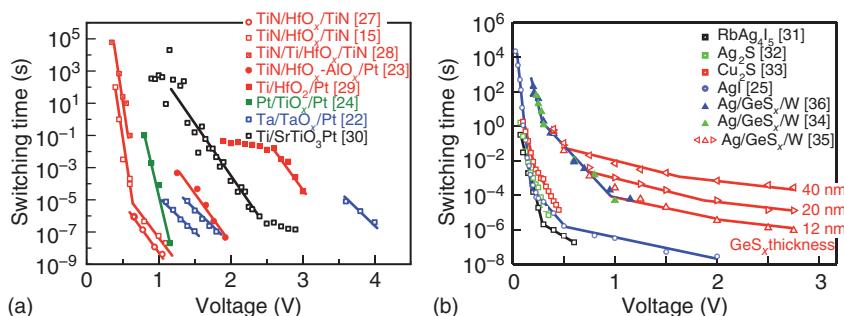


Figure 11.3 Nonlinear switching kinetics based on voltage pulse experiments for various VCM cells (a), namely HfO_x [15, 27–29], $\text{HfO}_x-\text{AlO}_x$ [23], TiO_x [24], TaO_x [22], and SrTiO_x [30], and for various ECM cells

(b), namely RbAg_4I_5 [31], Ag_2S [32], Cu_2S [33], AgI [25], and $\text{Ag}/\text{GeS}_x/\text{W}$ [34–36]. (Reproduced with permission from Ref. [37]. © 2015, Wiley-VCH.)

Poole–Frenkel carrier transport, thus increasing the rate exponentially. In other cases, different rate-limiting processes might govern the switching kinetics, depending on the voltage regime. For example, it has been demonstrated that the switching kinetics in AgI-based ECM cells are limited by a nucleation process at low voltages, by electron-transfer reactions at an intermediate voltage range and by a combination of ion hopping and electron-transfer reactions at high voltage [25]. This could explain the presence of different slopes in Figure 11.3b. All of the rate-limiting processes are connected to some activation barrier. Thus, they obey an exponential rate law, leading to a nonlinear switching kinetics. A more detailed discussion on the origin of the nonlinear switching kinetics is found in Chapter 14.

11.2

Explaining the Universal Switching of ReRAM

The observation of a universal switching of ReRAM, namely the unique dependence of the set-state properties on I_C (Figures 11.1 and 11.2) and the nonlinear switching kinetics (Figure 11.3), suggests that some fundamental general laws govern the physics of the switching processes in all ReRAM devices. Changing the ReRAM materials as shown in Figures 11.2 and 11.3 generally results in a weak shift of the position of the respective curve, with no significant change in the analytical expression. For instance, different values of $V_C = RI_C$ are found in Figure 11.2a by changing the ReRAM concept from VCM (e.g., HfO_x) to ECM; however, all the curves maintain a constant value of V_C . Understanding these general physical laws of ReRAM switching is therefore of paramount importance to address the device technology and engineering.

We first aim to explain the I_C -dependent programming shown in Figure 11.2, which can be interpreted according to two different scenarios. In the first approach, the different values of resistance in the set state are explained in terms of a variable diameter of the CF, as described in Figure 11.4a. In the second scenario, instead, the different resistance states are represented by a varying tunneling gap between the filament tip and the counter electrode as shown in Figure 11.4b. To compare these two possible approaches, the achievable resistance states are plotted in Figure 11.4c,d, respectively, as a function of the varied parameter, namely CF diameter ϕ and tunneling gap x , respectively. Assuming a resistivity of $400 \mu\Omega \text{ cm}$, corresponding to a metallic-Hf filament with enhanced resistivity according to the Fuchs–Sondheimer formula [5, 38, 39], the variable-diameter case covers 4 orders of magnitude in the diameter range between 1 and 100 nm, with a maximum value of about $100 \text{ k}\Omega$. Note that the CF conductivity can be significantly degraded in nanoscale filaments due to surface scattering and the presence of several defects, for example, the oxygen vacancies constituting the CF; therefore, much higher values of resistance might be achieved at small ϕ . On the other hand, the variable-gap model can cover more than 10 orders of magnitude with a minimum resistance of $>10 \text{ k}\Omega$, which coincides with the single atomic contact resistance $R_0 = 12.9 \text{ k}\Omega$.

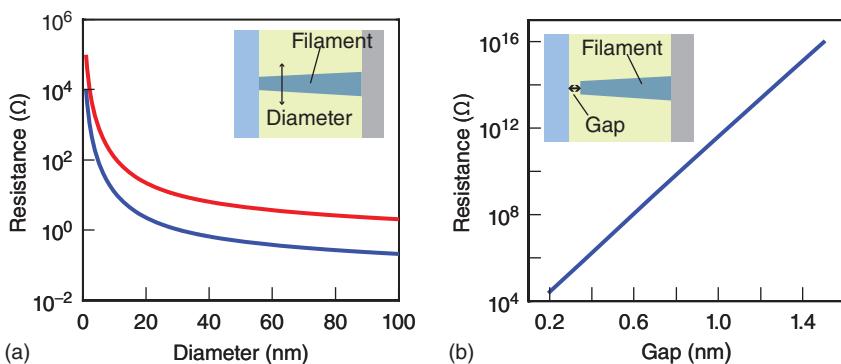


Figure 11.4 (a) Calculated resistance for variable-diameter type using Fuchs-Sondheimer formula [38, 39] and a resistivity of $400 \mu\Omega \text{ cm}$ (red) and $200 \mu\Omega \text{ cm}$ (blue), corresponding to metallic Hf and Cu filaments, respectively. (b) Calculated resistance

for the variable-gap type using Simmons equation [40] for electron tunneling. A barrier height of 3.6 eV, a filament diameter of 4 nm, and a relative electron effective mass of 1 are assumed.

In the following two sections, both scenarios are discussed based on 1D physical models.

11.3 Variable-Diameter Model

According to the variable-diameter model, the bipolar switching shown in Figure 11.1 can be understood by the localized migration of ionized defects, which leads to CF growth and dissolution under positive and negative voltages, respectively. The set transition is sketched in Figure 11.5, showing four snapshots of the CF formation during the set transition with positive bias applied to the top electrode [41]. Defects (blue spheres in the figure) are initially accumulated at the top electrode side, as shown in Figure 11.5a, due to the previous reset operation

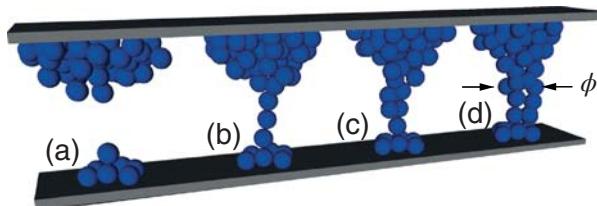


Figure 11.5 Schematic illustration of the set transition in a bipolar RRAM. In the reset state, defects are initially accumulated at the top electrode (a). Application of a positive voltage to the top electrode induces defect

migration toward the bottom electrode, resulting in CF formation (b) and growth (c,d). (Reprinted with permission from Ref. [41]. © 2013, Wiley-VCH.)

under negative voltage. The defects might be positively charged or neutral, although in the latter case they can be ionized by the high electric field and high temperature during set and reset processes. The ionized defects are modeled as cations, namely ions with a positive charge, as expected for oxygen vacancies or metallic impurities. Positive ions thus migrate along the field direction toward the bottom electrode, resulting in the formation of a CF as shown in Figure 11.5b [5], followed by growth as shown in Figure 11.5c,d.

The growth process results in a decrease in resistance due to the increase in CF diameter ϕ . The set transition can thus be described by a simple analytical equation for the rate of increase of ϕ , given by:

$$\frac{d\phi}{dt} = A \exp\left(-\frac{E_A}{k_B T}\right), \quad (11.1)$$

where A is a pre-exponential constant, E_A is the activation energy for defect migration, k_B is the Boltzmann constant, and T is the local temperature [5]. Equation (11.1) relies on an Arrhenius-type temperature activation, which is consistent with the Mott–Gurney model for ionized-defect hopping at localized states [36, 42, 43]. In the ion-hopping models, the energy barrier decreases as in the Poole–Frenkel transport [44], thus the activation energy in Eq. (11.1) can be written as:

$$E_A = E_{A0} - \alpha eV, \quad (11.2)$$

where E_{A0} is the energy barrier without any applied field, α is a constant describing the hopping distance [44], and V is the voltage across the ReRAM device. The maximum temperature T along the CF is also controlled by voltage through Joule heating according to:

$$T = T_0 + \frac{R_{th}}{R} V^2, \quad (11.3)$$

where T_0 is the ambient temperature and R_{th} is the effective thermal resistance of the CF [5]. Due to the voltage-dependent Eqs. (11.2) and (11.3), the CF growth rate in Eq. (11.1) is a sole function of the voltage across the ReRAM, thus resulting in a voltage-driven switching kinetics.

The voltage-driven kinetics of switching can be evidenced by a detailed study of the time evolution of resistance during the set process. Figure 11.6 shows the measured and calculated resistance R (a) as a function of time during the set transition for a HfO_x ReRAM [15]. A load resistance $R_L = 1 \text{ k}\Omega$ was positioned in series with the ReRAM as shown in Figure 11.7a, to serve as a compliance system limiting the maximum current. The ReRAM was initially prepared in the reset state with a resistance of about $5 \text{ k}\Omega$, then alternating set and read pulses were applied as shown in Figure 11.7b. As a result, R could be monitored during the set transition along eight decades of time, from 10 ns to 1 s . A constant voltage V_A from 1 to 3.5 V was applied for inducing the set process. The resistance as shown in Figure 11.6a decreases with time, due to the CF growth, and the R decrease is increasingly faster for increasing V_A due to the stronger driving force for CF growth, that is, electric field and/or local temperature [15].

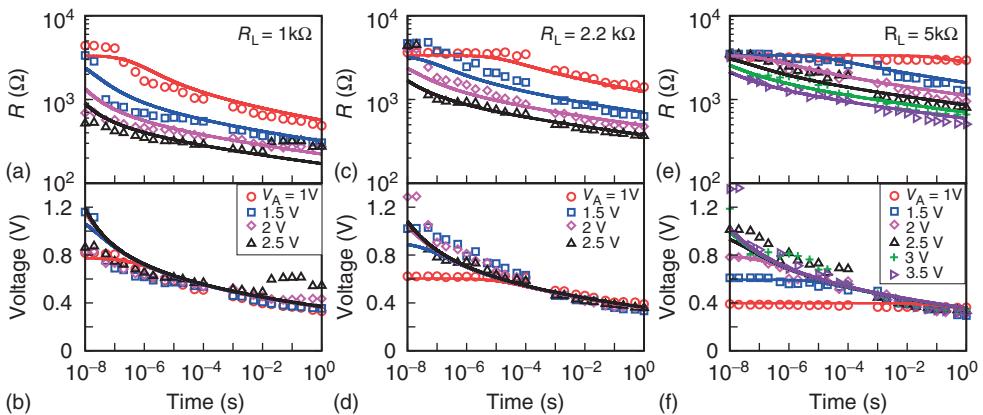


Figure 11.6 Measured and calculated R (a) and voltage across the ReRAM (b) as a function of time, during the set transition with a load resistance $R_L = 1\text{ k}\Omega$ for increasing voltage V_A . Data and calculations are also shown for $R_L = 2.2\text{ k}\Omega$ (c,d) and $R_L = 5\text{ k}\Omega$ (e,f).

The voltage across the ReRAM follows a universal behavior with time, supporting the voltage-driven model of Eqs. (11.1)–(11.3). (Reprinted with permission from Ref. [15]. © 2012, IEEE.)

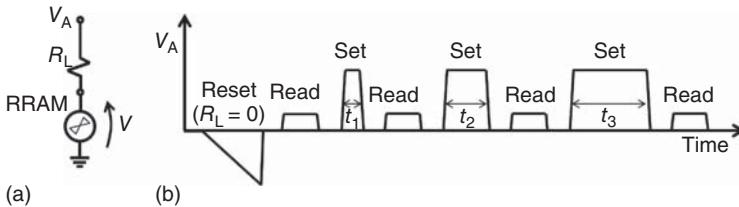


Figure 11.7 Experimental setup for data shown in Figure 11.6, consisting of a ReRAM in series with a load resistor R_L (a), and waveform of the applied voltage V_A (b). (Reprinted with permission from Ref. [15]. © 2012, IEEE.)

Figure 11.6b shows the measured voltage across the ReRAM during the set transition (see Figure 11.7a), showing a monotonic decrease with time from about 1 V at 10 ns to about 0.4 V at 1 s. The time evolution of the voltage is strikingly independent of the applied voltage V_A , suggesting a universal voltage behavior during the set transition. These results indicate that V always adjusts to the same characteristic value at any given time, irrespective of the external conditions (e.g., applied voltage and load resistance). For instance, a voltage 0.4 V is found around 1 s, which can be taken as a representative figure for the voltage needed to induce migration on the corresponding timescale of about 1 s.

Figure 11.6c,d show the measured R and V , respectively, for $R_L = 2.2\text{ k}\Omega$, while similar results are shown in Figure 11.6e,f for a load resistance $R_L = 5\text{ k}\Omega$. A larger R_L leads to a larger value of the CF resistance achieved at the end of the set transition around 1 s, as a result of the limitation to a lower current. Figure 11.6d,f confirm that the voltage across the ReRAM follows a universal time dependence, further supporting the voltage-driven model for CF growth in Eqs. (11.1)–(11.3).

Calculations from this analytical model are shown in Figure 11.6, using the parameters $E_{A0} = 1.2 \text{ eV}$, $\alpha = 0.3$ and $A = 1 \text{ ms}^{-1}$. The CF resistance was estimated by the formula:

$$R = \rho \frac{4t_{\text{ox}}}{\pi\phi^2}, \quad (11.4)$$

which was also used in the calculations of Figure 11.4c. In Eq. (11.4), ρ is the electrical resistivity and t_{ox} is the oxide thickness assumed to be equal to the CF length. The electrical resistivity was assumed to be $\rho = 400 \mu\Omega \text{ cm}$, that is, higher than the value for bulk metallic Hf ($\rho_{\text{bulk}} = 33 \mu\Omega \text{ cm}$) due to carrier scattering at the surface of the nanoscale CF. Scattering-degraded resistivity in nanoscale wires can be estimated by the Fuchs–Sondheimer approximation [38, 39]:

$$\rho = \rho_{\text{bulk}} \left(1 + \frac{3\lambda}{4\phi} (1 - p) \right), \quad (11.5)$$

where λ is the electron mean free path and p is the specularity factor describing the probability for elastic scattering at the CF surface. By assuming $\lambda = 28 \text{ nm}$ and $p = 0.5$ in Eq. (11.5), we obtain $\rho = 380 \mu\Omega \text{ cm}$ for Hf at $\phi = 1 \text{ nm}$. The resistivity can be further degraded by phonon scattering due to the high temperature during the set/reset transitions, surface roughness scattering, and scattering due to grain boundaries and other defects (vacancies, impurities) along the CF.

The effective thermal resistance R_{th} in Eq. (11.3) is defined as the ratio between the maximum temperature reached in the CF due to Joule heating and the dissipated power $P = VI = V^2/R$. The thermal resistance can thus be obtained by the Fourier equation in one dimension, namely:

$$k_{\text{th}} \frac{d^2T}{dz^2} + \rho F^2 = 0, \quad (11.6)$$

where k_{th} is the thermal conductivity, F is the electric field, and z is the axis coordinate along the CF. Equation (11.6) yields a parabolic T profile where the maximum temperature T can be obtained by Eq. (11.3) with $R_{\text{th}} = L/(8k_{\text{th}}\pi\phi^2/4)$ [5]. A thermal conductivity $k_{\text{th}} = 23 \text{ W m}^{-1} \text{ K}^{-1}$, thus equal to the bulk metallic conductivity in Hf, was used in the calculations. By substituting the electrical and thermal resistance formulas into Eq. (11.3), we obtain:

$$T = T_0 + \frac{V^2}{8\rho k_{\text{th}}}, \quad (11.7)$$

from which one can conclude that the local temperature is solely controlled by the applied voltage and does not depend on ϕ . This is because the Fourier equation in Eq. (11.6) only depends on the local electric field, which is the same for a given voltage for any value of ϕ . Finally, by substituting Eq. (11.7) into Eq. (11.1), we obtain:

$$\frac{d\phi}{dt} = A \exp \left(-\frac{E_{A0} - \alpha eV}{k_B T_0 (1 + V^2 / 8\rho k_{\text{th}} T_0)} \right), \quad (11.8)$$

highlighting once again the voltage-controlled nature of the growth dynamics in agreement with the data shown in Figure 11.6. Note that the same equation can

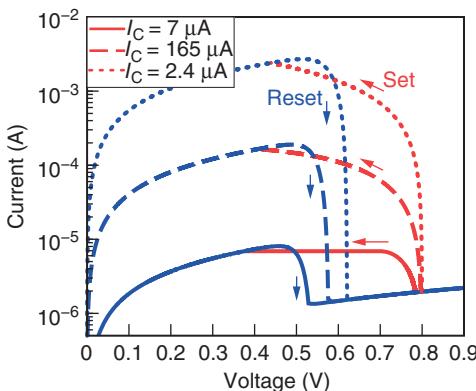


Figure 11.8 Calculated I - V curves using Eq. (11.8) for increasing compliance current, namely $I_C = 7 \mu\text{A}$, $165 \mu\text{A}$, and 2.4 mA . For set transition, a constant voltage of 0.8 V was applied to the cell in series with a transistor to limit the current to I_C . For the reset

transition, a voltage sweep from 0 to 0.9 V was applied. The resistance and reset current depend on I_C used in the set transition. (Reprinted with permission from Ref. [5].)

© 2011, IEEE.)

be used for the reset transition under negative voltage, provided that a minus sign, instead of a plus, is used to describe the diameter reduction instead of growth.

Figure 11.8 shows the calculated I - V curves using Eq. (11.8). In these calculations, the voltage was increased at a constant rate $dV/dt = 1 \text{ Vs}^{-1}$ and R was changed by Eq. (11.4), where ϕ was updated at any simulation step through Eq. (11.8). In this equation, plus and minus signs were used to describe growth and dissolution, respectively. The same physical parameters were used for set and reset processes, namely $E_{A0} = 1.2 \text{ eV}$, $A = 1 \text{ ms}^{-1}$, and $\alpha = 0.3$ [5]. Compliance current values $I_C = 7 \mu\text{A}$, $165 \mu\text{A}$, and 2.4 mA were used, resulting in a decrease in R and an increase in I_{reset} for increasing I_C . The I_C dependence of the CF resistance is due to a negative feedback resulting from the compliance limitation combined with the voltage-controlled switching kinetics: as the set transition takes place at the set voltage V_{set} , the growth-induced decrease in R first leads to a current increase, then to a voltage decrease as the current reaches I_C . The voltage decrease (or snap-back) is dictated by the universal time evolution shown in Figure 11.6; thus, the set process practically stops at the critical value V_C , which is the minimum voltage capable of inducing ion migration in the device. For the typical timescales of DC experiments of about 1 s , V_C is approximately equal to 0.4 V (see Figure 11.6); thus, the final set state is characterized by $V = V_C$, namely a resistance given by Ref. [5]:

$$R = \frac{V_C}{I_C}, \quad (11.9)$$

where V_C depends on the timescale of the set experiment. Equation (11.9) describes an inverse proportionality between I_C and the final resistance in the set state, which accounts for the R - I_C relationship in Figure 11.2.

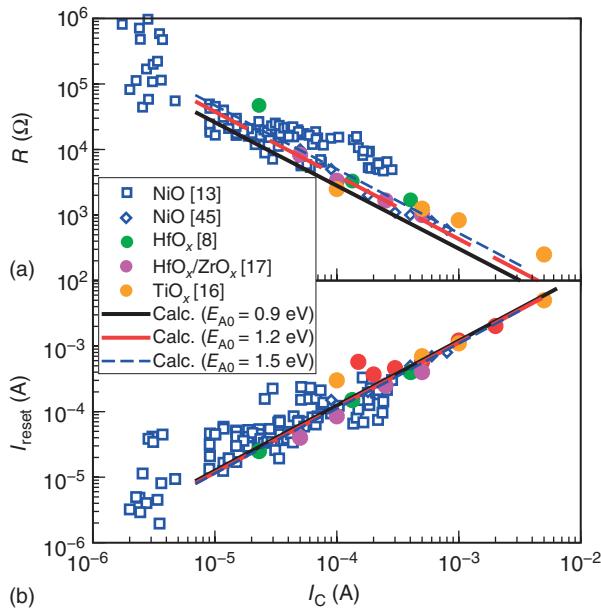


Figure 11.9 Measured and calculated set-state resistance R (a) and reset current I_{reset} (b) as a function of I_C . Several device materials are compared, that is, unipolar-switching ReRAM based on NiO [13, 45], bipolar-switching VCM based on HfO_x

[8], TiO_2 [16], and $\text{HfO}_x/\text{ZrO}_x$ [17]. Calculations agree with the experimental data and are consistent with the analytical formulas of Eqs. (11.9) and (11.10). (Reprinted with permission from Ref. [5]. © 2011, IEEE.)

Figure 11.8 also shows that the reset current I_{reset} for the onset of the reset process is given by:

$$I_{\text{reset}} = I_C \quad (11.10)$$

since the reset voltage $V_{\text{reset}} = RI_{\text{reset}}$ is equal to $V_C = RI_C$, namely, migration is equally activated at positive and negative polarities.

Figure 11.9 shows the measured and calculated R (a) and I_{reset} (b) as a function of I_C , comparing the results of Eq. (11.10) with experimental data in the literature for several material systems, namely unipolar ReRAM based on NiO [13, 45], bipolar-switching VCM based on HfO_x [8], TiO_2 [16], and $\text{HfO}_x/\text{ZrO}_x$ [17]. Calculations were done by the analytical model of Eq. (11.8) with variable E_{A0} . The results confirm Eqs. (11.9) and (11.10) with $V_C = 0.4$ V, thus supporting the voltage-driven migration model in Eq. (11.8). Although the model was developed for bipolar switching, namely ECM and VCM systems, unipolar switching also seems to agree with Eqs. (11.9) and (11.10), thus suggesting a similar voltage-driven switching dynamics with $V_{\text{reset}} \approx V_C \approx 0.4$ V also for unipolar-switching ReRAM. Calculations with lower E_{A0} and higher A could also account for ECM data with smaller V_C as shown in Figure 11.2 [46]. Resistance values larger than $100\text{ k}\Omega$ are obtained by the calculations in the figure by using unphysical values of the effective diameter

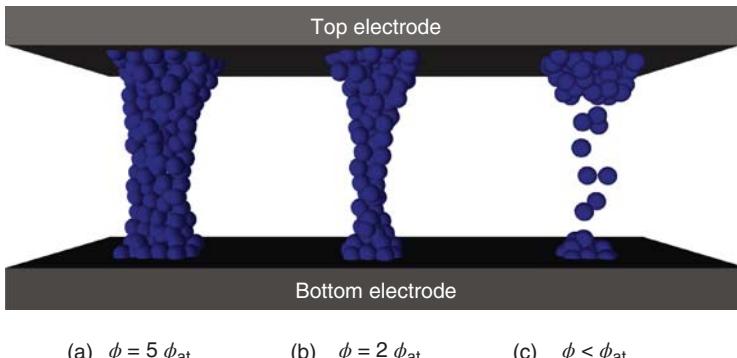


Figure 11.10 Sketch of a CF consisting of a localized conduction channel composed of point defects with atomic size ϕ_{at} . Depending on the number of defects belonging to the CF, the effective diameter can have a variable size, namely $\phi > \phi_{\text{at}}$ (a), $\phi \approx \phi_{\text{at}}$ (b), or finally $\phi < \phi_{\text{at}}$ (c). An effective CF size

smaller than ϕ_{at} can be understood for a discontinuous CF where defects act as local dopant, thus resulting in the local resistivity being higher than the metallic value even taking Fuchs–Sondheimer effects into account.

ϕ below the single-atom limit. Such small values of ϕ can be understood by “subatomic” filaments as sketched in Figure 11.10. Here, the CF consists of atomic-size defects, each with diameter ϕ_{at} . Depending on the number of defects in the filament, which is controlled by I_C , the effective diameter can be either higher than ϕ_{at} (a), or comparable to ϕ_{at} , (b), or smaller than ϕ_{at} (c). The latter case corresponds to atomic defects that are sparsely aligned along the CF, thus leading to a doping effect where the Fermi level remains below the conduction band edge of the semiconductor band structure [5, 47, 48]. Although the geometrical size is clearly still of the order of ϕ_{at} as shown in Figure 11.10c, the resistivity is significantly larger, because carrier hopping among defect states in the CF is necessary for electrical transport. Evidence for hopping transport mechanisms in small filaments is provided by nonlinear conduction characteristics, Poole–Frenkel effects, and semiconductor-like transport at high resistance in ReRAM [47]. In this regime, the variable-diameter model is thus similar to the variable-gap model discussed in the next section, where the high resistance is explained by tunneling effects instead of hopping conduction [6, 7].

The results shown in Figure 11.6 suggest that an increasing voltage is needed to induce migration at increasingly shorter times. This can be explained by ion migration being driven by electric field and temperature, both of which are controlled by the voltage across the device. This is also consistent with the results shown in Figure 11.3, where the switching times display a strong nonlinear dependence on voltage. To understand the universal switching times in ReRAM, Figure 11.11 shows the measured set time t_{set} , and reset times t_{reset} , which are defined in correspondence to a 50% decrease or increase of R , respectively [15]. Data below 10 ms were obtained from rectangular pulses, whereas data above 10 ms refer to

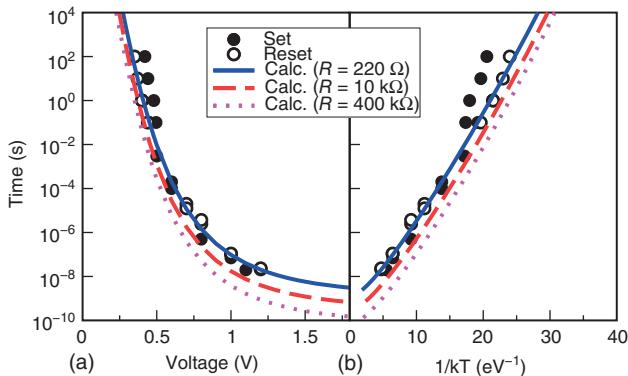


Figure 11.11 Measured and calculated set/reset times as a function of voltage (a) and as a function of $1/kT$, where T is the local temperature due to Joule heating (b). Set times are almost equal to reset times, except for a small overvoltage effect at

long times possibly due to nucleation. The Arrhenius plot indicates an almost linear dependence on $1/kT$, thus supporting the temperature-activated model for migration. (Reproduced with permission from Ref. [15]. © 2012, IEEE.)

triangular voltage sweeps [15]. Set and reset times follow a similar voltage dependence, irrespective of the pulse shape, thus suggesting that ion migration controls both processes in HfO_x -based bipolar-switching ReRAM, as already proposed [4, 5]. Set and reset times were calculated by:

$$t_{\text{set},\text{reset}} = \frac{\Delta\phi}{A^*} \exp\left(\frac{E_{\text{A}0} - e\alpha V}{k_B(T_0 + (R_{\text{th}}/R)V^2)}\right), \quad (11.11)$$

which was obtained from Eq. (11.8) with the approximation $d\phi/dt = \Delta\phi/t_{\text{set},\text{reset}}$. In the calculations, resistance values $R = 220 \Omega$, $10 \text{ k}\Omega$, and $400 \text{ k}\Omega$ were assumed for the study of the impact of CF size on the switching times. The value of 220Ω was the same as in the experimental data. The same parameters as those used in the calculations shown in Figure 11.6 were used. The switching times t_{set} and t_{reset} display a steep voltage dependence, covering about 10 decades in less than 1 V, suggesting a pseudo-exponential decrease in set/reset times at increasing voltage across the ReRAM [21, 23]. Such a large change in set/reset kinetics is important to allow for the necessary immunity to read and program operations, since the low voltage occurring in half-selected cells during program and selected cells during read does not lead to significant change in resistance because of low driving force for migration. The slightly longer t_{set} with respect to t_{reset} at low voltage shown in the figure might be due to the contribution of the filament formation process, which precedes growth in the set transition. At a higher voltage, the two switching times show approximately the same value, thus suggesting that filament formation is not the limiting step in this regime. This allows to neglect the formation process and to describe the observed time–voltage relationship based only on the growth model.

Figure 11.11b shows the measured and calculated t_{set} and t_{reset} (same as in Figure 11.11a) as a function of $1/kT$, where T is the local temperature in the CF estimated from Eq. (11.3). The Arrhenius plot indicates an almost linear behavior, thus further supporting the thermally activated growth model represented by Eq. (11.8). Both the given data and calculations show a weak nonlinearity in the Arrhenius plot, which can be attributed to the voltage-dependent activation energy represented by Eq. (11.2). The increase in V , in fact, causes a lowering of E_A on the basis of the decreasing slope shown in Figure 11.11b for decreasing $1/kT$.

11.4

Variable-Gap Model

In the variable-gap model, a tunneling gap x between a growing filament and its counter electrode determines the resistance state. This model has been developed for the simulation of ECM cells and applied to model the universal switching characteristics [6, 7], set switching kinetics [25], and complementary resistive switching (CRS) [49]. It covers all stages of filament formation in ECM cells: nucleation, electron-transfer reactions, ion hopping, and filamentary growth (cf. Figure 11.12a).

Figure 11.12b shows the equivalent circuit diagram of the model. It considers a growing cylindrical filament with area A_{fil} within an insulating switching layer of length L . It is discriminated between an ionic (I_{ion}) and an electronic current

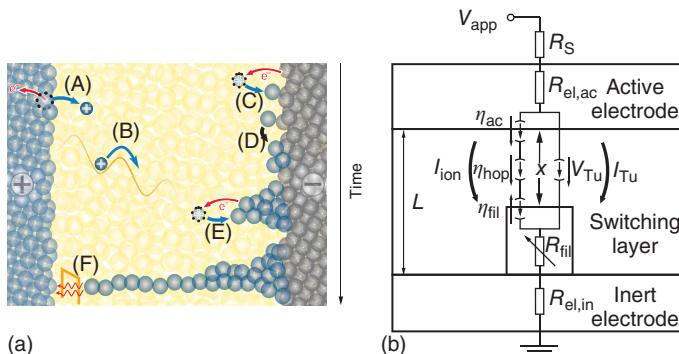


Figure 11.12 (a) Illustration of the electrochemical processes during set switching. A positive polarity is applied to the active Ag/Cu electrode (A) Oxidation of the Ag/Cu active electrode (charge-transfer reaction) and dissolution. (B) Migration of Ag/Cu cations under the applied electric field. (C) Reduction reaction at the inert electrode/solid film interface. (D) Nucleation

process prior to (E) filamentary growth driven by further reduction process.

(F) When the filament approaches the active electrode, significant electron tunneling current sets in and the cell switches to a low resistive state. (b) Equivalent circuit diagram of the simulation model. (From Ref. [25]. Reproduced by permission of the PCCP Owner Societies.)

path (I_{Tu}). The growth/dissolution of the filament under application of a voltage is described by Faraday's law:

$$\frac{dx}{dt} = -\frac{M_{\text{Me}}}{ze\rho_{\text{m,me}}} J_{\text{Me}^{z+}}. \quad (11.12)$$

Here, M_{Me} denotes the atomic mass, z the ion charge, e the elementary charge, $\rho_{\text{m,me}}$ the deposited metal mass density, and $J_{\text{Me}^{z+}}$ the ionic current density. The ionic current is limited by the electron-transfer reactions, that is, oxidation and reduction, occurring at the metal/insulator boundaries and the ion hopping process within the switching layer. Mathematically, the electron-transfer reactions is described by the Butler–Volmer equation

$$J_{\text{Me}^{z+}} = J_{\text{BV}}(\eta) = j_0 \left\{ \exp \left(\frac{(1-\alpha)ez}{kT} \eta \right) - \exp \left(-\frac{\alpha ez}{kT} \eta \right) \right\}, \quad (11.13)$$

where j_0 is the exchange current density, α the charge transfer coefficient, and η the overpotential. Note that the exchange current density is temperature-dependent and obeys an Arrhenius-type law. In Figure 11.12, the overpotentials are denoted as η_{ac} and η_{fil} corresponding to the relevant interface. The ion-hopping transport is driven by the hopping overpotential η_{hop} according to the Mott–Gurney law for ion hopping:

$$J_{\text{Me}^{z+}} = 2zeacf \exp \left(-\frac{E_{\text{hop}}}{k_B T} \right) \sinh \left(\frac{aze}{2k_B T} F \right). \quad (11.14)$$

In Eq. (11.14), c denotes the ion concentration, a the hopping distance, f the attempt frequency, and E_{hop} the migration barrier height. The electric field F depends on the tunneling gap according to $F = \eta_{\text{hop}}/x$. The electronic tunneling current is modeled using Simmons equation

$$I_{\text{Tu}} = \frac{eA_{\text{fil}}}{2\pi\hbar x^2} \left(\Delta W_0 - \frac{eV_{\text{Tu}}}{2} \right) \exp \left(-\frac{4\pi x}{h} \sqrt{2m_{\text{eff}}} \sqrt{\Delta W_0 - \frac{eV_{\text{Tu}}}{2}} \right) \\ - \frac{eA_{\text{fil}}}{2\pi\hbar x^2} \left(\Delta W_0 + \frac{eV_{\text{Tu}}}{2} \right) \exp \left(-\frac{4\pi x}{h} \sqrt{2m_{\text{eff}}} \sqrt{\Delta W_0 + \frac{eV_{\text{Tu}}}{2}} \right). \quad (11.15)$$

It depends on the tunneling barrier height ΔW_0 , the effective electron mass m_{eff} , Planck's constant h , and the tunneling voltage V_{Tu} . By applying Kirchhoff's laws to the equivalent circuit diagram in Figure 11.12, the ionic current density is calculated and the ordinary differential equation (11.12) is solved. To simulate the switching kinetics, Menzel *et al.* also included the nucleation time prior to the simulation of filamentary growth [25]. Figure 11.13 shows the corresponding simulation results for an AgI-based ECM cells. Based on the simulation results, three different regimes can be identified (I, II, and III). In regime I, the nucleation process limits the switching speed, and in regime II, the electron-transfer reaction occurring at the boundary becomes the rate-limiting step. Also for higher voltages, the ion migration becomes a limiting factor. Since the electron-transfer reaction also plays an important role in this regime, it is called *mixed-control regime*.

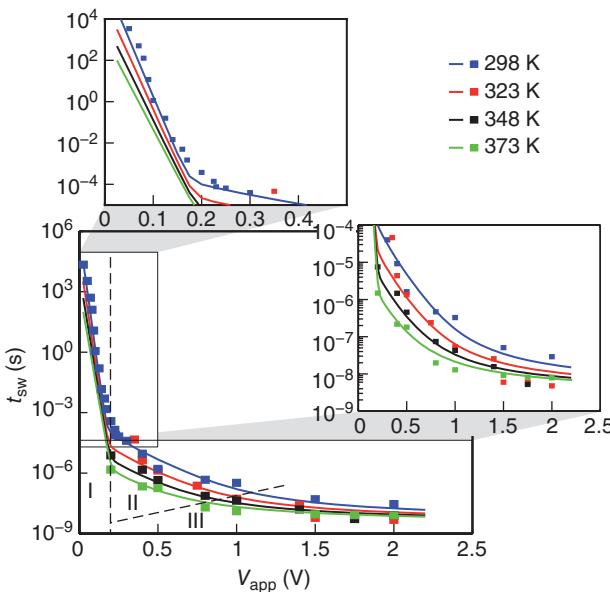


Figure 11.13 Pulsed set switching kinetics of the AgI-based ECM cell for different ambient temperatures $T = 298\text{ K}$ (blue), 323 K (red), 348 K (black), and 373 K (light green). The simulated data are displayed using solid lines and the experimental data

using squares. I, II, III mark the nucleation-limited, the electron-transfer-limited and the mixed-control regime, respectively. (From Ref. [25]. Reproduced by permission of the PCCP Owner Societies.)

Figure 11.14 illustrates that the simulated I - V characteristics based on this model fit very well to the experimental data [6]. Particularly, the cell voltage decrease (in blue) during current compliance captures the experimental data. The cell voltage drops to a value of approximately 0.4 V and then stays almost constant.

By changing the compliance current I_C , the device resistance can be precisely tuned as illustrated in Figure 11.15a. The described simulation model accurately describes this dependence over many orders of magnitude. Evidently, the device resistance obeys the empirical law $R = V_C/I_C$, whereas V_C equals the constant voltage level reached in the current compliance as discussed. As illustrated in Figure 11.15a, the voltage V_C also depends on the sweep rate $\beta = V_{\max}/t_{\text{rise}}$. It increases with increasing sweep rate as R increases.. To understand the growth dynamics during the current compliance, the growth velocity v during current control and the corresponding cell voltage are plotted against the tunneling gap for different current compliances and sweep rates. The set voltage corresponds to the point in voltage at which the current compliance is reached during the I - V sweep. Apparently, the v - x and V - x characteristics for different current compliances are the same and are shifted parallel to each other by a constant gap. The steep slope in the v - x characteristic is attributed to the highly nonlinear switching kinetics and the high sensitivity of the tunneling current

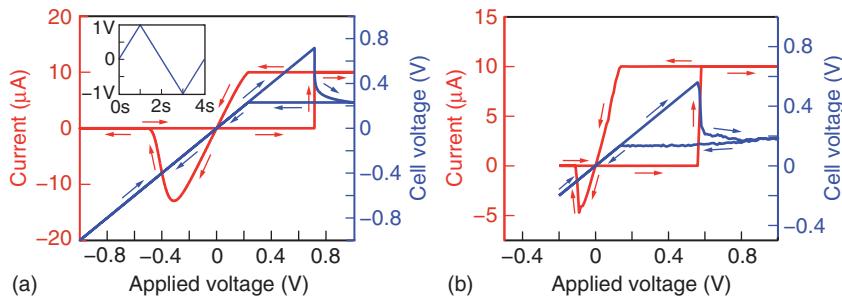


Figure 11.14 (a) Simulated I - V characteristic (red) and corresponding relation between the cell voltage and the applied voltage (blue). The applied voltage as a function of

time is shown as an inset. (b) Experimental I - V characteristic for a Cu/SiO₂/Pt cell. (Reproduced with permission from Ref. [6]. © 2011, American Institute of Physics.)

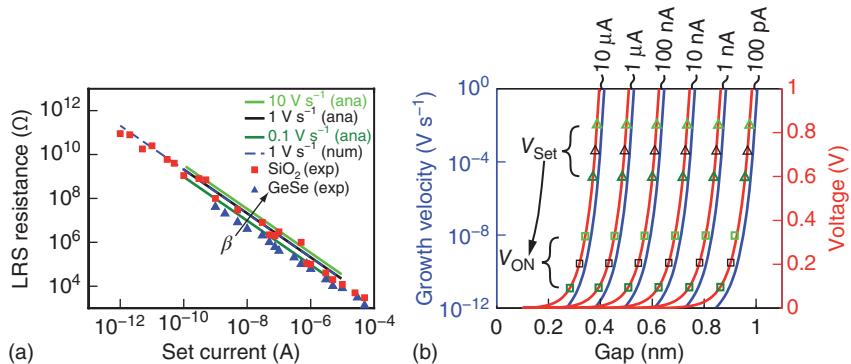


Figure 11.15 (a) Calculated resistance of the low-resistance state (LRS) as a function of set current using the analytical model for sweep rates $\beta = 0.1 \text{ V s}^{-1}$ (dark green), 1 V s^{-1} (black), and 10 V s^{-1} (light green). For comparison, the experimental data for a Cu/SiO₂-based ECM cell and an Ag/GeSe-based ECM cell are shown as red squares and blue triangles, respectively. Here, only a subset of the experimental data comparable

to the simulation range is shown. (b) Voltage (red) and growth velocity as a function of gap under current control for different current compliances. In addition, the set voltages V_{set} (triangles) and voltages V_C (squares) are shown for different sweep rates using the same color code as in (a). (From Ref. [7]. Reproduced by permission of The Royal Society of Chemistry.)

on the tunneling gap. When the gap is further reduced, the tunneling voltage needs to decrease strongly to maintain a constant current. In addition, the switching kinetics depends exponentially on the cell voltage. Thus, the growth velocity depends double-logarithmically on the tunneling gap, which explains the good controllability of the low-resistance state (LRS). The dependence of V_C on the sweep rate is also illustrated in Figure 11.15b. For faster sweep rates, the set voltage is higher. But during current control, the voltage drop during control is $V_{\text{set}} - V_C \approx 0.5 \text{ V}$ for all sweep rates and only depends on the kinetics during the current control. Thus, V_C is higher for increasing sweep rates.

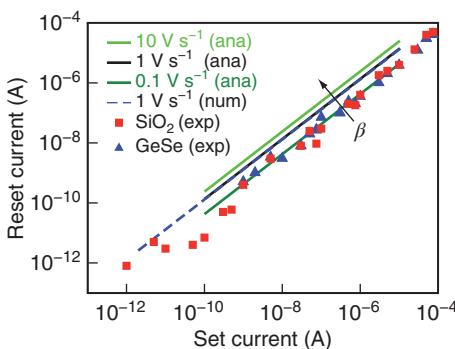


Figure 11.16 Calculated reset as a function of set current for sweep rates $\beta = 0.1 \text{ V s}^{-1}$ (dark green), 1 V s^{-1} (black), and 10 V s^{-1} (light green). For comparison, the experimental data for a Cu/SiO₂-based ECM cell and an Ag/GeSe-based ECM cell are shown

as red squares and blue triangles, respectively. Here, only a subset of the experimental data comparable to the simulation range is shown. (From [7]. Reproduced by permission of The Royal Society of Chemistry.)

For the analysis of the empirical reset law, an analytical expression for the reset current based on the described model has been derived [7] using the Lambert W function $W(\cdot)$:

$$I_{\text{RESET}} = \frac{V_{\text{RESET}}}{V_C} \exp \left(-W^{-1} \left(\frac{\alpha(1-\alpha)\beta \frac{ze}{k_B T}}{\frac{4\pi}{h} \sqrt{2m_{\text{eff}} \Delta W_0} \frac{M_{\text{Me}}}{zep_{\text{m,Me}}} j_0 \left(\frac{A_{\text{ac}}}{A_{\text{fil}}} \right)^{1-\alpha}} \right) \right) I_C. \quad (11.16)$$

It depends linearly on the compliance current as suggested by the empirical reset law. As previously explained, $V_C = \text{const}$ holds. Since the reset switching kinetics is also highly nonlinear, the reset voltage is almost constant for one specific sweep rate β . Thus, the reset current only depends on the current compliance and the sweep rate. Figure 11.16 illustrates this relation. The experimental data is reproduced very well by the simulation model and the analytically derived equation. With this model, the set/reset current asymmetry can be explained by a geometric and electrochemical asymmetry, thus favoring the oxidation reaction over the reduction [7].

In summary, it could be shown that the empirical set and reset law can be captured by a variable-gap model. The origin of these universal switching characteristics is the nonlinear switching kinetics of the ReRAM device. The model has been developed for ECM cells. The results, however, can be generalized to other ReRAM cells that can be described by a variable-gap model as long as the switching kinetics are nonlinear, for example, to the switching mechanism for oxide-based ReRAM described by Meyer *et al.* [50]. In the latter case, the device switches over the whole area rather than filamentary. But it has been demonstrated that the generic switching properties in the variable-gap model of Menzel *et al.* are independent of the filament diameter [6]. Thus, the conclusions are also valid for the area switching in oxide-based ReRAM.

11.5

Coexistence of Variable-Gap/Variable-Diameter States

Both the variable-diameter model and the variable-gap model can explain the universal switching characteristics. As shown in Figure 11.4, the two models may apply to different ranges of the set-state resistance, which suggests that there might be a transition between one switching mode and the other, depending on the different set/reset conditions.

The applicability of variable-diameter model and the variable-gap model can be assessed from numerical simulations and experimentally discriminated by electrical measurements. A numerical model for set/reset in oxide-based VCM was developed based on the migration of ionized defects, such as oxygen vacancies and metallic impurities [26]. In the model, migration is described through temperature-accelerated drift and diffusion processes, where the corresponding mobility and diffusivity are linked by the Einstein rule and are both thermally activated with the activation energy of 1 eV [26]. Figure 11.17a shows the contour plot of the calculated concentration of defects in the filament for an initial reset state, a set state with $I_C = 200 \mu\text{A}$, and a set state with a higher $I_C = 500 \mu\text{A}$. The initial reset state was obtained by simulating reset transition on an ideal cylindrical CF with a fixed diameter of 10 nm. Figure 11.17b shows similar results

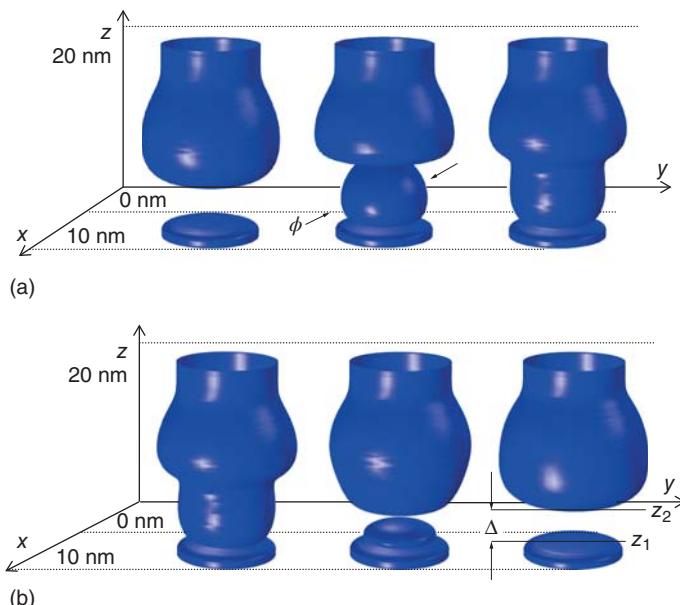


Figure 11.17 Contour plot of the defect concentration obtained from numerical simulations of set states at increasing compliance current (a) and reset states at increasing stop voltage (b). The filament shape

in set states can be described as continuous with increasing diameter ϕ , while reset states show a depleted gap with length Δ . (Reprinted with permission from [52]. © 2014, IEEE.)

for an initial set state, a reset states obtained at a voltage $V_{\text{stop}} = 0.7 \text{ V}$, and a reset state obtained at $V_{\text{stop}} = 1 \text{ V}$. Set states indicate an increase in the defect concentration in the whole gap, which can be described by an increase in the effective diameter ϕ within the gap region for increasing I_C . Reset states, instead, show a gradual opening of a gap during the reset transition, where the gap length Δ is controlled by the maximum voltage V_{stop} in the reset process [51].

The different shapes of the CF in set and reset states can be understood as follows: the reset transition takes place on an initially continuous CF, where migration results first in a small depletion usually in correspondence to the higher temperature along the filament. Once the gap is formed, Joule heating and the electric field both will concentrate at the gap. Therefore, migration will take place from the bottom positive-potential boundary of the gap toward the top boundary, causing depletion at the bottom side and accumulation at the top side. Defect migration from the bottom boundary to the top boundary thus results in a regular, gradual increase in the gap length Δ , as shown in Figure 11.17b. On the other hand, set transition takes place from the reservoir at the top boundary toward the depleted gap, causing an increase in the defect concentration in the gap, which can be described by an increasing effective ϕ .

The different shapes of the CF in set and reset states were confirmed experimentally by measuring the reset voltage V_{reset} , namely the minimum voltage in correspondence to the onset of the reset transition. Figure 11.18a shows the measured and calculated V_{reset} as a function of R for set states and reset states. Set states were obtained by the set process, while reset states were obtained from ReRAM cells with an initial resistance $R = 0.4$ and $1 \text{ k}\Omega$ in the set state. The resistance was increased by decreasing I_C in set states, while R was increased by increasing V_{stop} in reset states. Completely different behaviors are shown by V_{reset} in set and reset states in the figure, namely V_{reset} is approximately constant for set states, while it increases steeply in reset states. The constant V_{reset} for set states can be understood by the uniform CF shape, causing the voltage drop to be uniform as shown in Figure 11.18b. Irrespective of the filament diameter, the electric field and the maximum temperature are always the same; therefore, the voltage needed to trigger the driving force for migration does not depend on R . On the other hand, V_{reset} increases for increasing R in reset states because the voltage drops almost completely across the gap, where the field and the Joule heating are thus maximized (see Figure 11.18c). The field and heating are therefore decreased in the conductive regions of the CF, which are active from the viewpoint of migration. These data therefore allow for a direct discrimination between ReRAM programmed states on the basis of the CF shape.

In the variable-gap model, the current compliance controls the remaining gap. For very high current compliances, however, the gap closes completely [53]. This transition coincides with a resistance equivalent to one atomic contact resistance $R = 12.9 \text{ k}\Omega$ in thin metal wires. In fact, quantized conduction steps have been observed for $R \leq 12.9 \text{ k}\Omega$ [54, 55]. The resistance drop is then related to an increasing number of atoms creating a galvanic contact between the filament tip and the counter electrode. This can be regarded as an increase in filament diameter.

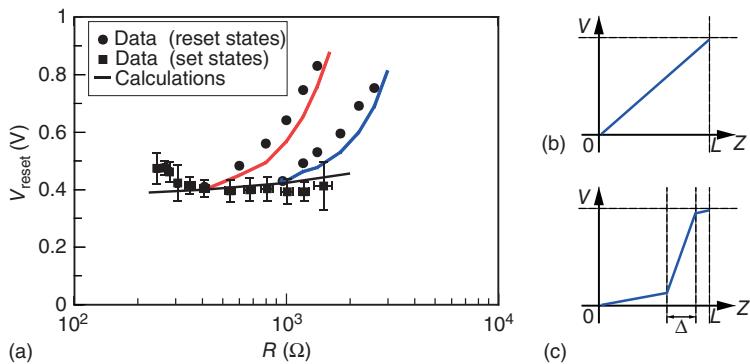


Figure 11.18 Measured and calculated V_{reset} as a function of R for set states obtained at variable I_C and reset states obtained at variable V_{stop} for two initial set states at $R=0.4$ and $1\text{k}\Omega$. (Reprinted with permission from [26]. © 2012, IEEE.)

In this regime, the reset mechanism changes to a lateral, thermally assisted self-dissolution of the conducting filament [53], that is, a variable-diameter mode. This process has been simulated using an axisymmetric 2D model [53]. The simulation results reveal that for both positive and negative polarities, a lateral dissolution sets in. The simulated I - V characteristics for a positive voltage sweep, that is, the nominal set polarity for bipolar operation, are shown in Figure 11.19a, and the corresponding time evolution of filament dissolution is shown in Figure 11.19b. First, the filament remains intact and the resistance increases due to the temperature increase. When the voltage is higher than 0.5 V , the filament starts to dissolve (C) instead of growing further. This has been explained by a thermally assisted self-dissolution of the filament. The redox reactions, that is, oxidation

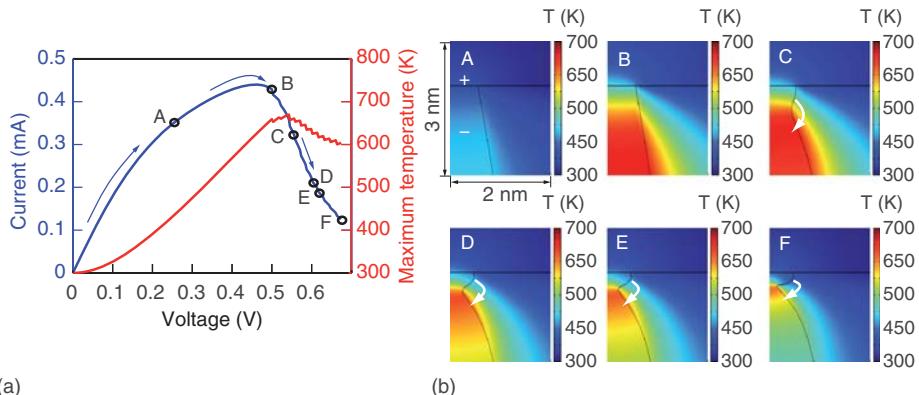


Figure 11.19 (a) Simulated I - V (blue) and T - V characteristic (red) using the multidimensional model. A positive polarity is applied to the top electrode. (b) Corresponding

temperature and shape evolution of the filament dissolution. (Reproduced with permission from [53]. © 2013, IEEE.)

and reduction, are highly temperature-dependent and, thus, will most probably be close to the hottest spot of the filament. In addition, the potential drop is highest at the constriction of the filament, which will also be the location where most Joule heat is generated. This combination results in an electrochemical dissolution and a succeeding redeposition both taking place along the filament close to the hottest spot (C–E) until the filament ruptures (F). Due to the thermal nature of the self-dissolution, the filament dissolution can occur for both positive and negative voltage polarities. A reset with the original set polarity, however, can lead to unstable current fluctuations [53].

11.6 Summary

In conclusion, filamentary-type ReRAM characteristics display several universal features, irrespective of the switching behavior (unipolar or bipolar type) and active material (binary oxides, chalcogenide-based electrolytes in ECMs). The origin of the universal switching characteristics has been discussed in the chapter, and the main conclusions can be summarized as follows:

- 1) For ReRAM cells, in general, the empirical law $R = V_C/I_C$ has its origin in the highly nonlinear switching kinetics that leads to a strong decrease in the driving force of resistance change under current control. As a result, the voltage across the device at the end of the set transition for a given pulse time is $V_C \approx \text{const}$, where V_C does not show a marked dependence on the operation current, on the active material, and on the switching mechanism (unipolar, bipolar, etc.). The condition $V_C \approx \text{const}$ reflects a voltage-driven switching kinetics.
- 2) The relation $I_{\text{reset}} \propto I_C$ results from the voltage-driven switching kinetics, which is in turn due to the strongly nonlinear rate of the microscopic switching mechanism. For instance, since ion migration rate is driven by voltage, both filament growth during set and filament dissolution during reset take place at similar (or proportional) voltages $V_C \propto V_{\text{reset}}$, thus leading to $I_{\text{reset}} \propto I_C$ for symmetric I - V curves.
- 3) The nonlinear switching kinetics appears as a strong dependence of set/reset times on the applied voltage, which is generally observed in all ReRAM devices.
- 4) The universal switching characteristics can be explained by analytical models where ReRAM resistance is controlled either by the filament diameter or by the filament gap. This chapter discusses the conditions where these models are more appropriate, according to the resistance value and the electrical process to obtain the ReRAM state, namely set or reset operation.

Acknowledgment

SM would like to acknowledge the support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

1. Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833–840.
2. Kwon, D.-H., Kim, K.M., Jang, J.H., Jeon, J.M., Lee, M.H., Kim, G.H., Li, X.-S., Park, G.-S., Lee, B., Han, S., Kim, M., and Hwang, C.S. (2010) Atomic structure of conducting nanofilaments in TiO_2 resistive switching memory. *Nat. Nanotechnol.*, **5**, 148–153.
3. Sawa, A. (2008) Resistive switching in transition metal oxides. *Mater. Today*, **11**, 28–36.
4. Ielmini, D., Nardi, F., and Cagli, C. (2011) Universal reset characteristics of unipolar and bipolar metal-oxide RRAM. *IEEE Trans. Electron Devices*, **58**, 3246–3253.
5. Ielmini, D. (2011) Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Devices*, **58**, 4309–4317.
6. Menzel, S., Böttger, U., and Waser, R. (2012) Simulation of multilevel switching in electrochemical metallization memory cells. *J. Appl. Phys.*, **111**, 014501/1–014501/5.
7. Menzel, S. and Waser, R. (2013) Analytical analysis of the generic SET and RESET characteristics of electrochemical metallization memory cells. *Nanoscale*, **22**, 11003–11010.
8. Lee, H.Y., Chen, P.S., Wu, T.Y., Chen, Y.S., Wang, C.C., Tzeng, P.J., Lin, C.H., Chen, F., Lien, C.H., and Tsai, M.J. (2008) Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM. IEEE International Electron Devices Meeting 2008, Technical Digest, pp. 297–300.
9. Kinoshita, K., Tsunoda, K., Sato, Y., Noshiro, H., Yagaki, S., Aoki, M., and Sugiyama, Y. (2008) Reduction in the reset current in a resistive random access memory consisting of NiO_x brought about by reducing a parasitic capacitance. *Appl. Phys. Lett.*, **93**, 033506.
10. Kund, M., Beitel, G., Pinnow, C.U., Roehr, T., Schumann, J., Symanczyk, R., Ufert, K.D., and Mueller, G. (2005) Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20 nm. IEDM Technical Digest, pp. 754–757.
11. Schindler, C., Meier, M., Waser, R., and Kozicki, M.N. (2007) Resistive switching in Ag-Ge-Se with extremely low write currents. Proceedings of NVMTS, pp. 82–85.
12. Tsunoda, K., Fukuzumi, Y., Jameson, J.R., Wang, Z., Griffin, P.B., and Nishi, Y. (2007) Bipolar resistive switching in polycrystalline TiO_2 films. *Appl. Phys. Lett.*, **90**, 113501/1–113501/3.
13. Nardi, F., Ielmini, D., Cagli, C., Spiga, S., Fanciulli, M., Goux, L., and Wouters, D.J. (2011) Control of filament size and reduction of reset current below $10\ \mu\text{A}$ in NiO resistance switching memories. *Solid-State Electron.*, **58**, 42–47.
14. Nardi, F., Larentis, S., Balatti, S., Gilmer, D., and Ielmini, D. (2012) Resistive switching by voltage-driven ion migration in bipolar RRAM—Part I: experimental study. *IEEE Trans. Electron Devices*, **59**, 2461–2467.
15. Ielmini, D., Nardi, F., and Balatti, S. (2012) Evidence for voltage-driven set/reset processes. *IEEE Trans. Electron Devices*, **59**, 2049–2055.
16. Park, J., Jung, S., Lee, J., Lee, W., Kim, S., Shin, J., and Hwang, H. (2011) Resistive switching characteristics of ultra-thin $\text{TiO}(x)$. *Microelectron. Eng.*, **88**, 1136–1139.
17. Lee, J., Shin, J., Lee, D., Lee, W., Jung, S., Jo, M., Park, J., Biju, K., Kim, S., Park, S., and Hwang, H. (2010) Diode-less nano-scale $\text{ZrO}_x/\text{HfO}_x$ RRAM device with excellent switching uniformity and reliability for high-density cross-point memory applications. IEDM Technical Digest, pp. 451–453.
18. Schindler, C. (2009) Resistive Switching in Electrochemical Metallization Memory Cells, RWTH Aachen.
19. Ielmini, D. (2011) Filamentary-switching model in RRAM for time, energy and scaling projections. 2011 IEEE International Electron Devices Meeting – IEDM '11, pp. 409–412.

20. Waser, R., Bruchhaus, R., and Menzel, S. (2012) *Nanoelectronics and Information Technology*, 3rd edn, Wiley-VCH Verlag GmbH.
21. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011) Origin of the ultra-nonlinear switching kinetics in oxide-based resistive switches. *Adv. Funct. Mater.*, **21**, 4487–4492.
22. Nishi, Y., Schmelzer, S., Böttger, U., and Waser, R. (2013) Weibull analysis of the kinetics of resistive switches based on tantalum oxide thin films. Proceedings of the 43rd European Solid-State Device Research Conference (ESSDERC), pp. 174–177.
23. Yu, S., Wu, Y., and Wong, H. (2011) Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory. *Appl. Phys. Lett.*, **98**, 103514/1–103514/3.
24. Alibart, F., Gao, L., Hoskins, B.D., and Strukov, D.B. (2012) High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology*, **23**, 75201/1–75201/7.
25. Menzel, S., Tappertzhofen, S., Waser, R., and Valov, I. (2013) Switching kinetics of electrochemical metallization memory cells. *Phys. Chem. Chem. Phys.*, **15**, 6945–6952.
26. Larentis, S., Nardi, F., Balatti, S., Gilmer, D.C., and Ielmini, D. (2012) Resistive switching by voltage-driven ion migration in bipolar RRAM-Part II: modeling. *IEEE Trans. Electron Devices*, **59**, 2468–2475.
27. Koveshnikov, S., Matthews, K., Min, K., Gilmer, D., Sung, M., Deora, S., Li, H., Gausepohl, S., Kirsch, P., and Jammy, R. (2012) Real-time study of switching kinetics in integrated 1 T/ HfO_x 1R RRAM: intrinsic tunability of set/reset voltage and trade-off with switching time. Technical Digest – International Electron Devices Meeting, IEDM, pp. 20.4.1–20.4.3.
28. Diokh, T., Le-Roux, E., Jeannot, S., Gros-Jean, M., Candelier, P., Nodin, J.F., Jousseaume, V., Perniola, L., Grampeix, H., Cabout, T., Jalaguier, E., Guillermet, M., and De Salvo, B. (2013) Investigation of the impact of the oxide thickness and RESET conditions on disturb in HfO_x 2-RRAM integrated in a 65 nm CMOS technology. 2013 IEEE International Reliability Physics Symposium (IRPS), pp. 5E.4.1–5E.4.4.
29. Cao, M.G., Chen, Y.S., Sun, J.R., Shang, D.S., Liu, L.F., Kang, J.F., and Shen, B.G. (2012) Nonlinear dependence of set time on pulse voltage caused by thermal accelerated breakdown in the Ti/HfO₂/Pt resistive switching devices. *Appl. Phys. Lett.*, **101**, 203502/1.
30. Fleck, K., Böttger, U., Waser, R., and Menzel, S. (2014) Interrelation of sweep and pulse analysis of the SET process in SrTiO₃ resistive switching memories. *IEEE Electron Device Lett.*, **35**, 924–926.
31. Valov, I., Waser, R., Jameson, J.R., and Kozicki, M.N. (2011) Electrochemical metallization memories-fundamentals, applications, prospects. *Nanotechnology*, **22**, 254003/1–254003/22.
32. Nayak, A., Tamura, T., Tsuruoka, T., Terabe, K., Hosaka, S., Hasegawa, T., and Aono, M. (2010) Rate-limiting processes determining the switching time in a Ag₂S atomic switch. *J. Phys. Chem. Lett.*, **1**, 604–608.
33. Nayak, A., Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) Switching kinetics of a Cu₂S-based gap-type atomic switch. *Nanotechnology*, **22**, 235201/1–235201/7.
34. Palma, G., Vianello, E., Molas, G., Cagli, C., Longnos, F., Guy, J., Reyboz, M., Carabasse, C., Bernard, M., Dahmani, F., Bretegnier, D., Liebault, J., and De Salvo, B. (2013) Effect of the active layer thickness and temperature on the switching kinetics of GeS₂-based conductive bridge memories. *Jpn. J. Appl. Phys.*, **52**, UNSP 04CD02/1.
35. Jameson, J.R., Gilbert, N., Koushan, F., Saenz, J., Wang, J., Hollmer, S., and Kozicki, M.N. (2011) One-dimensional model of the programming kinetics of conductive-bridge memory cells. *Appl. Phys. Lett.*, **99**, 063506.
36. Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A.L., and Kozicki, M.N. (2009) Study of multilevel programming in programmable metallization cell (PMC) memory. *IEEE Trans. Electron Devices*, **56**, 1040–1047.

37. Menzel, M., Salinga, U., Böttger, and M. Wimmer, Physics of the switching kinetics in resistive memories, *Adv. Funct. Mater.* (2015) DOI: 10.1002/adfm.201500825.
38. Fuchs, K. (1938) The conductivity of thin metallic films according to the electron theory of metals. *Proc. Cambridge Philos. Soc.*, **34**, 100–108.
39. Sondheimer, E.H. (1952) The mean free path of electrons in metals. *Adv. Phys.*, **1**, 1–42.
40. Simmons, J.G. (1963) Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.*, **34**, 1793–1803.
41. Balatti, S., Larentis, S., Gilmer, D.C., and Ielmini, D. (2013) Multiple memory states in resistive switching devices through controlled size and orientation of the conductive filament. *Adv. Mater.*, **25**, 1474–1478.
42. Mott, N.F. and Gurney, R.W. (1940) *Electronic Processes in Ionic Crystals*, Clarendon Press, Oxford.
43. Yu, S. and Wong, H.-S. (2011) Compact modeling of conducting-bridge random-access memory (CBRAM). *IEEE Trans. Electron Devices*, **58**, 1352–1360.
44. Ielmini, D. (2008) Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses. *Phys. Rev. B*, **78**, 035308.
45. Tsunoda, K., Kinoshita, K., Noshiro, H., Yarnazaki, Y., Lizuka, T., Ito, Y., Takahashi, A., Okano, A., Sato, Y., Fukano, T., Aoki, M., and Sugiyama, Y. (2007) Low power and high speed switching of Ti-doped NiO ReRAM under the unipolar voltage source of less than 3 V. 2007 IEEE International Electron Devices Meeting, pp. 767–770.
46. Ambrogio, S., Balatti, S., Choi, S., and Ielmini, D. (2014) Impact of the mechanical stress on switching characteristics of electrochemical resistive memory. *Adv. Mater.*, **26**, 3885–3892.
47. Ielmini, D., Nardi, F., and Cagli, C. (2011) Physical models of size-dependent nanofilament formation and rupture in NiO resistive switching memories. *Nanotechnology*, **22**, 254022/1–254022/12.
48. Yang, J.J., Strukov, D.B., and Stewart, D.R. (2013) Memristive devices for computing. *Nat. Nanotechnol.*, **8**, 13–24.
49. Linn, E., Menzel, S., Ferch, S., and Waser, R. (2013) Compact modeling of CRS devices based on ECM cells for memory, logic and neuromorphic applications. *Nanotechnology*, **24**, 384008.
50. Meyer, R., Schloss, L., Brewer, J., Lambertson, R., Kinney, W., Sanchez, J., and Rinerson, D. (2008) Oxide dual-layer memory element for scalable non-volatile cross-point memory technology. Proceedings NVMTS, pp. 54–58.
51. Ielmini, D., Balatti, S., and Larentis, S. (2013) Filament evolution during Set and Reset transitions in oxide resistive switching memory. *Jpn. J. Appl. Phys.*, **52**, UNSP 04CD10/1.
52. Ambrogio, S., Balatti, S., Cubeta, A., Calderoni, A., Ramaswamy, N., and Ielmini, D. (2014) Statistical fluctuations in HfO₂ resistive-switching memory: part I – Set/reset variability. *IEEE Trans. Electron Devices*, **61**, 2912–2919.
53. Menzel, S., Adler, N., van den Hurk, J., Tappertzhofen, S., Valov, I., and Waser, R. (2013) Simulation of polarity independent RESET in electrochemical metallization memory cells. 2013 5th IEEE International Memory Workshop (IMW), pp. 92–95.
54. Tappertzhofen, S., Valov, I., and Waser, R. (2012) Quantum conductance of AgI based resistive switches: towards an atomic scale memory. Nature Conferences, pp. 208–209.
55. Tsuruoka, T., Hasegawa, T., Terabe, K., and Aono, M. (2012) Conductance quantization and synaptic behavior in a Ta₂O₅-based atomic switch. *Nanotechnology*, **23**, 435705.

12

Quasistatic and Pulse Measuring Techniques

Antonio Torrezan, Gilberto Medeiros-Ribeiro, and Stephan Tiedke

12.1

Brief Introduction to Electronic Transport Testing of ReRAM

One of the most interesting aspects of resistive random-access memory (ReRAM) is the fact that the device characteristics are never static. In essence, unlike transistors, diodes, resistors, and the majority of electronic devices, ReRAM (see Chapter 1 for definitions) devices change their intrinsic properties under external stimuli. These changes are dependent upon the amplitude of the stimuli and the associated timescale. This fact poses several interesting questions – what is the device impedance? Can one develop a SPICE model for it? Are the changes permanent? How can one model a device in order to use it in a circuit?

Over the past several years, a number of studies addressed ReRAM properties primarily from a quasistatic characteristic point of view, utilizing commercial, off-the-shelf, semiconductor parameter analyzers (SPAs). Critical time-dependent phenomena are severely overlooked when characterizing these devices from this somewhat simplistic point of view.

In fact, the precise nature of the switching process has been under debate, where the resistance change could arise from a diversity of mechanisms: Schottky barrier height changes [1], current limiting due to trap-controlled space charge [2], diffusion of oxygen vacancies and associated bulk and interface contributions [3], and make and break of metallic filaments [4], to name a few. The aforementioned experiments illustrate that a thorough characterization to assess switching mechanisms cannot rely solely on the *I-V* characteristics obtained from standard SPAs. Despite the multiplicity of proposals for mechanisms, the common issue that those experiments address is the time structure.

Traditionally, impedance and admittance spectroscopy have been employed to study defects and trapping phenomena [5] of deep-level states in semiconductors. Similarly to a Debye relaxation process in dielectrics [6], the key ingredient of admittance/impedance spectroscopy is the measurement of real and imaginary parts of the dielectric constant, and from those, one can derive the characteristic times that can be ascribed to electronic, ionic, dielectric polarization, or any other

process that creates an electric dipole in the system. To arrive at an interpretation that unequivocally ascribes the actual mechanisms for switching is a challenge. Yet, Menke *et al.* [3] were able to develop a microscopic form for different resistance states and nature of the conducting channels.

From a different perspective, Yu *et al.* [4] investigated the noise spectra for low- and high-resistance states (HRSs) of HfO_2 devices. For this particular case, the noise spectra revealed changes in the electronic structure of the conductance channels rather than on the mechanism itself. Nevertheless, the possibility of using noise spectra to understand the transport mechanisms of ReRAM devices adds the time degree of freedom to the experimentalist and device modeling scientist, similarly to dielectric spectroscopy.

After an overview of quasistatic measurement for *I-V* characteristics, here we address two different strategies for coping with device dynamics. The first one targets control of the runaway current during switching by means of current compliance. This particular strategy has the benefit that most SPAs have that as an option, but one can also implement current compliance by using field-effect transistors (FETs), which in and of itself embodies a 1T-1R type cell. The second strategy is simply to stimulate and measure the device response as fast as the state-of-the-art equipment allows us. This particular strategy does not impose any *a priori* constraint over the stimulus; however, it does constrain the time frame for the dynamics evolution. For this particular case, the resistance change at the subnanosecond regime can provide invaluable, real-time insights into the device evolution.

12.2

Quasistatic Measurement of Current–Voltage Characteristics

The basic functionality of a resistive-switching random-access memory cell can be determined from its current–voltage characteristics, also referred as an *I-V* curve, which can be obtained in a quasistatic fashion. In a quasistatic characterization, also referred as quasi-DC (direct current), the voltage applied to the device is slowly swept while monitoring the current through the cell. Alternatively, the current applied to the device can be swept while the voltage across it is measured. The apparatus topology for a quasistatic measurement is shown in Figure 12.1a and is found commercially in the form of SMUs (Source Measurement Units) with a range of current, voltage, and power capabilities available [7]. SMUs can also be utilized as stand-alone current and voltage sources as well as a stand-alone ammeter, voltmeter, ohmmeter, and electronic load.

The quasistatic characterization can be performed in a two-wire or a four-wire configuration as indicated in Figure 12.1b,c, respectively. In the former case, the source voltage applied to the device is recorded in the *I-V* curve and may differ from the voltage across the resistive cell depending on how the ReRAM cell resistance compares with the resistance of the interconnect R_{interc} linking the source to the device. In a four-point probe measurement, also known as Kelvin

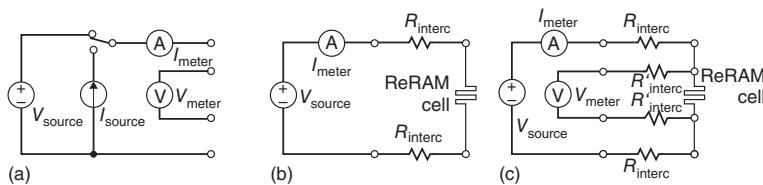


Figure 12.1 (a) Basic topology of an SMU. (b) Configuration of a two-point probe measurement and (c) a four-point probe measurement driven by a voltage source. R_{interc} and R'_{interc} denote interconnect resistances.

sensing, an additional pair of wires connected to the cell bottom electrode (BE) and top electrode (TE) enables the device voltage to be measured without the effect of the interconnect resistance as almost no current flows in the voltage sensing pair.

Resistive memory cells exhibit a pinched hysteresis loop confined to the first and third quadrants of the I - V plane with nonvolatile resistance states [8] referred to as the low-resistance state (LRS) or ON state and the HRS or OFF state. Basic switching parameters can be obtained from the I - V characteristics illustrated in Figure 12.2. A switching event from HRS to LRS is called a SET process or ON switching and is characterized by a SET voltage V_{SET} . Conversely, the reverse switching from LRS to HRS is called a RESET process or OFF switching and is identified by a RESET voltage V_{RESET} . In the cases where the state transition is gradual as seen on an I - V curve, V_{SET} (V_{RESET}) may be specified at a certain percentage decrease (increase) in resistance [9].

The switching types for resistive memory can be categorized into two modes: bipolar switching and unipolar switching. In bipolar switching, the SET process occurs at one polarity while the RESET process occurs at the opposite polarity as indicated in Figure 12.2a. In unipolar switching, SET and RESET processes are independent of the polarity of the switching voltage as depicted in Figure 12.2b. Experimental results for unipolar devices typically show RESET occurring at a lower voltage and higher current than the SET process under controlled switching current [10, 11]. For each switching type, current compliance (CC) can be utilized

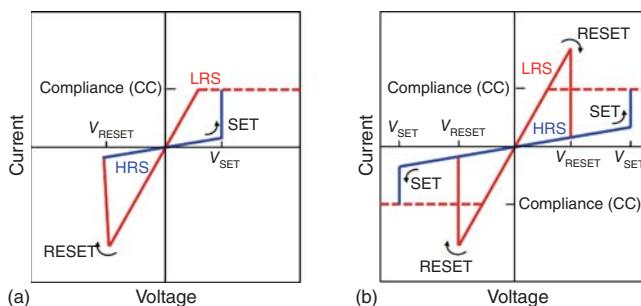


Figure 12.2 Generic representation of an I - V curve for (a) bipolar switching and (b) unipolar switching. Experimental I - V curves may deviate considerably from these illustrations.

during ON switching to prevent irreversible cell breakdown and to control the LRS for multilevel cell (MLC) operation [12]. Current compliance capability may be provided by SMUs in commercial SPAs, and the effectiveness of this and other approaches to constrain the device current during switching will be discussed later in this chapter. Experimental I - V characteristics may deviate considerably from the representations in Figure 12.2.

ReRAM cells may require an initial electroforming step in which the as-fabricated virgin device is conditioned to a repeatable resistance switching by applying a forming voltage or forming current higher than the ones required for subsequent switching. For bipolar devices with symmetric material stack, the voltage polarity of the electroforming step usually determines the switching polarity of the SET and RESET processes. On the contrary, asymmetric bipolar devices have their switching polarity typically determined by the asymmetry of the material stack instead of the polarity of the electroforming voltage [13]. Figure 12.3 shows the forming process of a bipolar TiO_x -based ReRAM device with switching polarity defined by its asymmetric material stack. Applying an electroforming step with the SET polarity formed the device in the LRS (Figure 12.3a), while forming the device with the opposite polarity leads to an initial HRS (Figure 12.3b) [14]. For both cases, the magnitude of the forming voltage was higher than the required V_{SET} or V_{RESET} for the next switching cycle.

Besides the switching and forming parameters, other information can be obtained from quasistatic I - V curves such as the nonlinearity of the LRS and the HRS. In cross-point memory arrays without selectors, nonlinearity plays a role in mitigating sneak path leakage currents and in determining the maximum array size, which is also a function of other parameters including device state resistance, the resistance ratio between states, and the interconnect resistance between memory cells [15].

Fitting the I - V characteristics below the switching voltage combined with other characterization techniques such as varying temperature has also been used to identify possible conduction mechanisms for the LRS and HRS. Candidates

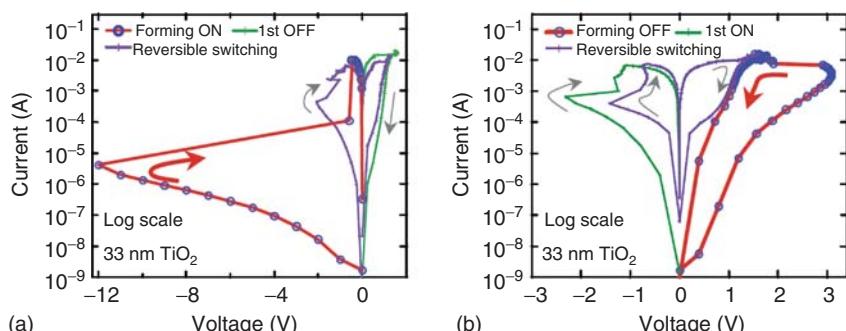


Figure 12.3 Forming of an asymmetric bipolar TiO_x -based ReRAM cell in the (a) LRS and in the (b) HRS. (Reproduced with permission from Ref. [14].)

for conduction mechanisms include ohmic conduction, Schottky emission, Poole–Frenkel emission, space-charge-limited conduction, Fowler–Nordheim tunneling, direct tunneling, among others [16, 17].

12.2.1

Dependence of Switching Parameters on Sweep Rate

The switching voltages V_{SET} and V_{RESET} seen in I - V curves have a dependence on the sweep rate at which the data is obtained. This dependence has been observed for fast sweep rates as well as slow sweep rates in quasistatic measurements. Figure 12.4a shows the measured I - V characteristics of HfO_x -based resistive memory devices with V_{SET} and V_{RESET} increasing as the sweep rate is increased [9]. Similar trend has been verified for V_{SET} in $\text{Cu}-\text{SiO}_2$ -based resistive cells as seen in Figure 12.4b for quasistatic sweep rates [18]. As the sweep rate increases, the resistive cell has less time to switch and does so at a higher voltage as observed experimentally.

The dependence of the apparent switching voltage on the sweep rate has also been captured by device models derived from the static (nonswitching) and dynamical (switching) behavior of ReRAM devices [19, 20]. Figure 12.5 shows a simulated I - V curve for a TaO_x -based resistive cell with increasing V_{SET} and V_{RESET} for faster sweep rates. Also observed in this graph is the collapse of the pinched hysteresis loop to a single trace at the highest sweep frequency, a feature characteristic of memristors [8, 21, 22]. Even though a change in switching

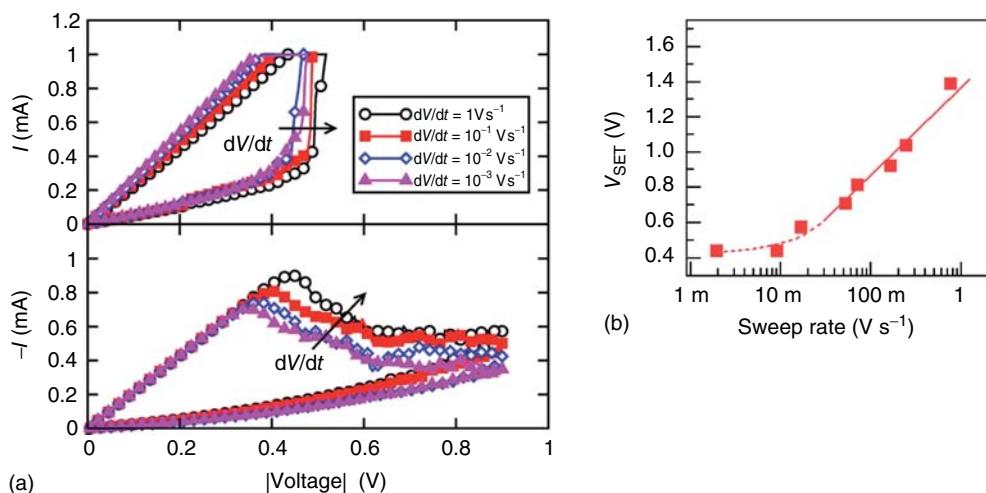


Figure 12.4 (a) Measured I - V characteristics at quasistatic sweep rates for a HfO_x -based ReRAM device turning ON (top plot) and turning OFF (bottom plot). (Reproduced with permission from Ref. [9].) (b) Switching

SET voltage V_{SET} as a function of quasistatic sweep rates for a $\text{Cu}-\text{SiO}_2$ -based ReRAM cell. (Reproduced with permission from Ref. [18].)

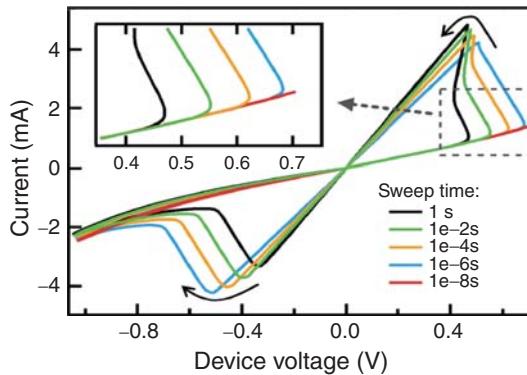


Figure 12.5 Simulated I - V curve with varying sweep rates for a TaO_x ReRAM device.
(Reproduced with permission from Ref. [19].)

voltage with sweep rate can be observed in quasistatic I - V characteristics, limited information about switching dynamics is provided by these curves even using time stamps for each measurement point.

12.3

Current Compliance and Overshoot Effects

Current compliance for the voltage source driving the resistive memory element is needed for two reasons. On the one hand, current has to be limited in order to prevent the sample from reaching irreversible breakdown during electroforming or from overswitching as exemplified in Figure 12.6. On the other hand, control of the maximum current allows tuning of the device resistance state, leading to an MLC as seen in Figure 12.7.

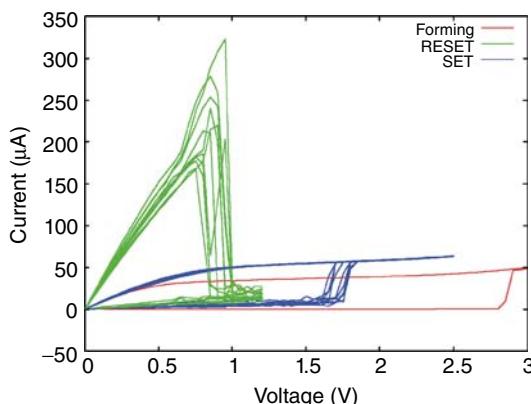


Figure 12.6 I - V characteristics for forming, SET, and RESET of a NiO_x -based unipolar ReRAM device.

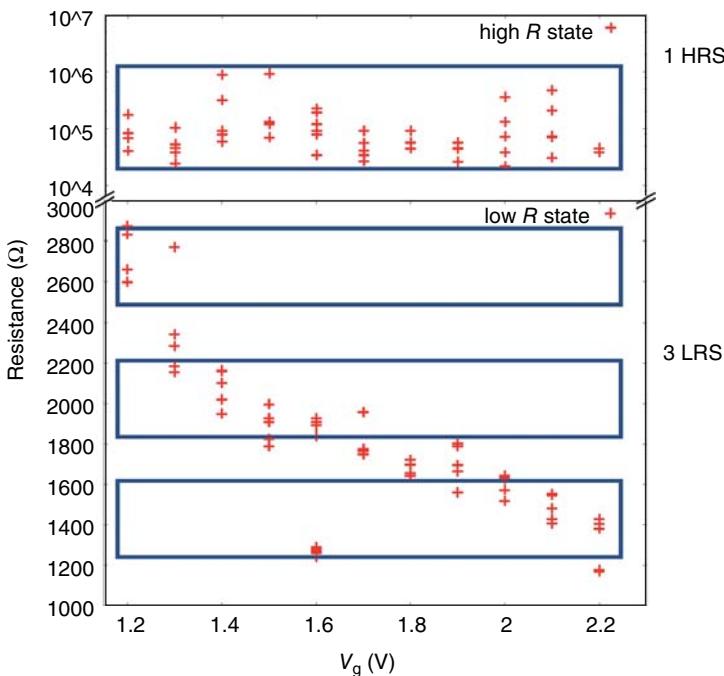


Figure 12.7 Multilevel NiO_x ReRAM cell (2 bits per cell) in a 1T-1R configuration programmed by the gate voltage V_g of a field-effect transistor providing current compliance.

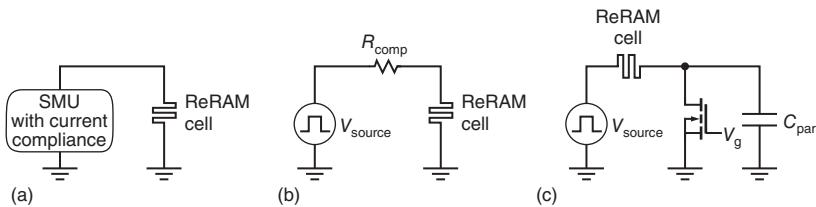


Figure 12.8 Implementations of current compliance using (a) an SMU with integrated compliance circuit, (b) a resistor R_{comp} to limit the maximum current, and (c)

a transistor controlled by a gate voltage V_g . In (c), C_{par} denotes the parasitic capacitance between the ReRAM cell and the transistor.

There are various approaches to establish current compliance for a voltage source, which in general exhibit a low internal resistance. One of these approaches is the built-in current compliance in an SMU illustrated in Figure 12.8a. This current limiter is based on an internal electronic circuit that transforms the SMU voltage source into a constant current source after sensing a current surge above the compliance value I_{comp} . The resulting voltage across the device under test (DUT) is calculated as $V_{\text{DUT}} = R_{\text{DUT}} \times I_{\text{comp}}$. Although this method is flexible in letting the user set the compliance value electronically, the response time of the compliance circuit is approximately 60 ns at best (Keithley, Germany, and Rainer

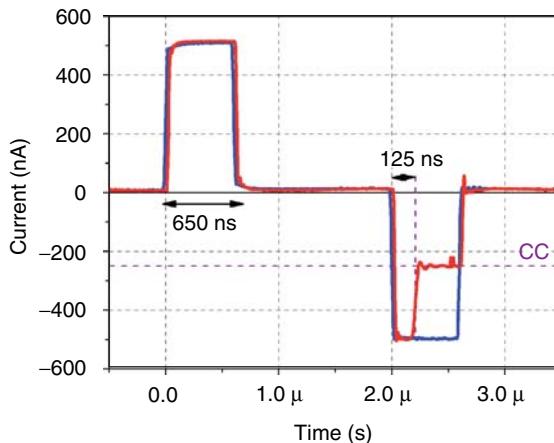


Figure 12.9 Transient device current (red curve) in a NiO_x ReRAM cell as measured by an SMU with integrated current compliance.

Waser's Group, 2011, private communication) and slightly load-dependent. During this response time, the current can significantly exceed the target compliance value. Figure 12.9 shows a typical delay in the response of an SMU with integrated current compliance. A significant current overshoot is observed at the beginning of the pulse at $2\ \mu\text{s}$ with the current compliance being enforced after a delay of $125\ \text{ns}$.

One approach to possibly circumvent the observed current overshoot is to place a compliance resistor R_{comp} in series with the memory cell as indicated in Figure 12.8b. In theory, this implementation would limit the maximum current to the value $V_{\text{source}}/R_{\text{comp}}$. However, due to the dynamical nature of resistive switching, the compliance element R_{comp} can be bypassed by the charge and discharge of parasitic capacitances existent in the test setup. Parasitic capacitance originates, for instance, from scope probes and interconnects such as cables, probe needles, printed circuit board (PCB) traces, and device electrodes, to name a few. For example, the capacitance for a $50\ \Omega$ coaxial cable is about $1\ \text{pF cm}^{-1}$. While minimizing the parasitic capacitance as much as possible is desirable, another avenue to reduce device current overshoot is to ensure that any remaining parasitic capacitance sees a low impedance outside the device current path. The plot in Figure 12.10a shows the transient device current during the forming of a TiO_x -based ReRAM cell. The observed current overshoot occurred as the parasitic capacitance C_{par} , associated mainly with a setup cable, discharged through the sample and an oscilloscope, bypassing the compliance resistor $R_{\text{comp}} = 1\ \text{k}\Omega$ [23]. Moving the compliance resistor close to the sample proved to be effective in reducing the observed overshoot as indicated in Figure 12.10b. This rearrangement not only decreased the parasitic capacitance between the device and the compliance element, but also made the main parasitic capacitance C_{par} to be in parallel with the low impedance of the voltage source. Although the compliance level could be adjusted in this implementation by using a bank of

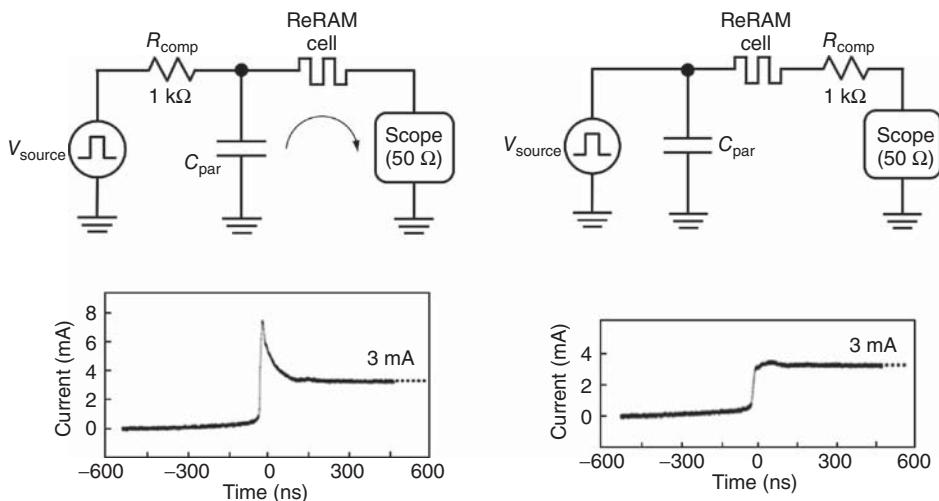


Figure 12.10 (a) Measured device current for a TiO_x ReRAM cell during electroforming. The schematic of the experimental setup using a compliance resistor R_{comp} is also shown. (b) Measured device current for a TiO_x ReRAM cell during electroforming. The

schematic of the experimental setup is also shown with the compliance resistor R_{comp} being placed close to the memory cell. (Reproduced and adapted with permission from Ref. [23].)

discrete resistors and a multiplexer, this compliance method is not as flexible as an SMU with regard to set and fine-tune current compliance levels.

Besides a resistor, another possibility to limit current is to use a transistor in series with the ReRAM sample as depicted in Figure 12.8c, embodying a 1T-1R cell. For the FET shown in Figure 12.8c, the current compliance level is adjusted by setting the transistor gate voltage V_g to a corresponding level of saturated drain current. Similarly to the case with the resistor, the effectiveness of the transistor as a compliance element relies upon minimizing the parasitic capacitance C_{par} between the ReRAM cell and the transistor, especially at low compliance levels. As the device forms or turns on, the transistor drain voltage increases, leading to a transient current that charges C_{par} and flows through the device without any control by the transistor. Kinoshita *et al.* [11] showed that the expected relationship $I_{\text{RESET}} \approx I_{\text{comp}}$ between the RESET current I_{RESET} and the compliance current I_{comp} in unipolar ReRAM devices was deviated by the transient overshoot current flowing through C_{par} during SET or forming. Figure 12.11 shows their experimental results for a NiO_x unipolar memory cell in two 1T-1R configurations that differ from each other by an extra monitor pad between the device and the transistor. This results in a higher parasitic capacitance C_{par} in configuration A compared to the one in configuration B. As observed in the graph, the higher the C_{par} , the higher the I_{RESET} required, deviating from the relationship $I_{\text{RESET}} \approx I_{\text{comp}}$, particularly at low I_{comp} . In another experiment also utilizing a NiO_x device, Nardi *et al.* [24] demonstrated that I_{RESET} could be scaled below 10 μA in an integrated 1T-1R cell with extremely low parasitic capacitance.

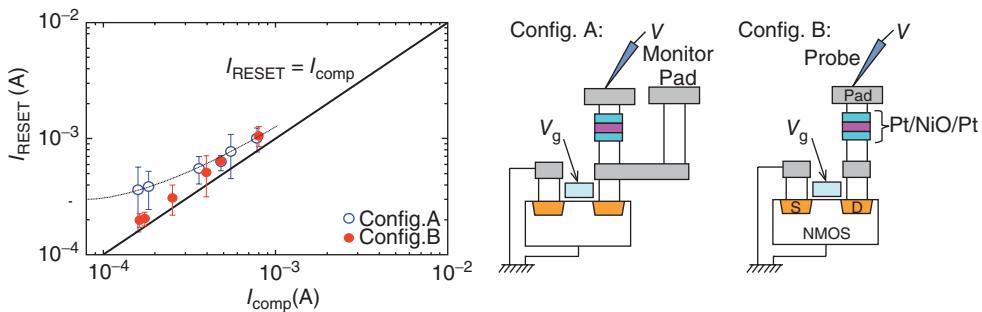


Figure 12.11 Mean RESET current I_{RESET} for 30 switching cycles as a function of the compliance current I_{comp} for a unipolar NiO_x ReRAM cell in two 1T-1R configurations. (Adapted with permission from Ref. [11].)

12.4

Pulsed Measurements for the Study of Switching Dynamics

The study of switching dynamics of resistive memory cells and the evaluation of memory metrics such as switching time and switching energy require monitoring the device voltage and current in response to a stimulus, which can take the form of a voltage pulse or a voltage step. The minimum observable timescale is dictated by the slowest part of the setup, and it could be determined by the rise and fall times of the voltage driver, the bandwidth of the measuring system to track the switching voltage and current in real time, or intrinsic RC delays from parasitic resistance and capacitance associated with the setup interconnect and the device structure.

The basic setup that has been utilized to observe ReRAM switching dynamics is shown in Figure 12.12. The setup consists of a pulse generator and may include an oscilloscope to measure the voltage across the device and a current-limiting element such as a series resistor or a transistor in a location that may differ from the one illustrated in Figure 12.12. A current reading capability may be provided by the $50\ \Omega$ input impedance of an oscilloscope, a combination of a resistor and an amplifier, or another method suited for the target bandwidth of the setup. Even when there is a means to observe the device current during switching, the apparatus may not be sensitive enough to precisely determine HRSs. In this case, an alternative method to read the device state before and after a pulse is necessary and that could be in the form of a SPA to perform quasistatic I - V sweeps or an AC (alternating current) method.

An example of measured switching dynamics in response to a voltage step applied to a TaO_x resistive memory cell is seen in Figure 12.13a for the SET and in Figure 12.13b for the RESET operation [25]. With the measured device voltage and device current, other dynamical variables can be extracted. Figure 12.14a,c show the evolution of the device conductance and the energy differential of conductance, respectively, for ON switching. Similar plots are displayed for OFF switching in terms of resistance as shown in Figure 12.14b,d. It can be observed

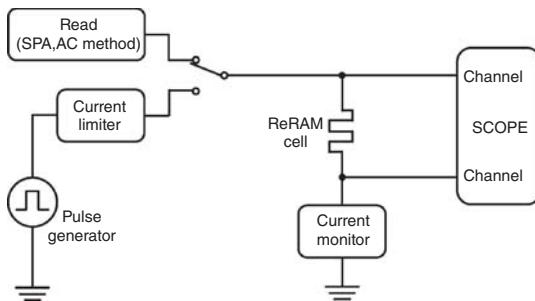


Figure 12.12 Basic pulse measuring setup to study ReRAM device dynamics. The location of the current limiting element may differ from the one in the illustration.

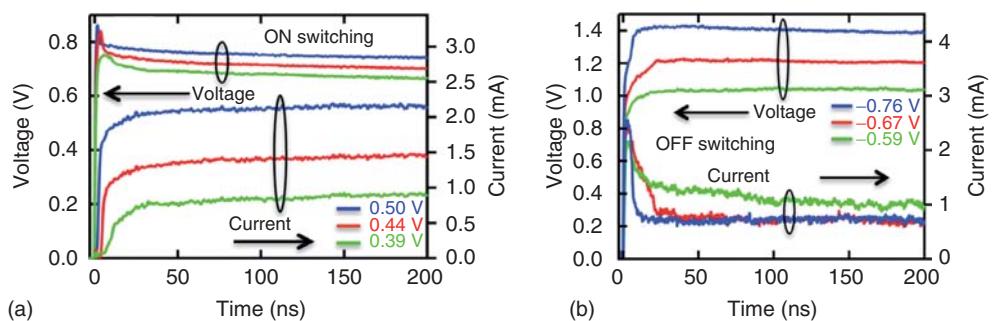


Figure 12.13 Current and voltage dynamics for (a) ON switching and (b) OFF switching of a TaO_x ReRAM cell. For each case, three different pulse amplitudes (measured with a 50Ω load) were applied to the device. (Reproduced with permission from Ref. [25].)

from Figure 12.14 that the device conductance (resistance) changes considerably with the initial energy expenditure, but after that, it barely changes with any additional energy injected. This results in energy waste and possibly device damage and favors the use of shorter pulse widths to optimize the energy cost and to limit potential cell damage. The apparatus utilized for these experiments is detailed in Section 12.4.2.

The study of switching dynamics in oxide-based ReRAM devices also led to the observation of a steep decrease in switching time and energy as a higher voltage is applied [9, 19, 26–28]. Therefore, reduced write energy can be obtained by using short write pulses with higher voltage amplitude. Figure 12.15a shows that an increase of about 1 V in the magnitude of the applied pulse results in a decrease of 4 orders of magnitude in the pulse width required to achieve a target resistance state for a $\text{HfO}_x/\text{AlO}_x$ bilayer cell [26]. As shown in Figure 12.15b, an increase in the SET applied voltage by a factor of 5 leads to a decrease in the SET switching time by 9 orders of magnitude in an SrTiO_3 resistive cell [27]. The setup for the experiment shown in Figure 12.15b was limited by an estimated RC time constant of 20 ns due to a large top disk electrode with 50 μm diameter. The extreme non-linearity verified in the switching kinetics of oxide-based ReRAM devices may be

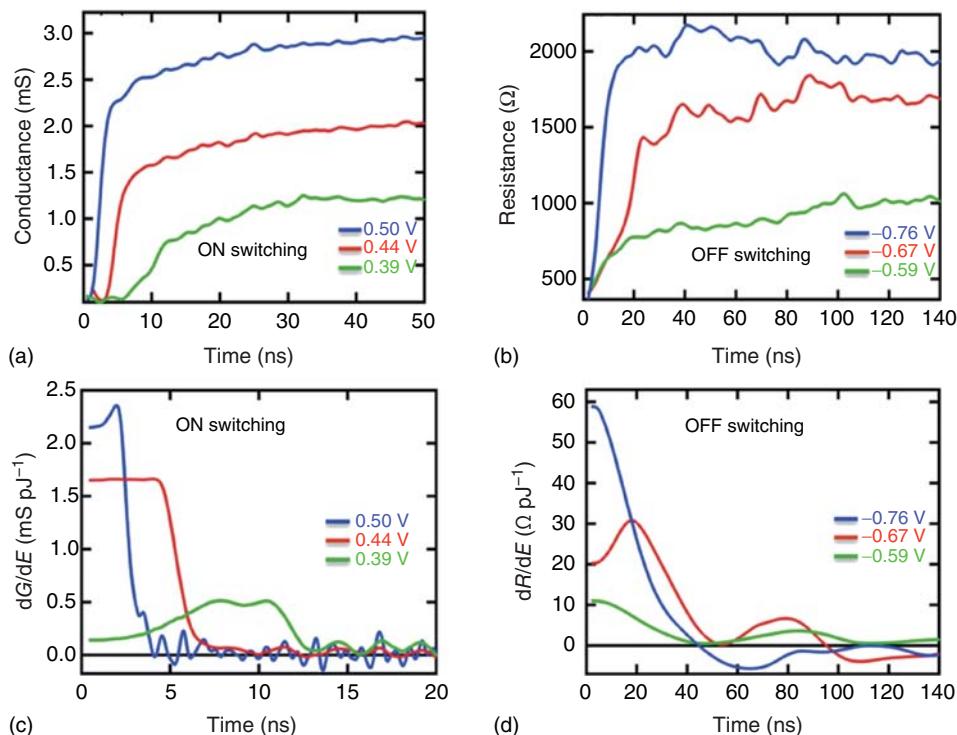


Figure 12.14 (a) Conductance dynamics and (c) energy differential of conductance for ON switching. (b) Resistance dynamics and (d) energy differential of resistance for

OFF switching. For each case, three different pulse amplitudes (measured with a 50Ω load) were applied to a TaO_x ReRAM device. (Reproduced with permission from Ref. [25].)

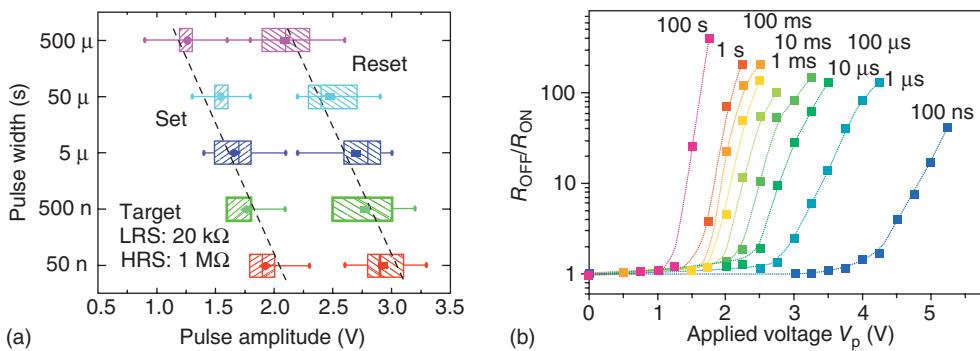


Figure 12.15 (a) Required pulse amplitude to successfully set and reset a $\text{HfO}_x/\text{AlO}_x$ bilayer ReRAM device at some fixed pulse widths. Successful switching was defined as reaching a predefined target LRS/HRS ($20\text{k}\Omega/1\text{M}\Omega$). (Reproduced with permission

from Ref. [26].) (b) ON switching dynamics of an SrTiO_3 ReRAM cell plotted as the HRS/LRS ($R_{\text{OFF}}/R_{\text{ON}}$) ratio as a function of the applied pulse amplitude and pulse width. (Reproduced with permission from Ref. [27].)

attributed to the enhanced drift velocity of oxygen vacancies caused by the high applied electric field and acceleration due to local heating [26, 27, 29]. Besides the programming pulse, another avenue to decrease the switching energy is to fabricate devices with high resistance in the LRS [9].

12.4.1

Experimental Setup and Results for Nanosecond Switching with Real-Time Monitoring of Device Dynamics

Expanding on the basic frame of a pulse measuring system outlined in Figure 12.12, this section details a setup, seen in Figure 12.16a, that was devised to observe voltage and current dynamics in ReRAM cells on the nanosecond timescale [30–32]. Although single-digit nanosecond switching in ReRAM devices has been reported, monitoring of the full device dynamics, particularly the device current, is not commonplace [33–35].

The setup shown in Figure 12.16a consists of a pulse generator able to deliver pulses as short as 1 ns at full width at half maximum (FWHM). The generator is followed by an impedance-matching network that produces an attenuated replica of the generator signal in the pick-off port. While the picked-off signal is sampled by a fast scope with gigahertz bandwidth, the input signal from the pulse generator passes the matching network with little attenuation toward one electrode of the device. The other device electrode is connected to a shunt resistor that is utilized to monitor the device current through two stages of operational amplifiers. The op-amp of choice was an OPA847 with a closed-loop gain of 20 and bandwidth of about 300 MHz. The output of each op-amp and all scope channels are terminated by a $50\ \Omega$ impedance, resulting in an overall voltage-to-voltage gain of $10\times$ and $200\times$ for the first and second amplification stages, respectively, as seen by the scope.

Measuring the device current of high-resistance cells using the devised broadband setup is challenging, so an AC method was added to evaluate the device state before and after a fast pulse. The AC method is comprised of a lock-in amplifier,

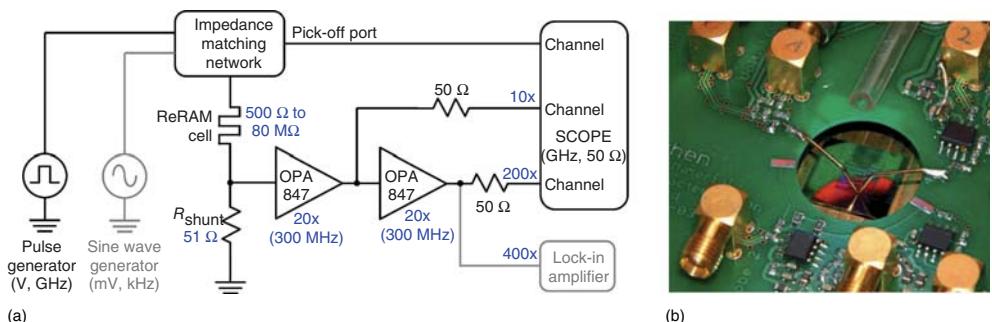


Figure 12.16 (a) Schematic and (b) picture of pulse measuring setup developed to study nanosecond switching dynamics in ReRAM devices. (Reproduced and adapted with permission from Refs. [30, 31].)

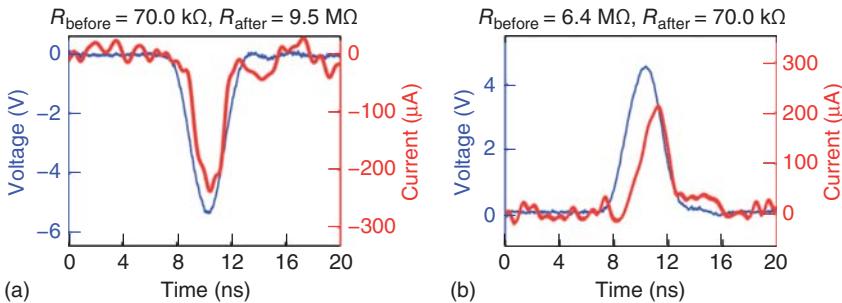


Figure 12.17 Current and voltage dynamics for (a) RESET and (b) SET operations for a nanocrossbar ($100 \text{ nm} \times 100 \text{ nm}$) TiO_2 ReRAM cell. (Reproduced with permission from Ref. [32].)

located at the $400\times$ output of the second stage gain, and a sine-wave generator delivering a millivolt signal at a kilohertz frequency to a memory cell without disturbing its state. Device resistance up to approximately $80 \text{ M}\Omega$ can be measured by the implemented AC method.

The setup was assembled on a PCB with a 3 cm diameter opening that allows probing needles to access the sample as seen in Figure 12.16b. The probing needles were made as short as possible to minimize parasitic capacitance and make the probes electrically short for the frequencies of interest. Besides minimizing parasitic capacitance, any remaining capacitor in the sample path should require a low equivalent resistor wherever it is possible to reduce the RC loading. For instance, the sensing probe in the setup (right probe shown in Figure 12.16b) is connected to the 51Ω shunt resistor.

Nanosecond switching while monitoring the device voltage and current was demonstrated for nanocrossbar TiO_2 devices ($100 \text{ nm} \times 100 \text{ nm}$) using the described setup [32]. The results are displayed in Figure 12.17a for OFF switching and in Figure 12.17b for ON switching, with the measured device state before and after the write pulse using the aforementioned AC method. The measured high peak current could lead to a high device temperature by Joule heating, which in turn increases the mobility of oxygen vacancies and may enable the observed switching time [32]. Switching time of 1 ns has also been observed in GeTe phase-change memory cells utilizing the apparatus described [36].

12.4.2

Experimental Setup and Results for Subnanosecond Switching with Real-Time Monitoring of Device Dynamics

After observing switching on the order of 1 ns, the next step is to verify whether ReRAM devices can perform even faster, reaching subnanosecond write speeds comparable to volatile static random-access memory (SRAM) and faster than volatile dynamic random-access memory (DRAM) [37]. Switching below 1 ns has been reported for ReRAM devices, for instance, in HfO_x -based cells without

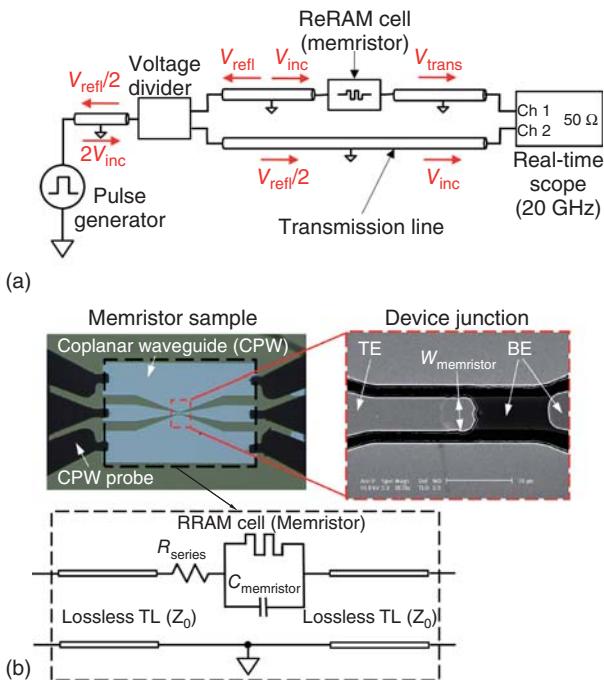


Figure 12.18 (a) Broadband testing setup to study ReRAM device dynamics with subnanosecond time resolution. (b) Optical image and scanning electron microscope image of a ReRAM cell (memristor)

integrated with a tapered coplanar waveguide (CPW). The equivalent circuit for this configuration is also shown. (Adapted with permission from Ref. [39].)

tracking full device dynamics [38], and in TaO_x -based memory cells with observation of full device dynamics [39].

The setup utilized in Ref. [39] for the observation of subnanosecond switching with real-time monitoring of device current and voltage from a single-shot measurement is shown in Figure 12.18a. The experimental setup is composed of a fast pulse generator (or a combination of a fast pulse generator with a fast amplifier) delivering pulses with 20–80% rise time $t_{\text{rise}(20-80)}$ below 100 ps and a 20-GHz real-time scope with channels terminated by a 50Ω impedance. The fast scope monitors not only the transmitted voltage pulse V_{trans} through the ReRAM DUT but also the incident voltage pulse V_{inc} from the pulse generator and the reflected voltage pulse V_{refl} from the DUT. In order to ensure that the fast pulses reach the DUT with reduced pulse broadening and negligible ringing from multiple reflections, low-loss transmission lines (TLs) with characteristic impedance $Z_0 = 50\Omega$ were utilized as interconnects with impedance-matched source and scope loads.

Although most of the TL in the setup consists of external coaxial cables, the TL takes the form of a tapered coplanar waveguide (CPW) on the ReRAM cell wafer followed by CPW probes to interface with the external cabling as seen in Figure 12.18b. The devised approach overcomes broadening associated with

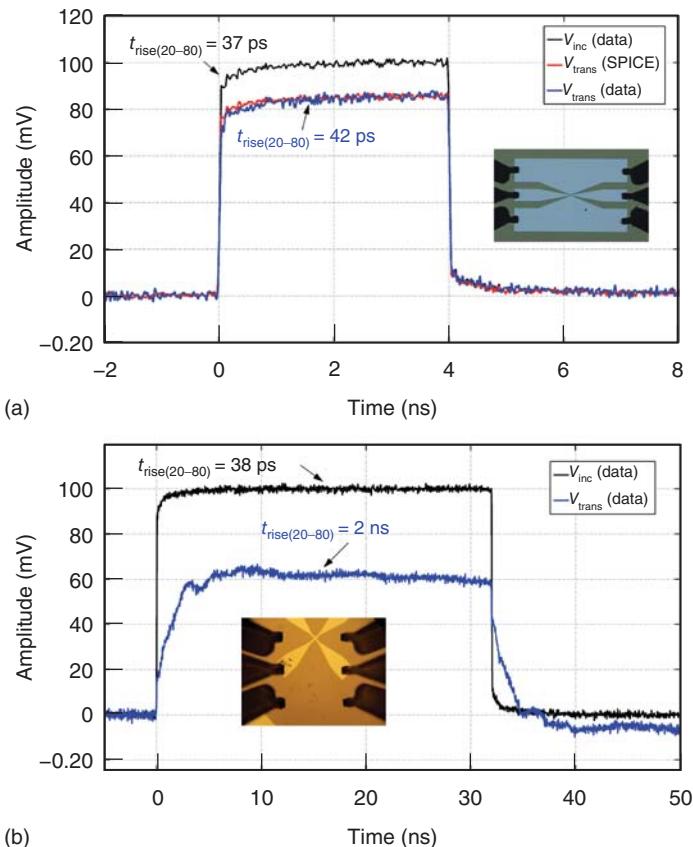


Figure 12.19 Measured incident pulse (black curve) and transmitted pulse (blue curve) through (a) a tapered coplanar waveguide and (b) a conventional crossbar

electrode for ReRAM cells. The series resistance is 18Ω for the CPW and 65Ω for the crossbar electrode. For both cases, there is no ReRAM device connected.

parasitic capacitance and discontinuities in the local return current path existent in conventional electrode structures. Figure 12.19 compares the performance of the designed CPW with the one of a conventional crossbar structure. While an incident pulse with $t_{\text{rise}}(20-80)$ of 38 ps is broadened to the nanosecond scale in the case of the crossbar (Figure 12.19b), the transmitted pulse through the tapered CPW without a ReRAM device connected does not experience significant broadening (Figure 12.19a). The attenuation observed in the transmitted signal is due to a series resistance R_{series} of 18Ω for the CPW and 65Ω for the crossbar electrode.

The current through the device $I_{\text{memristor}}$ and the voltage across the device $V_{\text{memristor}}$ can be obtained from the single-shot measurement of V_{inc} and V_{trans} . Following the equivalent circuit for the ReRAM cell in series with the tapered

CPW as seen in Figure 12.18b, the device current and voltage are given by:

$$I_{\text{memristor}} = I_{\text{sample}} - I_C \quad (12.1)$$

$$V_{\text{memristor}} = V_{\text{sample}} - R_{\text{series}} I_{\text{sample}}, \quad (12.2)$$

where $I_{\text{sample}} = V_{\text{trans}}/Z_0$, $V_{\text{sample}} = 2*(V_{\text{inc}} - V_{\text{trans}})$, R_{series} is approximately equal to the DC series resistance of the tapered CPW, and $I_C = C_{\text{memristor}} dV_{\text{memristor}}/dt$ is the current through the parasitic capacitance $C_{\text{memristor}}$ associated with the geometry and material stack of the memory cell. The parasitic elements R_{series} and $C_{\text{memristor}}$ can be determined, for instance, from frequency-domain measurements using a vector network analyzer [39].

Figure 12.20 shows the first ON switching of a $2\text{ }\mu\text{m} \times 2\text{ }\mu\text{m}$ TaO_x ReRAM device ($w_{\text{memristor}} = 2\text{ }\mu\text{m}$ as indicated in Figure 12.18b) with retrieved device voltage and current from a single-shot measurement. The current through the virgin memristor cell is initially capacitive until the device turns on after 700 ps in the applied pulse as verified by an increase in device current and decrease in device voltage. The parasitic elements for this particular memory cell are $R_{\text{series}} = 67\text{ }\Omega$ and $C_{\text{memristor}} = 0.12\text{ pF}$.

Faster and repeatable switching was achieved for this TaO_x device, and the results are seen in Figure 12.21 [39]. Successful SET operations were achieved by applying an incident pulse with amplitude of $+2.0\text{ V}$ and FWHM = 105 ps (Figure 12.21a). For RESET operations, an incident pulse with amplitude of -3.3 V and FWHM = 120 ps was required (Figure 12.21b). Change in the device state was confirmed by performing low bias $I-V$ sweep up to 0.1 V before and after each write pulse. Based on the measured real-time device dynamics, the energies to set and reset the device could be computed by integrating the product $V_{\text{memristor}} \times I_{\text{memristor}}$ during the pulse and were determined to be 1.9 and 5.8 pJ ,

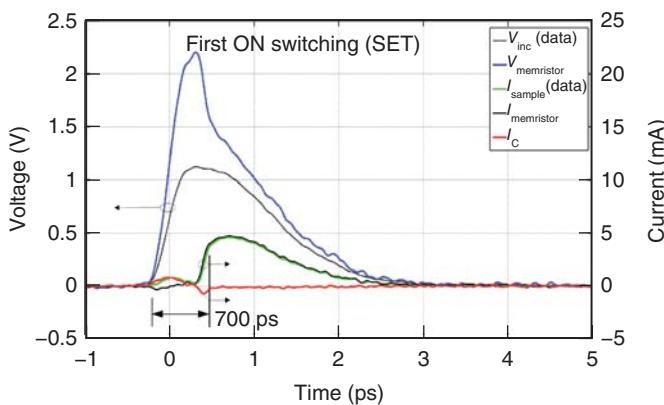


Figure 12.20 First ON switching of a $2\text{ }\mu\text{m} \times 2\text{ }\mu\text{m}$ TaO_x ReRAM cell ($w_{\text{memristor}} = 2\text{ }\mu\text{m}$) occurring within 700 ps in the applied pulse. The plot shows the

measured incident voltage V_{inc} , the measured sample current I_{sample} , the ReRAM cell current $I_{\text{memristor}}$ and voltage $V_{\text{memristor}}$, as well as the parasitic capacitive current I_C .

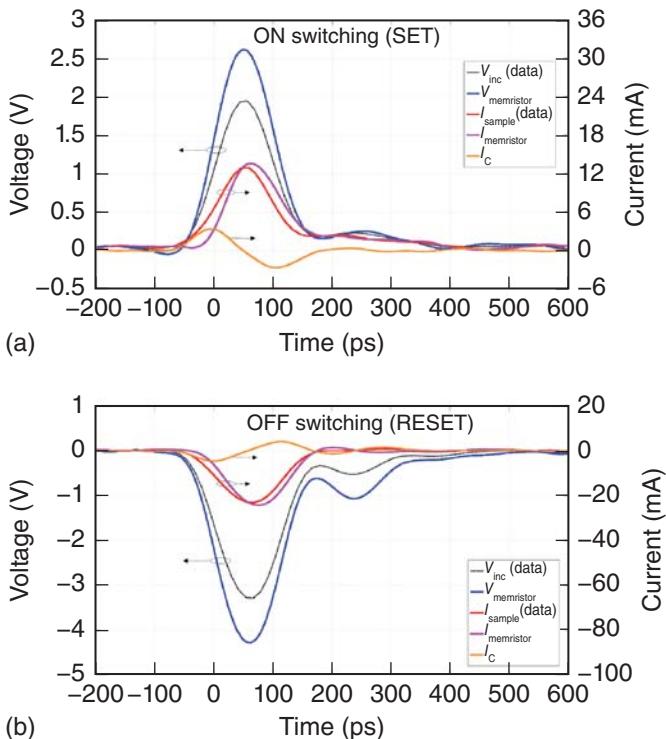


Figure 12.21 Subnanosecond (a) SET and (b) RESET of a $2\text{ }\mu\text{m}$ TaO_x -based ReRAM device with monitored device dynamics. (Reproduced with permission from Ref. [39].)

respectively. By comparison, using ns pulses to set ($V_{\text{inc}} = +1.0\text{ V}/1.5\text{ ns}$) and reset ($V_{\text{inc}} = -1.5\text{ V}/1.5\text{ ns}$) the same device required lower voltage amplitude and higher write energy, 6.6 pJ for SET and 14.3 pJ for RESET, in line with the trend discussed earlier in this chapter.

12.5

Conclusions

In summary, here, we addressed the issues pertaining to electrical characterization of ReRAM devices, focusing on the device behavior upon current compliance and device time dependence down to subnanosecond timescale. The challenges for proper dynamical device characterization such as ReRAM go beyond simplistic analysis on what is commercially available. Parasitics de-embedding, impedance matching, and detailed knowledge of driver performance are of fundamental importance for correctly assessing the device parameters. Our nonlimiting examples indicate the most salient issues that are present in routine device characterization, which should aid in developing parametric device structures and test apparatus.

Acknowledgment

The authors would like to acknowledge the numerous discussions with many of our colleagues: J. Borghetti, F. Barachati, F. Miao, J. Nickel, M. D. Pickett, J. P. Strachan, J. J. Yang, W. Yi, R. Waser, and R. S. Williams.

References

1. Sawa, A., Fujii, T., Kawasaki, M., and Tokura, Y. (2004) Hysteretic current–voltage characteristics and resistance switching at a rectifying $Ti/Pr_{0.7}Ca_{0.3}MnO_3$ interface. *Appl. Phys. Lett.*, **85**, 4073.
2. Chang, W.-Y., Liao, J.-H., Lo, Y.-S., and Wu, T.-B. (2009) Resistive switching characteristics in $Pr_{0.7}Ca_{0.3}MnO_3$ thin films on $LaNiO_3$ -electroded Si substrate. *Appl. Phys. Lett.*, **94**, 172107.
3. Menke, T., Meuffels, P., Dittmann, R., Szot, K., and Waser, R. (2009) Separation of bulk and interface contributions to electroforming and resistive switching behavior of epitaxial Fe-doped $SrTiO_3$. *J. Appl. Phys.*, **105**, 066104.
4. Yu, S., Jeyasingh, R., Wu, Y., and Philip Wong, H.-S. (2012) Characterization of low-frequency noise in the resistive switching of transition metal oxide HfO_2 . *Phys. Rev. B*, **85**, 045324.
5. Losee, D.L. (1975) Admittance spectroscopy of impurity levels in Schottky barriers. *J. Appl. Phys.*, **46**, 2204.
6. Froehlich, H. (1949) *Theory of Dielectrics*, Oxford University Press.
7. Keithley (2013) Technical Information: Source Measurement Unit (SMU) Instruments.
8. Chua, L.O. (2011) Resistance switching memories are memristors. *Appl. Phys. A*, **102**, 765–783.
9. Ielmini, D., Nardi, F., and Balatti, S. (2012) Evidence of voltage-driven set/reset processes in bipolar switching RRAM. *IEEE Trans. Electron Devices*, **59** (8), 2049–2056.
10. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) Redox-based resistive switching memories – nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21**, 2632–2663.
11. Kinoshita, K., Tsunoda, K., Sato, Y., Noshiro, H., Yagaki, S., Aoki, M., and Sugiyama, Y. (2008) Reduction in the reset current in a resistive random access memory consisting of $NiOx$ brought about by reducing a parasitic capacitance. *Appl. Phys. Lett.*, **93**, 033506.
12. Miao, F., Yi, W., Goldfarb, I., Yang, J.J., Zhang, M.-X., Pickett, M.D., Strachan, J.P., Medeiros-Ribeiro, G., and Williams, R.S. (2012) Continuous electrical tuning of the chemical composition of $TaOx$ -based memristors. *ACS Nano*, **6** (3), 2312–2318.
13. Yang, J.J., Strukov, D.B., and Stewart, D.R. (2013) Memristive devices for computing. *Nat. Nanotechnol.*, **8**, 13–24.
14. Yang, J.J., Miao, F., Pickett, M.D., Ohlberg, D.A.A., Stewart, D.R., Lau, C.N., and Williams, R.S. (2009) The mechanism of electroforming of metal oxide memristive switches. *Nanotechnology*, **20**, 215201.
15. Liang, J. and Philip Wong, H.-S. (2010) Cross-point memory arrays without cell selectors – device characteristics and data storage pattern dependencies. *IEEE Trans. Electron Devices*, **57** (10), 2531–2538.
16. Yu, S., Guan, X., and Philip Wong, H.-S. (2011) Conduction mechanism of $TiN/HfO(x)/Pt$ resistive switching memory: a trap-assisted-tunneling model. *Appl. Phys. Lett.*, **99**, 063507.
17. Kim, K.M., Choi, B.J., Lee, M.H., Kim, G.H., Song, S.J., Seok, J.Y., Yoon, J.H., Han, S., and Hwang, C.S. (2011) A detailed understanding of the electronic bipolar resistance switching behavior in $Pt/TiO_2/Pt$ structure. *Nanotechnology*, **22**, 254010.

18. Schindler, C., Staikov, G., and Waser, R. (2009) Electrode kinetics of Cu–SiO₂-based resistive switching cells: overcoming the voltage-time dilemma of electrochemical metallization memories. *Appl. Phys. Lett.*, **94**, 072109.
19. Strachan, J.P., Torrezan, A.C., Miao, F., Pickett, M.D., Yang, J.J., Yi, W., Medeiros-Ribeiro, G., and Williams, R.S. (2013) State dynamics and modeling of tantalum oxide memristors. *IEEE Trans. Electron Devices*, **60** (7), 2194–2202.
20. Sheridan, P., Kim, K.-H., Gaba, S., Chang, T., Chen, L., and Lu, W. (2011) Device and SPICE modeling of RRAM devices. *Nanoscale*, **3** (9), 3833–3840.
21. Chua, L.O. and Kang, S.M. (1976) Memristive devices and systems. *Proc. IEEE*, **64** (2), 209–223.
22. Chua, L.O. (2012) The fourth element. *Proc. IEEE*, **100** (6), 1920–1927.
23. Lu, Y.M., Noman, M., Chen, W., Salvador, P.A., Bain, J.A., and Skowronski, M. (2012) Elimination of high transient currents and electrode damage during electroformation of TiO₂-based resistive switching devices. *J. Phys. D: Appl. Phys.*, **45**, 395101.
24. Nardi, E., Ielmini, D., Cagli, C., Spiga, S., Fanciulli, M., Goux, L., and Wouters, D.J. (2011) Control of filament size and reduction of reset current below 10 μA in NiO resistance switching memories. *Solid-State Electron.*, **58**, 42–47.
25. Strachan, J.P., Torrezan, A.C., Medeiros-Ribeiro, G., and Williams, R.S. (2011) Measuring the switching dynamics and energy efficiency of tantalum oxide memristors. *Nanotechnology*, **22**, 505402.
26. Yu, S., Wu, Y., and Philip Wong, H.-S. (2011) Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory. *Appl. Phys. Lett.*, **98**, 103514.
27. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011) Origin of the ultra-nonlinear switching kinetics in oxide-based resistive switches. *Adv. Funct. Mater.*, **21** (23), 4487–4492.
28. Pickett, M.D., Strukov, D.B., Borghetti, J.L., Yang, J.J., Snider, G.S., Stewart, D.R., and Williams, R.S. (2009) Switching dynamics in titanium dioxide memristive devices. *J. Appl. Phys.*, **106**, 074508.
29. Strukov, D.B. and Williams, R.S. (2009) Exponential ionic drift: fast switching and low volatility of thin-film memristors. *Appl. Phys. A*, **94**, 515–519.
30. aixACCT Systems GmbH (2011) Catalog Brochure.
31. Bruns, G. (2012) Electronic switching in phase-change materials. PhD dissertation. RWTH Aachen University, Aachen.
32. Hermes, C., Wimmer, M., Menzel, S., Fleck, K., Bruns, G., Salina, M., Böttger, U., Bruchhaus, R., Schmitz-Kempen, T., Wuttig, M., and Waser, R. (2011) Analysis of transient currents during ultrafast switching of TiO₂ nanocrossbar devices. *IEEE Electron Device Lett.*, **32** (8), 1116–1118.
33. Ignatiev, A., Wu, N.J., Chen, X., Liu, S.Q., Papagianni, C., and Strozier, J. (2006) Resistance switching in perovskite thin films. *Phys. Status Solidi B*, **243** (9), 2089–2097.
34. Yoshida, C., Tsunoda, K., Noshiro, H., and Sugiyama, Y. (2007) High speed resistive switching in Pt/TiO₂/TiN film for nonvolatile memory application. *Appl. Phys. Lett.*, **91**, 223510.
35. Zhang, X.T., Yu, Q.X., Yao, Y.P., and Li, X.G. (2010) Ultrafast resistive switching in SrTiO₃:Nb single crystal. *Appl. Phys. Lett.*, **97**, 222117.
36. Bruns, G., Merkelbach, P., Schlockermann, C., Salina, M., Wuttig, M., Happ, T.D., Philipp, J.B., and Kund, M. (2009) Nanosecond switching in GeTe phase change memory cells. *Appl. Phys. Lett.*, **95**, 043108.
37. ITRS International Technology Roadmap for Semiconductors (2012) ITRS 2012 Edition, <http://www.itrs.net> (accessed 09 July 2015).
38. Lee, H.Y., Chen, Y.S., Chen, P.S., Gu, P.Y., Hsu, Y.Y., Wang, S.M., Liu, W.H., Tsai,

- C.H., Sheu, S.S., Chiang, P.C., Lin, W.P., Lin, C.H., Chen, W.S., Chen, F.T., Lien, C.H., and Tsai, M.-J. (2010) Evidence and solution of over-reset problem for HfO_x based resistive memory with sub-ns switching speed and high endurance. 2010 IEEE International Electron Devices Meeting, San Francisco, CA, p. 19.7.1.
39. Torrezan, A.C., Strachan, J.P., Medeiros-Ribeiro, G., and Williams, R.S. (2011) Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, **22**, 485203.

13

Unipolar Resistive-Switching Mechanisms

Ludovic Goux and Sabina Spiga

13.1

Introduction to Unipolar Resistive Switching

Today, a large family of emerging nonvolatile memory concepts is based on the switching between a low-resistive state (LRS) and a high-resistive state (HRS) within a two-terminal memory device. These LRS and HRS states, sometimes called *set and reset*, or even *ON and OFF states*, are programmed by electrical stimulation [1–3] and have been demonstrated in a wide variety of material systems. Between the numerous mechanisms potentially at the origin of resistive-switching effects, nanoionic transport and redox-reaction mechanisms taking place at the nanometer scale [4] have been clearly identified as accounting for the switching functionality of various systems. This is, for example, the case of electrochemical metallization memory (ECM) and valence change memory (VCM) cell concepts, as detailed in different chapters of this book.

Interestingly, due to the nanoscale character of the switching activated by these mechanisms, the associated so-called resistive random-access memory (RRAM) technology has been so far perceived as highly scalable [4, 5] and has thus been developed intensively. Over the past decade, the outstanding progress of RRAM in terms of compatibility with CMOS integration [3], memory characteristics, and reliability has in turn aroused considerable interest in the fields of nanoionics and resistive switching. Particularly, nanoscale ionics and redox mechanisms are intensively investigated in oxide materials that form VCM systems, for which the switching relies mainly on the migration/drift of oxygen-ionic species [4]. Interestingly, for other specific oxide systems, the switching is not dominated by migration but by thermally controlled diffusion and redox reaction. This distinct class of switching mechanism is called thermochemical memory (TCM) mechanism [6].

This major microscopic difference between VCM and TCM mechanisms reflects in a different switching operation. Indeed, while the migration-dominated origin of the switching in VCM systems necessitates the use of different voltage polarities for programming LRS and HRS states, change in polarity is not required in TCM systems. Therefore, TCM concepts are inherently *unipolar*

resistive-switching (URS) mechanisms, as opposed to bipolar resistive-switching (BRS) mechanisms in VCM systems.

The specific class of TCM systems is the focus of this chapter, whose purpose is to provide a state-of-the-art understanding of URS mechanisms. In Section 13.2, we briefly describe the typical cell structure and basic switching operation in various material systems investigated in the literature. Section 13.3 is dedicated to Pt\NiO\Pt systems, for which the underlying microscopic mechanisms and proposed models for unipolar switching are detailed together with their impact on memory properties. Finally, the influence of oxide and electrode materials on switching mechanisms and device characteristics is discussed in Section 13.4.

13.2

Principle of Unipolar Switching

13.2.1

Basic Operation of Unipolar Memory Cells

Unipolar switching is usually observed in simple metal–insulator–metal (MIM) structures (Figure 13.1a), where the “I” and “M” elements typically constitute an oxide layer and a noble metallic layer, respectively. Fresh cells are most often in a very HRS requiring the application of a large “forming” voltage (V_f), which allows to turn the cell for a first time in LRS. After this so-called “forming” conditioning step, the cell may be reversibly reset-switched to HRS and set-switched to LRS, as illustrated in Figure 13.1b.

The forming operation is generally regarded as a kind of electrical breakdown [7], which critically needs to be current-limited to preserve the subsequent switching functionality of the cell. Current limitation may be achieved using the current compliance (CC) function of a commercial semiconductor parameter analyzer (SPA). However, it is important to note that the SPA response is slower than the forming transient, so that a spike current overshoot may not be avoided in practice [8, 9]. The transient forming current is more efficiently limited in the so-called 1-transistor/1-resistor (1T1R) integration scheme (Figure 13.1c), whereby the current flowing through the MIM cell is controlled by fixing an appropriate gate potential (V_{Gate}) [1, 7, 9]. Figure 13.1c shows typical URS I - V traces obtained in a 1T1R cell [10]. By using a relatively low V_{Gate} , both the forming and set currents are limited by the saturation characteristics of the transistor. On the other hand, a larger V_{Gate} is applied during the reset operation due to the relatively high current required to reset the cell ($I_{reset} > I_{set}$). The reset voltage (V_{reset}) is generally lower than the set voltage (V_{set}). Typically, the current drop at V_{reset} is sharp and may allow several decades of resistance difference between LRS and HRS. Note, however, that $V_{set} < V_f$ (Figure 13.1c), which indicates that the reset operation does not correspond to a return to the fresh state.

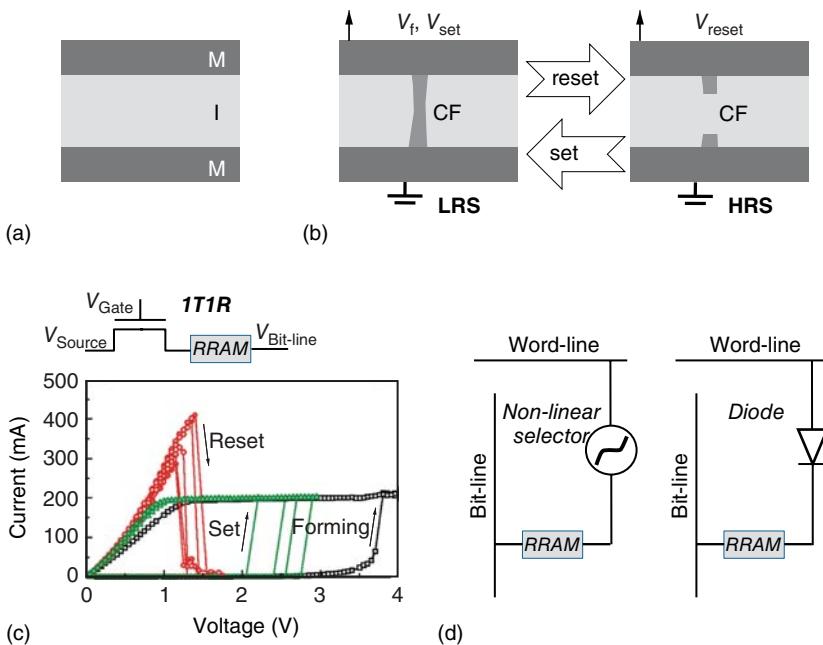


Figure 13.1 (a) Metal–insulator–metal (MIM) structure of a unipolar resistive-switching cell. (b) Formation/rupture of a conductive filament (CF) through the cell induced by set/reset operations respectively. (c) Unipolar switching I - V characteristics

of a 1-transistor/1-resistor (1T1R) cell. (d) Cross-point architecture of an RRAM array, using a nonlinear two-terminal cell selector, which may be a unidirectional diode in the case of a unipolar RRAM.

13.2.2 Structure of Unipolar Memory Arrays

The typical architecture of unipolar memory arrays is the well-known crossbar organization, where each MIM element is associated with a nonlinear selector in each cross-point (Figure 13.1d) (see also Chapter 22, dedicated to selectors). The role of the selector element is both to select the addressed memory cell and to electrically isolate the nonaddressed cells. This means that the selector should both (i) drive a sufficiently high current at the “program” voltage to allow effective switching of the addressed cell and (ii) limit the current at the voltage dropping in nonaddressed cells to certain levels, in order to avoid programming disturbs.

It follows from these requirements that I_{reset} will be the limiting parameter for scaling because the required driving current for reset will limit the scaling of the selector. On the other hand, V_{reset} appears as the limiting parameter for reading, as the read voltage V_{read} should be significantly lower than V_{reset} to ensure sufficient disturb immunity. Finally, the minimum voltage needed for unipolar operation will be determined by the forming operation. However, consistently with a breakdown-type process, V_f has been shown to scale with the

oxide thickness [11] or to be engineered to lower values using fresh oxides with increased density of defects [8]. Such approaches may allow obtaining forming-free systems, that is, where the forming operation may not be distinguished from subsequent set operations [8, 12], or even systems that are “initially ON,” that is to say, in which the defect density has reached a level that necessitates a reset operation to initialize the cell [12].

As for the choice of selector device, the unipolar nature of the switching allows the use of a unipolar selector, such as a two-terminal diode (Figure 13.1d). This peculiarity has driven a large interest for unipolar systems during the past 10 years because of the inherently better scalability of diode selectors compared to three-terminal transistor-based selectors. Their integration in crossbar arrays allows, indeed, a *footprint down to $4F^2$* , where F denotes the minimum feature size of a specific technology. More integration details are addressed in Chapter 20, focusing on integration technology and cell design.

13.2.3

Experimental Evidences of Filamentary-Switching Mechanism

In BRS systems, both localized filamentary switching (in HfO_2 cells [13]) and uniform areal switching (in SrTiO_3 cells [1]) have been demonstrated. On the contrary, uniform areal switching *has never been evidenced in URS systems*.

Similarly to any electrical breakdown, the forming operation triggers the formation of a local path of increased electrical conductivity through the oxide layer. In the field of resistive switching, this path is usually referred to as conductive filament (CF). As sketched in Figure 13.1b, reset/set operations are related to the rupture/restoration of this CF by fuse/antifuse thermal effects as detailed in Section 13.3. As said earlier, the fresh state is not recovered, implying that the CF rupture is only partial.

The filamentary switching in unipolar devices has been demonstrated by several experiments including electrical and physical analyses. The first evidence of filamentary switching is given by the local conduction in LRS that has been revealed by means of conductive atomic force microscopy (C-AFM) in NiO -based URS systems [14–19] by various authors. Secondly, filamentary switching has been pointed out in several reports due to the cell-size independence of the electrical-switching parameters [7, 20]. Particularly, reports showed that the resistance of fresh NiO cells is inversely proportional to the cell size, while no significant size dependence is observed for LRS and HRS resistances (Figure 13.2) and for V_{reset} and V_{set} parameters down to $0.18 \mu\text{m}$ cell size [7, 20]. These characteristics point to a switching localized in a region much smaller than the cell size, so that the current is limited by the CF both in LRS and in HRS. This provides great potential in terms of cell scalability. Note that the observation of a clear cell-size dependence of state resistances has already been reported, in particular for very small cells [21]. To account for such effects, we may invoke a possible impact of multiple-filament programming [22] or specific effects of scaling to size ranges of the order of the

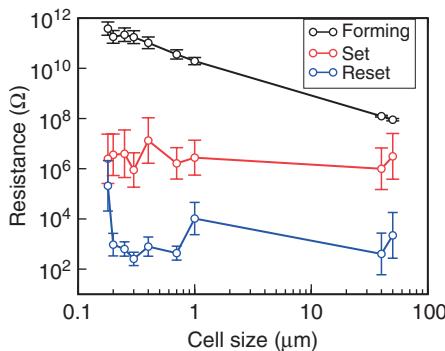


Figure 13.2 Cell-size dependence of the resistance of W\NiO\Pt unipolar RRAM cells in fresh LRS and HRS states, showing a size dependence only for the fresh state.

CF size that may reach the order of ~ 10 nm for very low LRS resistance (R_{LRS}) [23, 24].

There is a debate in the literature whether URS mechanism involves single or multiple CF. For instance, Kim *et al.* attributed intermediate states and anomalous fluctuations during set switching of NiO cells to the formation and rupture of additional CF paths [25]. On the other hand, other reports evidenced clear relationship between discrete resistance fluctuations and variations in the number of defects within a single CF [26]. Multiple filaments have been observed by C-AFM for NiO layers in LRS [14, 17, 15]. Particularly, Son and Shin [15] used a Hg drop as electrode on the surface of the NiO cells to program LRS and HRS states. After removal of the Hg drop, numerous CF paths were revealed by C-AFM preferentially at grain boundaries (GBs) in LRS while significantly less were mapped in HRS. Overall, both single- and multiple-CF situations may be achieved, depending on the thickness and microstructure of the oxide layer, as well as on the forming conditions.

Finally, as detailed in Section 13.3, the conductivity of single and multiple CF may vary between metallic and semiconducting types of conduction, depending on forming CC and cell microstructure [4, 8, 20, 27].

13.2.4

Typical Materials Used in Unipolar-Switching Cells

Resistive-switching and negative differential resistance phenomena were first reported in the 1960s in binary *transition metal oxides* (TMOs). At that time, rather thick oxide materials were studied, including NbO_x [28], Al_2O_3 [29], SiO_x [30], TiO_2 [31], and NiO [32] layers. This early period of research has been comprehensively reviewed in Refs. [33–35].

Since the 2000s, there has been a renewed interest in resistive switching in TMO systems driven by their potential industrial application as RRAMs. In this respect,

the URS mode has received large consideration due to possible integration with a diode, as already mentioned.

A huge amount of TMO materials have been investigated for URS in the past 10 years, focusing on thin films, in order to ensure compatibility with CMOS circuits. Reference [3] gives a recent review of these developments.

Between these materials, URS has been mostly studied in NiO systems [7, 36] and, to a lesser extent, in TiO₂ [37, 38], Gd₂O₃ [39], Al₂O₃ [40, 41], Nb₂O₅ [42, 43], ZrO₂ [44, 45], HfO₂ [46, 47], CuO_x [48, 49], Fe₂O₃ [50], CoO [50, 51], ZnO [52], and SiO_x [53–55]. All these studies have shown that URS is of filamentary type and related to TCM mechanism.

RRAM cells based on SiO_x exhibit a unique behavior. Indeed, as reported in several recent studies [53–55], their set operation is self-compliant, and they show $V_{\text{reset}} > V_{\text{set}}$. Moreover, they have demonstrated promising endurance and fast switching, so that the URS system operating without CC is a serious candidate to be considered in the URS class.

Even though major attention has been devoted to binary oxides due to easier CMOS-compatible integration [5], it is important to mention that a few papers in the literature have also reported on URS in more complex perovskite systems that usually exhibit BRS behavior [1]. For example, Li *et al.* reported on URS in Pr_{0.7}Ca_{0.3}MnO₃ (PCMO) systems [56]. The structure and composition of the oxide layer were identified as primary factors determining URS properties.

To summarize, many binary or perovskite systems have shown URS properties. On the other hand, most experimental results and basic understanding of URS have been obtained on NiO-based cells, in particular on Pt/NiO/Pt cells. The large interest in NiO mainly originates from the outstanding pioneering works by Baek *et al.*, showing excellent switching control and reliability [7, 57]. Pt/NiO/Pt cells have thus ever been considered as model systems for apprehending TCM mechanisms.

As a consequence, the following section dedicated to the microscopic origin and the modeling of URS focuses on the knowledge built up for this prototypical system.

13.3

Unipolar-Switching Mechanisms in Model System Pt/NiO/Pt

13.3.1

Microscopic Origin of Switching in NiO Layers

13.3.1.1 Defect Chemistry

Although NiO is a simple binary oxide having rock-salt crystalline structure, its electronic properties are complex. It shows antiferromagnetic ordering, and it is a strongly correlated electron system, which is predicted to be a metal by the conventional band theory but is a charge-transfer insulator, instead, as a result of the strong Coulomb repulsion between electrons occupying Ni d

orbitals [58]. The states at valence-band maximum (VBM) are composed of a mixture of nickel 3d lower Hubbard band and oxygen 2p band, while the states at conduction-band minimum (CBM) are composed of the nickel 3d upper Hubbard band.

As the switching operation relies on the creation and recovery of defects constituting the CF, it is of prime importance to identify energetically favorable defects created in ideal NiO cubic system. Several first-principle calculations were reported to this aim [58–60]. As the formation energy of vacancies is reportedly lower than that of interstitials [58], we focus here on the introduction of either Ni or O vacancies.

After introducing a Ni vacancy (V_{Ni}) in a $\text{Ni}_{32}\text{O}_{32}$ supercell, the Fermi level ($\text{FL} = E_F$) enters the oxygen 2p band and the Coulomb repulsion causes a band splitting [58]. As a result, both the CBM and VBM are primarily of the oxygen 2p band character [58]. Holes are created, with a binding energy to the nickel vacancy of ~ 0.2 eV [60], which thus accounts for a p-type conductivity of the oxide. Importantly, although the Ni vacancy is known to be a major point defect in bulk NiO [5], it was shown that the formation energies of oxygen-vacancy (V_O) defects in the proximity of VBM are lower than those of the nickel vacancies [60], the positively charged V_O^{2+} being the most stable [59]. Depending on the charge of the defect (V_O or V_{Ni}), additional states are obtained in the band gap up to the CBM [59, 60].

Hence, it can be expected from first-principle calculations that V_O and/or V_{Ni} point defects are generated during fabrication of nickel oxide layers. In the literature, NiO layers are most often prepared by means of sputtering [16, 27, 61, 62] or atomic layer deposition (ALD) [16, 43] techniques. The NiO composition is generally reported to be stoichiometric or Ni-deficient [27, 62]. On the other hand, NiO layers with oxygen deficiency have also been fabricated by sputtering under low oxygen pressure [63], by metal-organic chemical vapor deposition (MOCVD) [11] or by thermal oxidation of Ni layers [12, 64].

Interestingly, controlled nonvolatile URS has been reported for high O deficiency ($\text{O}/\text{Ni} \sim 0.8$ in Ref. [12]) as well as for high Ni deficiency ($\text{O}/\text{Ni} \sim 1.15$ in Refs. [27, 65]). However, Ni-deficient NiO layers also typically exhibit threshold-switching properties [58], meaning that the conductive state generated by the forming operation is not stable after power switch-off. On the other hand, excessive oxygen deficiency results in highly conductive NiO layers [65]. These different aspects suggest that the CF is more likely locally constituted of V_O and/or metallic Ni species rather than V_{Ni} defects. Note that the reported controlled URS for Ni-deficient NiO layers does not mean the layer only contains V_{Ni} defects. For example, Jung *et al.* also identified metallic Ni by X-ray photoelectron spectroscopy (XPS) measurement in $\text{NiO}_{1.15}$ layers [27].

From the aforementioned considerations, the LRS state may thus result from the creation of V_O^{2+} -defect chain and/or metallic Ni chain constituting a CF through the NiO layer. In agreement with this observation, Lee *et al.* have shown by *ab initio* calculation that the generation of V_O^{2+} species in a $\text{Ni}_{64}\text{O}_{64}$ supercell resulted in a structure with a minimum of energy after the O-vacancy defects

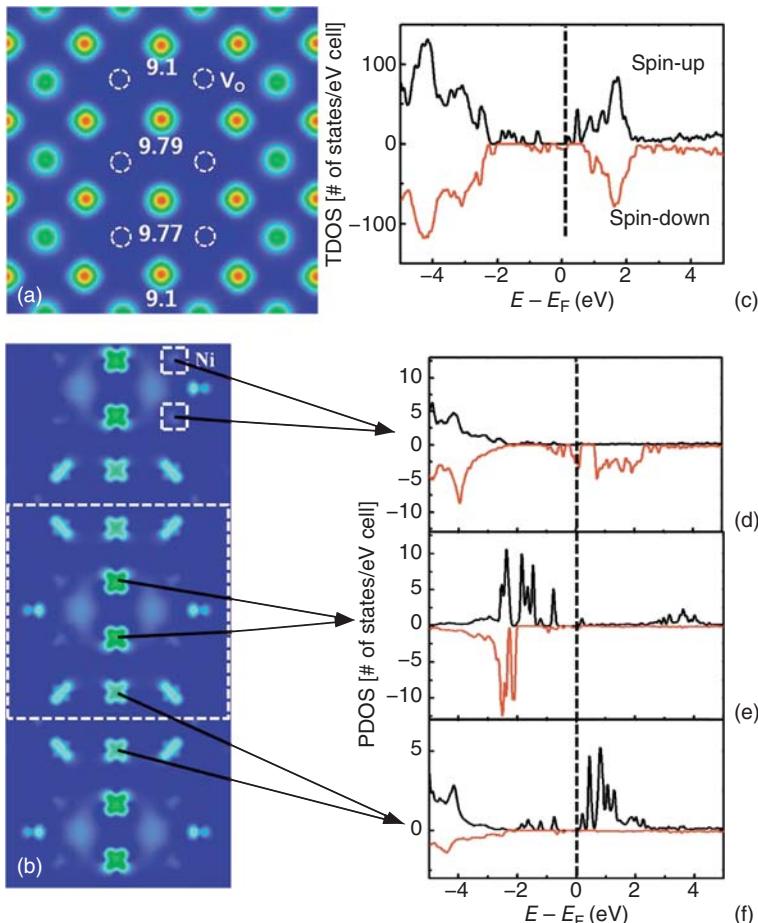


Figure 13.3 (a) Charge density of an $\text{Ni}_{64}\text{O}_{64}$ supercell with six oxygen vacancies. (b) Partial charge density within $E_F \sim E_F + 0.3 \text{ eV}$ in (100) plane including oxygen vacancies and Ni metal chain. (c) Total density of states of the supercell, and (d)–(f) partial density of states of d orbitals

at each Ni atom; FL is set to 0 eV (c)–(f); dotted circle in (a) refers to oxygen vacancy site and square in (b) to Ni site; the number located at each Ni atom of the filament in (a) represents charge from Bader charge analysis: 8.68e for Ni and 7.32e for O in perfect NiO .

have clustered locally, leading to partial reduction of neighboring Ni species from Ni^{2+} into Ni^{1+} or metallic Ni (Figure 13.3a,b). This structure results in a distribution of defect levels over the whole range of the band gap (Figure 13.3c), turning NiO into a metal [59]. The detailed partial density of states (DOS) analysis showed that the states below E_F are from Ni atoms away from the CF, while those between FL and CBM are from Ni atoms inside the CF (Figure 13.3d–f).

13.3.1.2 Microscopic Mechanism of the Switching

By applying statistical time-dependent dielectric-breakdown methodology to NiO cells, Buh *et al.* have evidenced the exponential dependence of forming time on applied voltage, indicating that the forming operation is not a spontaneous process at some critical voltage, but a breakdown-like time-dependent dynamical process resulting from random generation of voltage-stress-induced defects [66]. Several other reports have also shown the NiO-thickness dependence of V_f , pointing to a field control [7, 9, 11].

As mentioned earlier, the generation of V_O^{2+} species is energetically favorable. At voltage levels applied in forming operation, these species are generated by removal of O-anion species. Upon forming transient, high temperature rise is locally expected as induced by the high transient current density, so that the charged O species are expected to migrate under the electric field, as well as to diffuse sideways. According to this scenario, the CF is expected to be made of an O-deficient NiO material. In the literature, several reports have shown experimental evidences of oxygen migration [11, 16] or V_O -based conductivity in NiO programmed in LRS state [67].

However, when O ions reach the anodic interface during forming transient, Ni-cation species may also be generated according to the redox reaction $\text{NiO} \rightarrow \text{Ni}^{2+} + 2\text{e}^- + 1/2\text{O}_2(\text{g})$ and drift toward the cathode. When this process is dominant, a metallic Ni CF is expected to form, as reported by several authors [5].

Thus, it appears that the nature of the CF depends on the mobility of the different species generated at high temperature. From the previous experimental investigations on NiO single crystals, O'Keeffe *et al.* reported that the diffusivity of nickel is larger by several decades than that of oxygen in NiO at 1500 °C [68]. However, elemental diffusivities are expected to be strongly enhanced in polycrystalline layers due to GBs and local disorder generated by forming, so that it is not straightforward to determine the species that shows lower local barrier to diffusion, which in turn explains the various results obtained in the literature.

Whether the CF is made of metallic Ni or O-deficient NiO, the forming operation results in a local reduction mechanism of NiO, activated by the field and the large temperature generated by the forming power. Note that this reduction process at high temperature is in agreement with the Ellingham diagrams showing an increase in the free energies of oxide formation in metals in general as a function of the temperature.

The reset-switching operation corresponds to the rupture of the CF, resulting in a drop in conductivity. Let us assume first that the CF consists of a V_O chain. Upon reset operation, the current density through the CF has been shown to reach high values, generating the necessary thermal energy to activate the migration of oxygen from O-rich regions outside the CF toward O-deficient regions inside the CF. According to this scenario, the reset switching corresponds to the local reoxidation of the CF. This requires that the forming-generated O species move back to the V_O sites, and as the switching is unipolar in nature, this process is expected to take place mainly in the lateral direction by diffusion mechanisms.

In agreement with this scenario, by means of time-of-flight secondary ion mass spectroscopy (TOF-SIMS), Peng *et al.* have evidenced a significant increase in the O/Ni ratio after reset operation in NiO cells [58].

Assuming now that the CF consists of a metallic Ni chain, the thermal dissolution of Ni upon reset switching might be a more realistic mechanism for CF rupture. In this scenario, the diffusion mechanism is expected to occur radially from inside the Ni-rich CF toward the surrounding Ni-poor regions [16].

In both situations, the physics of the reset operation is overall expected to depend on the spatial distribution of defects, on local fields, as well as on temperature profiles. The physics-based modeling of the electrical-switching characteristics of reset and set operations is addressed in the next subsection.

13.3.2

Physics-Based Electrical Models

13.3.2.1 Modeling of the Reset Switching

In this section, the LRS state is assumed to consist of a single cylindrical CF of diameter ϕ and height equal to the thickness of the NiO film. By using this simple model, Russo *et al.* have developed a powerful technique allowing the evaluation of the Joule heating and CF temperature T_{CF} during reset switching [69]. This technique works in the following sequence:

- 1) Firstly, the electrical resistance (R) of the CF in the LRS state is measured as a function of the device temperature (T_0); as the CF shows a metallic behavior, the LRS resistance can be described as $R(T_0) = R(T)[1 + \alpha(T_0 - T)]$, where R is the CF resistance at the reference room temperature T and α is the resistance temperature coefficient.
- 2) After solving the steady-state Fourier heat flow equation along the CF, a parabolic T_{CF} profile is calculated when an electrical current flows, with $T_{\text{CF}} = T_{\max}$ in the middle of the CF and $T_{\text{CF}} = T_0$ at the interface with the electrodes acting as heat sinks.
- 3) Then, the macroscopic CF resistance is computed by integrating the nonuniform CF resistivity (ρ_{CF}) along the CF during electrical-induced heating, which gives $R = R_0[1 + \gamma \cdot \alpha(T_{\max} - T_0)]$, where R_0 and ρ_0 are the low-field resistance and resistivity at T_0 , respectively, and $\gamma = 2/3$ is a geometrical coefficient accounting for nonuniform T and ρ along the CF. Thus, after calculating the R - V plot during a reset I - V sweep, as shown in Figure 13.4a,b, and using the extracted experimental α parameter, T_{\max} may be calculated up to the point of reset as shown in Figure 13.4c. From this technique, the device resistance thus acts as a thermometer for the local temperature of the CF, where $T_{\max} = T_0 + V^2 \cdot R_{\text{th}} / R$, with R_{th} being the equivalent thermal resistance.
- 4) Finally, the temperature $T_{\max} = T_{\text{crit}}$ is extracted at $V = V_{\text{reset}}$, where T_{crit} is considered as a physical parameter describing the stability of the CF with respect to its temperature-activated dissolution mechanism. Figure 13.4c shows that $T_{\text{crit}} \sim 600$ K for NiO cells studied by Russo *et al.* [69].

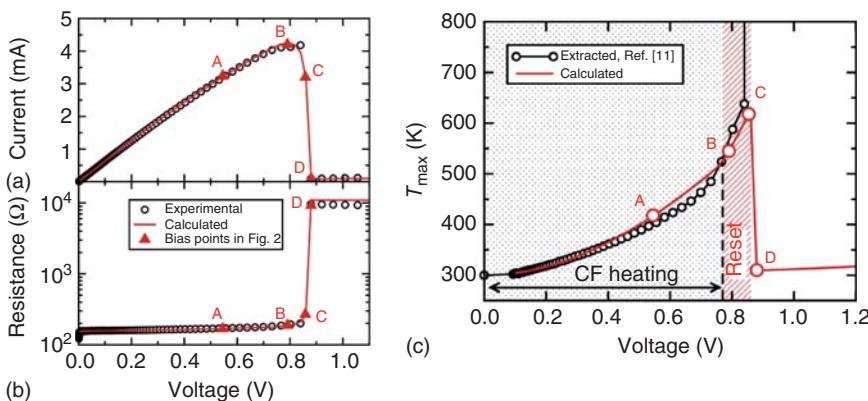


Figure 13.4 (a) Measured I - V characteristic of a NiO RRAM cell in LRS, showing reset transition at about 0.83 V. (b) Corresponding R - V characteristic, obtained from (a) as $R = V/I$. (c) Experimentally extracted

and calculated CF temperature as a function of applied voltage. Each figure shows a red curves and four bias states A–D, corresponding to the electrothermal simulations shown in Figure 13.5 Ref. [69].

Interestingly, by using the CF electrical and thermal parameters extracted with this technique, it is possible to simulate the thermal dissolution of the CF during a reset I - V sweep. Russo *et al.* developed a numerical model where the Joule heating and the temperature inside the CF were self-consistently computed in a 2-D cylindrical symmetry [70].

Figure 13.5 shows different steps of the electrothermal simulation corresponding to points A–D in Figure 13.4a–c. Since the electrodes act as heat sinks, the reset dissolution first occurs in the middle of the CF, leading to a local reduction of the CF cross section, which in turn leads to a local increase in current density and, hence, to an increased Joule heating. Therefore, the reset transient appears as a *self-accelerated process* taking place in a small voltage range from points B to D, which abruptly slows down once the electrical current collapses due to CF rupture (point D).

13.3.2.2 Modeling of the Set Switching

After reset, a large range of HRS resistances may be obtained experimentally, as only a partial CF rupture is achieved after the CF dissolution mechanism slows down due to current collapse. Therefore, HRS states are most often considered not to correspond to a return to the pristine state of the cell. As a result, while the conduction in pristine NiO material is typically attributed to hole-polaron hopping between Ni^{2+} and Ni^{3+} sites [27], the conduction in HRS states has been modeled by different mechanisms.

For example, for TiN/NiO/Ni cells, the conduction in different HRS states achieved in the course of the reset process has been modeled by the point-contact theory, revealing the lateral character of the reset process [11].

On the other hand, for W/NiO/Pt cells, Ielmini *et al.* have regarded the HRS states as doped-semiconductor states [36]. Supporting this view, the activation

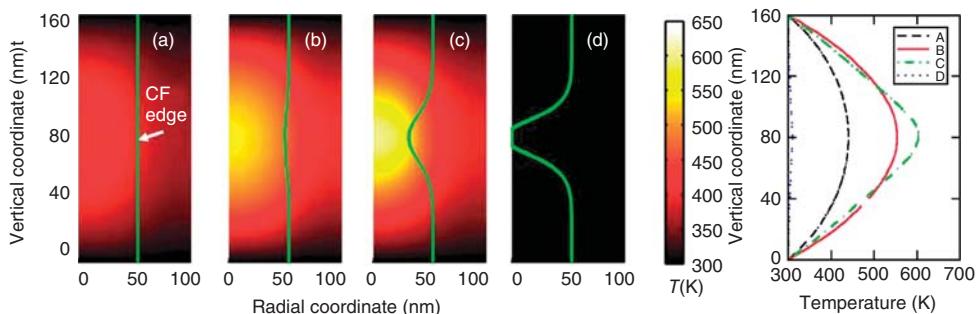


Figure 13.5 Electrothermal calculations of the temperature map along a 100 nm-wide CF formed in a 160 nm-thick NiO cell; the calculations are shown for the four bias points shown in Figure 13.4: (a) 0.53, (b)

0.78, (c) 0.85, and (d) 0.87 V. The CF edge is also shown, indicating the thermally activated shrinking effect. The temperature profile along the CF longitudinal axis is shown on the right.

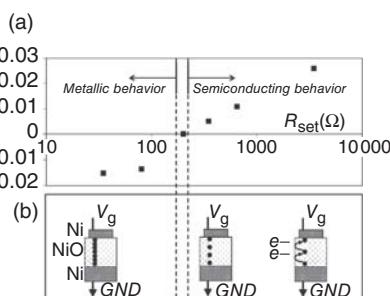
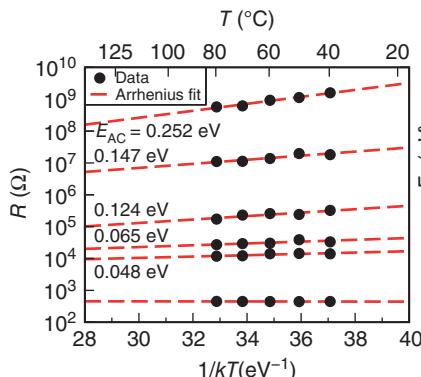


Figure 13.6 (a) Arrhenius plot of measured resistance of a W\NiO\Pt cell in different states, from full LRS ($R < 1 \text{ k}\Omega$) to full reset ($R > 1 \text{ G}\Omega$), showing the activation energy E_{AC} (or E_a) for each states. (b) E_a extractions for different LRS (or set) states obtained in a Ni\NiO\Ni cell, together with associated sketches of amounts of oxygen vacancies

created during set switching, affecting the conduction mechanism. Both experimental results indicate a metallic conduction for very low LRS resistances, while higher LRS resistance ranges and HRS states show semiconductor conduction with positive activation energy.

energy (E_{AC}) of the conduction was shown to gradually change from negative for LRS states (metallic CF) to positive for HRS states (see Figure 13.6a). The increasingly semiconductor-like behavior of the CF associated to the increase in E_{AC} with the increase in the HRS resistance (R_{HRS}) was related to the modulation of E_F with respect to the nearest conducting band edge due to gradual recovery of the defects constituting the CF. This picture is consistent with *ab initio* results presented in Figure 13.3.

As a typical conduction mechanism in defective semiconductors, the Poole–Frenkel (PF) conduction model was applied to these different semiconductor-like

states by Ielmini *et al.* [36]. Based on this modeling, the defect concentration (N_T) was extracted from measured set I - V sweeps and was shown to consistently decrease with the increase in R_{HRS} .

For the same NiO cells, Ielmini *et al.* showed by time-resolved experiments that the set switching is triggered by very fast (<1 ns) purely electronic threshold switching [71]. Similar electronic threshold switching has also been observed in chalcogenide glasses and has been explained by the high-field energy gain of carriers hopping through localized states [72].

Hence, a threshold switching model was developed based on energy gain of carriers hopping along a chain of PF defect states [36]. The energy gain at high field is quantified by the expression $E_F - E_{F0}$, where E_F and E_{F0} are the quasi-Fermi and the equilibrium FLs, respectively. The threshold set power P_T is then approximated by assuming the following critical condition for threshold switching: $E_F = E_{F0} + k_B T$, where k_B is the Boltzmann constant. The calculated P_T was found to show a fair match with experimental P_{set} data obtained in a wide range of R_{HRS} values [36]. On the other hand, as $P_T \propto R_{HRS}^{-0.5}$, the relations $I_{set} \propto R_{HRS}^{-0.75}$ and $V_{set} \propto R_{HRS}^{0.25}$ were derived, giving some guidelines on how set programming parameters may be tuned through targeting a specific R_{HRS} range. Note that the model shows increasing deviations for larger R_{HRS} ranges where the cell conduction is no more dominated by the CF conduction but by the cell leakage.

Once set switching is triggered by the electronic threshold switching, the large increase in the current density induces a strong temperature rise, which in turn increases the conduction as a positive feedback loop. During this set transient, similarly to the breakdown process discussed earlier, the severe current, field, and temperature conditions locally generate additional defects, which further increase the current and result in a structural modification of the CF (creation of V_O defects and/or metallic Ni) that is stable after power switch-off. These permanent modifications are regarded as local reduction mechanisms of the semiconductor-like material [11, 73].

As already addressed in Section 13.2, the positive feedback loop between the defect generation and the increase in the local current density necessitates careful current control during the set transient. This is achieved efficiently when the RRAM cell is integrated in a 1T1R scheme, with a MOSFET current limiter being able to react in the nanosecond time range from the threshold switching event and thus able to limit the maximum transient current (I_{max}) accurately by means of the gate potential. In the next subsection, we assume a controlled $I_{max} = I_C$ during set operation.

13.3.3

Model Implications on the Device Level

13.3.3.1 CF Size and R_{LRS} Scaling with c

Due to the filamentary nature of the switching, the conduction in LRS and HRS states is dominated by the CF conduction, except for large R_{HRS} ranges where the

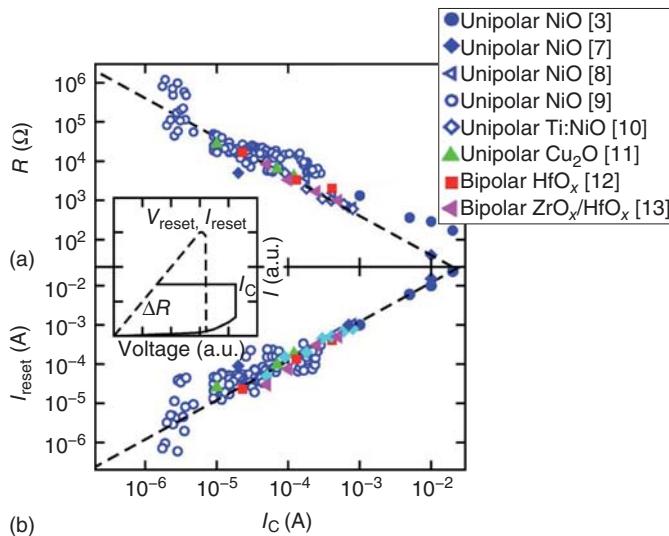


Figure 13.7 Measured (a) LRS resistance R and (b) I_{reset} as a function of I_C . Data are shown for several unipolar and bipolar RRAM devices with different active materials, including NiO, Cu₂O, HfO_x, and ZrO_x/HfO_x.

The inset shows the schematic I - V curves for (solid) unipolar set and (dashed) reset transitions, defining parameters I_C , V_{reset} , I_{reset} , and $R = R_{LRS}$ Ref. [75].

current is screened by the background uniform leakage through the whole cell area, as discussed earlier.

Assuming a single cylindrical metallic CF of diameter ϕ , $R_{LRS} = R_{CF}$ scales with ϕ^{-2} . On the other hand, a universal empirical law has been established for several materials operated in unipolar or bipolar mode, showing $R_{LRS} = V_0/I_C$, with $V_0 \sim 0.4$ V (Figure 13.7a) [74, 75], which means that I_C allows controlling ϕ and thus R_{LRS} . Note again that for large R_{LRS} ranges, the CF may no longer be assumed to be metallic (Figure 13.6) [8, 75].

13.3.3.2 I_{reset} Scaling with CF Size Scaling

In turn, Figure 13.7b shows that I_{reset} also scales with I_C as $I_{reset} \sim I_C$. From the steady-state solution of the Fourier equation for Joule heating, $I_{reset}^2 = (T_{crit} - T_0)/(R_{LRS} \cdot R_{th})$. Assuming that R_{th} is mainly controlled by conduction in the metallic CF, R_{th} is then proportional to R_{LRS} due to the Wiedemann–Franz law for metals. Thus, in order to reach the same T_{crit} , the empirical law $I_{reset} \propto R_{LRS}^{-1}$ is generally observed, which also means that V_{reset} does not vary significantly, provided R_{LRS} is in a range $> \sim 100$ Ω [76].

However, Figure 13.7b shows a deviation for larger R_{LRS} ranges ($> 10^5$ Ω). Indeed, in practice, for narrow filament, the increase in R_{th} with ϕ scaling is moderate due to the relatively large heat losses through the oxide surrounding the CF, which weakly depends on ϕ . Note that large R_{LRS} may also result in V_{reset} increase due to less efficient heating.

Hence, the reduction in I_C allows operation at lower I_{reset} . In practice, current levels in the range below 20 μA have been achieved using a 1T1R configuration [9, 77], and a few other reports demonstrated effective URS using less than 10 μA [10, 22, 78]. Further current scaling has been demonstrated by C-AFM experiments on NiO layers at nanometer scale. I_{reset} has been demonstrated to scale down to ~ 1 nA range [21]. Note, however, that no repeatable and stable switching has been demonstrated so far at the nanoampere level.

13.3.3.3 Switching Speed

So far, only quasistatic switching (in the range ~ 0.1 s) has been considered. As said earlier, the threshold switching during set occurs in the subnanosecond range, and the net switching time is generally observed to take only a few nanoseconds or less [57]. However, the microscopic origin of the reset operation is believed to be controlled by the radial diffusion of Ni or O species, leading to CF dissolution or oxidation, respectively. Hence, the reset time t_{reset} may be estimated by the time for radial diffusion through a distance ϕ , which should follow an Arrhenius dependence on temperature as $t_{\text{reset}} = \phi^2 \cdot D_0^{-1} \cdot \exp(E_{\text{AD}}/k_B T_{\text{crit}})$, where E_{AD} is the activation energy for diffusion and D_0 is the diffusion prefactor. In agreement with this dependence, Russo *et al.* observed that reset I - V sweeps carried out using faster voltage ramps resulted in larger V_{reset} [69], which is expected because larger $T_{\text{crit}} \propto T_0 + V_{\text{reset}}^2$ should be reached to allow for diffusing species to cross the same distance in shorter times. From experiments using various ramp rates, Russo *et al.* have extracted $E_{\text{AD}} = 1.1$ eV [69] by means of the Kissinger method [79].

A larger $E_{\text{AD}} \sim 2.3$ eV was reported for Ni diffusion at high temperature in NiO single crystals [68]; however, similar values are expected in the presence of GBs either for Ni [80] or O species [81], which makes sense physically due to the large structural disorder in the CF region. For the same reason, D_0 is expected to be several orders of magnitude larger compared to single crystals. In agreement with these considerations, pulse-switching investigations have reported $t_{\text{reset}} \sim 1 - 10$ μs [9, 57, 78, 82]. On the other hand, the relation $t_{\text{reset}} \propto \phi^2$, which is equivalent to the relation $t_{\text{reset}} \propto R_{\text{LRS}}^{-1}$ obtained from electrothermal simulation by Sato *et al.* [83], predicts faster reset for narrower CF assuming constant T_{crit} . Indeed, reset switching in the range < 10 ns has been achieved by Lee *et al.* for $R_{\text{LRS}} \sim 10$ k Ω , as demonstrated in Figure 13.8 [16]. Ahn *et al.* also achieved reset switching within 10 ns for $R_{\text{LRS}} \sim 100$ k Ω in Ti-doped NiO cells [22], which might also originate from the improved heating efficiency resulting from the Ti doping.

Indeed, efficient CF heating is key to achieve fast reset in the nanosecond range, and in practice, such short reset time requires significant increase in V_{reset} (~ 1 V in Figure 13.8) in order to activate the diffusion by higher T_{crit} . Note that this increase in V_{reset} is not related to the time needed for the temperature to saturate after voltage impulse, which was estimated in the range < 1 ns by electrothermal simulations [83].

However, the increase in V_{reset} for shorter pulse times has been shown by Ielmini [84] to generate reset/set instabilities (RSIs). This issue relates to the

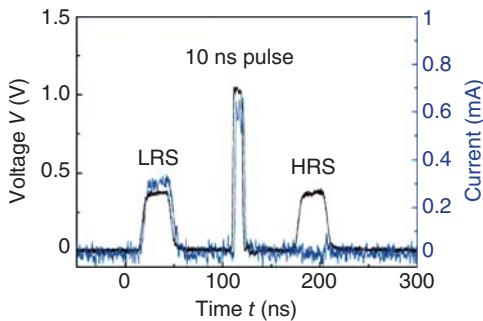


Figure 13.8 Pulse-switching experiment carried out on a $0.1 \mu\text{m}^2$ Pt/NiO/Pt cell; a monitoring pulse is input before and after a reset pulse, showing effective cell resetting for a 10 ns-long pulse of 1 V.

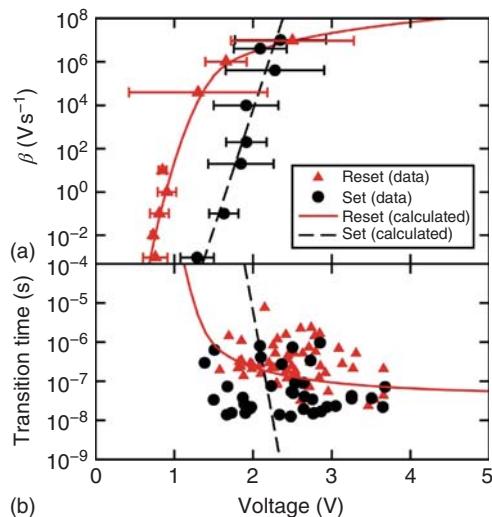


Figure 13.9 Measured and calculated relationships between the set/reset voltage at the cell and (a) the sweep rate β or (b) the set/reset transition times. The crossover between set and reset at high β /short transition times is clear.

parasitic occurrence of set switching during reset pulse [85]. Figure 13.9 shows the crossover between set and reset times as a function of the voltage amplitude. In order to avoid RSI and thus to allow a reliable operation, it is advised to use lower V_{reset} at the expense of longer pulse width [85, 86].

As shown in Figure 13.4c, reset switching in quasistatic experiments typically occurs at a temperature T_{crit} of the order of 600 K. Based on the Arrhenius dependence of $t_{\text{reset}}(T)$, the LRS state is expected to be less stable than the HRS state during accelerated retention tests at moderately elevated temperatures $T_R < 600$ K. This was confirmed by a specific investigation of the isothermal retention characteristics of NiO cells [87], from which a similar activation energy E_{AR} was extracted

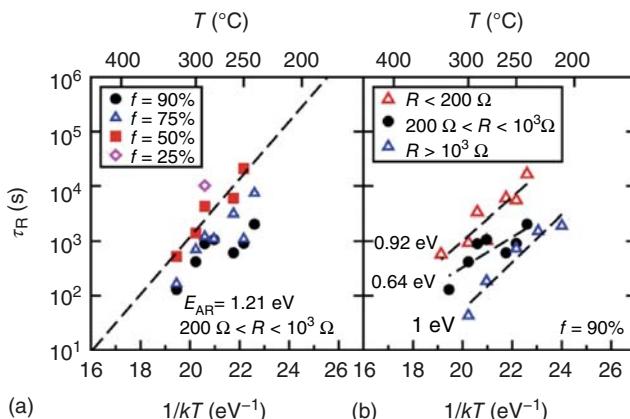


Figure 13.10 (a) Arrhenius plot of the measured retention time τ_R (defined as the time for a $\times 10$ increase of R in the average data-loss characteristics) for different percentiles, namely $f = 25, 50, 75$, and 90% , within a

fixed resistance range $0.2 < R < 1 \text{ k}\Omega$. (b) Arrhenius plot of measured τ_R for three different resistance ranges, namely $R < 0.2 \text{ k}\Omega$, $0.2 < R < 1 \text{ k}\Omega$, and $R > 1 \text{ k}\Omega$, at $f = 90\%$.

to the E_{AD} value obtained from quasistatic experiments (Figure 13.10a) [69]. On the other hand, consistently with the lower CF diameter ϕ for larger R_{LRS} , shorter retention lifetimes were achieved for larger R_{LRS} ranges (Figure 13.10b) [87]. These results are very important from a device viewpoint, because they imply that a *trade-off needs to be found between a low operating current and a long retention lifetime ($I_{reset} \propto R_{LRS}^{-1}$)*.

We note that some reports also mention less stable HRS states, in particular in the case of oxygen-deficient NiO fabricated by oxidation process [8, 88], showing that the set switching may also take place thermally. This composition-dependent effect has been observed in various oxides [88] and may be related to spontaneous V_0 -chains rearrangement of a dense V_0 network [89].

13.4

Influence of Oxide and Electrode Materials on Unipolar-Switching Mechanisms

In Section 13.3, we have presented physics-based switching models describing the URS behavior of NiO cells most often sandwiched between Pt electrodes. The models were developed considering a single cylindrical CF bridging both electrodes.

However, specific switching parameters may be significantly affected by material characteristics. For example, the switching speed depends on the microscopic kinetics of diffusing species, most often oxygen. The reset current may also be significantly modulated by the type of oxide used, as shown by Baek *et al.* [7], likely due to different thermal and electrical characteristics.

Overall, the single cylindrical CF model is, in principle, extendable to any material system. However, we describe alternative models accounting for specific material effects, in this Section 13.4. The influence of the nature and microstructure of the oxide is addressed first, followed by the impact of the electrode materials

13.4.1

Influence of the Oxide Material

13.4.1.1 The Specific Case of TiO_2

TiO_2 is a well-known model system for unipolar switching; however, TiO_2 material properties are fairly different from those of NiO . While NiO does not show any stable suboxide phase, the so-called multiple stable Magneli phases exist for O-deficient TiO_{2-x} . Although Ti interstitials may exist in TiO_{2-x} phases, V_{O} species were recently evidenced as major defects involved in the URS behavior [90]. Interestingly, Magneli phases may be formed from the rutile phase by V_{O} creation and subsequent Ti ion migration along the (121) crystallographic shear plane. As a result, *CF formation in TiO_2 can be understood as a phase transition*. By using high-resolution transmission electron microscopy (HRTEM), Kwon *et al.* clearly evidenced the Ti_4O_7 Magneli phase in LRS [91]. HRTEM images clearly revealed a *conical* CF shape instead of cylindrical [91], which in principle should move the “hot” switching spot from the bulk toward one of the electrodes. In fact, the CF disruption was evidenced at the bottom-electrode (BE) interface [91].

13.4.1.2 Influence of the Oxide Microstructure

Specific models were proposed elsewhere to describe conical CF configurations [92]. One of these models relies on the influence of the oxide microstructure. The URS property does not require a specific microstructure, for instance, URS has been shown in various amorphous materials, including, for example, Nb_2O_5 [42, 45], ZrO_2 [44], and many more materials [33]. As for NiO materials, URS has been observed both in polycrystalline and in single-crystalline layers [81, 93]. However, Park *et al.* reported lower V_{reset} variability for polycrystalline NiO layers, which they associated to the decreased E_{AD} of oxygen obtained at GBs by first-principles calculation [81]. In agreement with a preferential CF formation at GB, Lee *et al.* [16] and Son and Shin [15] observed CFs at GBs by means of HRTEM and C-AFM techniques respectively. Park *et al.* also showed the formation of Ni filament at the GB of a polycrystalline NiO films by electron energy-loss spectroscopy [62].

However, another report from Yoo *et al.* [94] suggests the generation of CF paths in the bulk, which are additional to the preformed CF paths at GBs and create new GBs during switching (Figure 13.11). As a matter of fact, these results may be understood by considering the specific property of NiO material, showing no stable suboxide phase as mentioned earlier, which thus makes this structure prone to defect percolation during CF formation [94]. A percolation-based model thus appears as a relevant approach to better describe these generated multiple-CF networks.

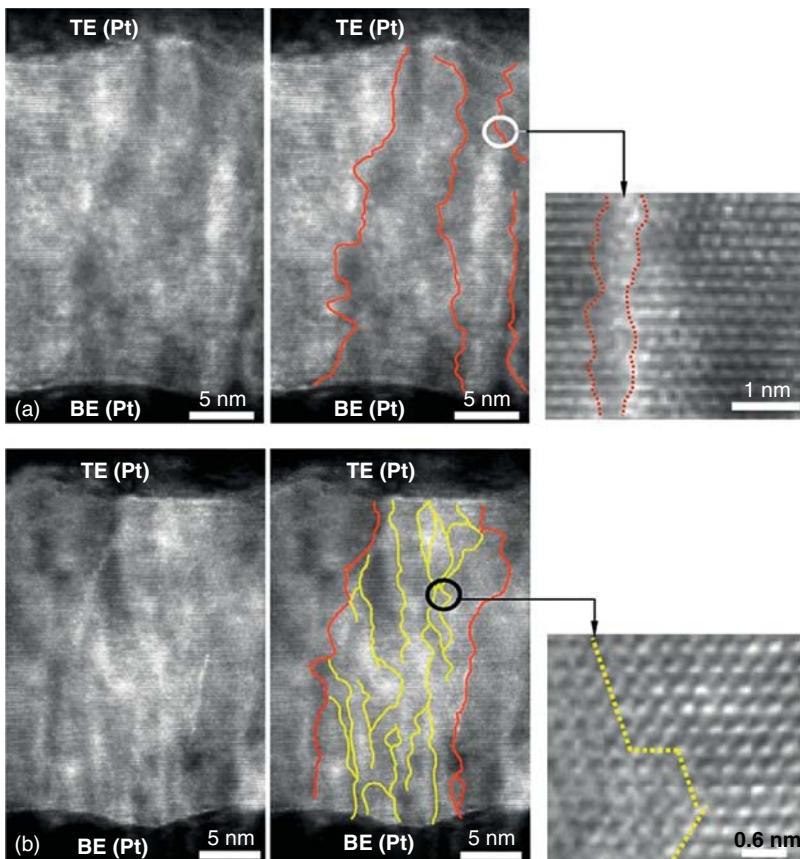


Figure 13.11 Observation of soft breakdown boundaries in NiO thin films during resistance switching. (a) A grain boundary is indicated by red lines in the virgin NiO thin film. (b) New boundaries with yellow

lines appeared after resistance switching. It is believed that conducting clusters (percolating network) are formed in both grain boundaries and breakdown boundaries during resistance switching

13.4.1.3 Random Circuit Breaker Model

The random circuit breaker (RCB) model initially proposed by Chae *et al.* [24] relies on the percolation theory, to describe the voltage-induced formation and disruption of single-CF or multiple-CF network. The RCB network consists of elementary bonds, called circuit breakers, which are initially set randomly in low- or high-resistance states and which may switch to a different state if a voltage pulse applied to the network distributes such that the local potential becomes larger than the threshold voltages for local reset/set processes. In an extended version, to better account for the thermal origin of URS, a thermal component was added so that a single breaker may change state depending on its local temperature (T_{loc}) (see Figure 13.12) [95].

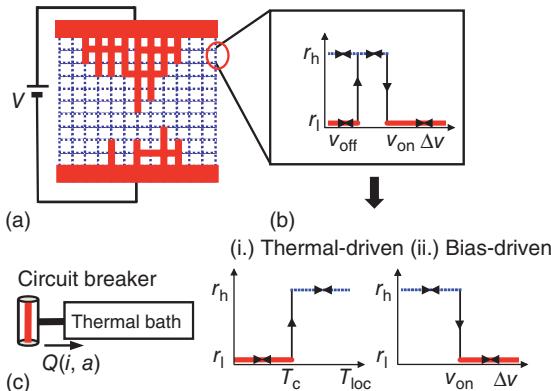


Figure 13.12 (a) Schematic diagram of a random circuit breaker (RCB) network composed of circuit breakers. The thick solid (red) and thin dotted (blue) lines represent the on and off states, respectively. (b) Field-induced switching rules, complemented in (c) by the thermal component where the heat Q of each circuit breaker is

assumed to be dissipated to a thermal bath of temperature T_b . When its local temperature T_{loc} becomes higher than a threshold temperature T_c , it will change its resistance from r_l to r_h (a thermally driven process). On the other hand, if the applied bias becomes larger than v_{on} , the state with r_h will change to r_l , independent of T_{loc} .

As a result, forming operation typically results in a percolating cluster through the cell, while reset operation typically turns off one or several circuit breakers. The morphology of the CF network critically depends on the local competition between Joule heating and thermal dissipation. This local competition successfully accounts for the occurrence of abnormal switching behavior, that is, parasitic set during reset sweep [96], which is not described by single CF models. RCB models were also used to explain the R_{LRS} dependence of V_{reset} , observed in some systems to clearly deviate from single-CF predictions [46].

13.4.1.4 Coexistence of Bipolar and Unipolar Switching

As discussed in the previous subsection, large densities of GBs provide fast-diffusion paths during resistive switching. Interestingly, Goux *et al.* have evidenced the coexistence of URS and BRS modes for NiO layers having a defect-rich microstructure [12]. While NiO is intrinsically a thermochemical material showing URS, this disordered/defective microstructure proved favorable to ionic migration and electrochemical redox mechanisms, allowing for reversible selection of any switching mode between URS and BRS [12].

A defective microstructure is not a prerequisite for a mixed URS/BRS functionality. Indeed, the forming step may suffice, as it locally generates a highly disordered network in the CF region, which is likely to act as a migration channel. For example, coexisting URS and BRS operation modes were obtained after forming operation in TiN/HfO₂/Pt stacks where a high-quality amorphous HfO₂ layer was prepared by ALD [89]. Coexisting URS and BRS properties were also evidenced in TiO₂ cells [97] as well as in perovskite systems [98].

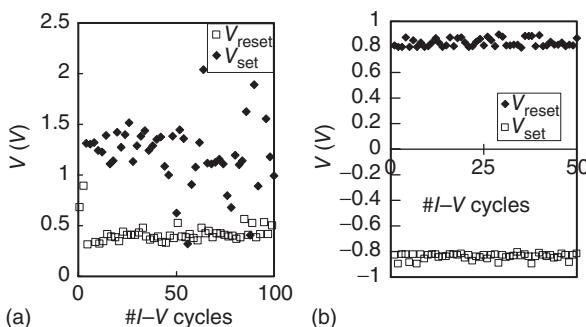


Figure 13.13 V_{set} and V_{reset} parameters extracted from consecutive unipolar (a) and bipolar (b) switching $I-V$ cycles obtained in the same Ni/NiO/Ni cell exhibiting coexisting URS and BRS switching properties; the

switching variability is clearly larger for the URS switching mode. (Reproduced with permission from Ref. [12]. © 2010, AIP Publishing LLC.)

13.4.1.5 Switching Variability and Endurance

The switching variability was investigated in various reports [9, 57]. In general, the switching-voltage distribution is wider for URS than for BRS, as illustrated in Figure 13.13. Indeed, as the reset mechanism is a self-accelerated process for URS mechanisms, it may be argued that the programming of the HRS state is difficult to control, leading to different cycle-to-cycle R_{HRS} and subsequent V_{set} parameters (Figure 13.13) [12]. Previous experiments also suggested that the variability degrades in a configuration of multiple-CF network [99]. Consistently, RCB simulations clearly suggest that the switching “hot” spot is likely to move within the network from cycle to cycle.

Hence, not surprisingly, write-endurance lifetime is clearly shorter for URS than for BRS. Several explanations were proposed for the decreased endurance, including (i) the higher temperature required for URS reset compared to BRS, possibly inducing electromigration effects [75], and (ii) the gradual loss of species involved in the radial diffusion, or along the vertical direction due to the use of a single programming polarity. Nevertheless, cycling endurance of the order of 10^6 set/reset cycles was demonstrated by Baek *et al.* (Figure 13.14) [7].

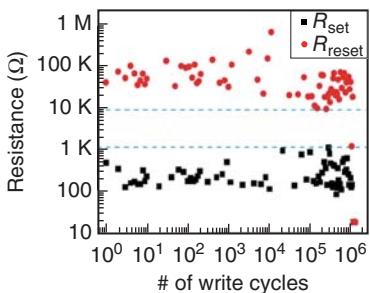


Figure 13.14 Unipolar set/reset cycling endurance obtained for $0.3 \times 0.7 \mu\text{m}^2$ cell.

Importantly, the clearly *improved cycling reliability of BRS systems has been a determining factor explaining the preferred development of BRS over URS systems during the past few years.*

13.4.2

Impacts and Roles of Electrodes

As URS is based on thermally activated diffusion and redox reaction in a MIM cell, inert metals are typically preferred as electrodes to avoid any parasitic redox reactions. In this respect, Pt has been the electrode of choice so far. The next subsection shows that other electrode materials have been tested and specifically addresses the impact of the anode material during reset operation.

13.4.2.1 Anode-Mediated Reset Operation

As said earlier, Kwon *et al.* reported a conical Ti_4O_7 CF shape in TiO_2 cells [91]. The narrow cross section of this conic CF is located at the BE interface, which was used as anode in this experiment [91]. This is commonly attributed to the migration of V_0^{2+} defects toward the cathode during forming, followed by the nucleation and growth of a so-called virtual cathode up to the anode. From this conical shape, it is expected that the Ti_4O_7 CF disrupts *at the “hot” constriction, thus closer to the anode interface*. Anode-interface-localized switching has been observed in n-type materials in general, such as Al_2O_3 [100] or TiO_2 and ZnO [101], while mixed results are reported for p-type NiO , as the CF growth direction is claimed to be reversed in this case [102]. However, some reports clearly show CF oxidation at the anode interface for NiO [11, 88].

From these results, the careful engineering of the anode interface appears critical. Goux *et al.* reported strong modulation of the R_{HRS} level induced by nanometer-thick inserted layers at the Pt–anode interface of $\text{TiN}/\text{HfO}_2/\text{Pt}$ cells [103]. Lower R_{HRS} was obtained for insert layers having large O affinity (Ti, La), which was attributed to the gettering of O-species preventing effective CF oxidation [103]. Consistently with this scenario, degraded switching was also observed for NiO cells made of easy-to-oxidize anode metals [11]. As a matter of fact, the O affinity of the anode electrode should be lower than that of the CF material, as also deduced from other experiments [46].

These results may not be solely related to the narrow CF constriction being closer to the anode, because similar results are observed for very thin oxide layers for which the cross-section variation of the CF along its length should be minor and for which the actual location of the hot spot should highly rely on the thermal properties of both electrodes.

On the other hand, different combinations of cathode and anode materials have been reported in the literature. Functional unipolar properties have been shown for symmetric $\text{Pt}/\text{NiO}/\text{Pt}$ or $\text{Ni}/\text{NiO}/\text{Ni}$ cells, however, not for $\text{TiN}/\text{NiO}/\text{TiN}$ cells [78] due to competing oxidation of TiN. Considering asymmetric $\text{TiN}/\text{HfO}_2/\text{Pt}$ [89], $\text{TiN}/\text{NiO}/\text{Pt}$, or $\text{TiN}/\text{NiO}/\text{Ni}$ [11, 78] cells, effective unipolar operation was observed only in the case where the Pt and Ni electrodes, respectively,

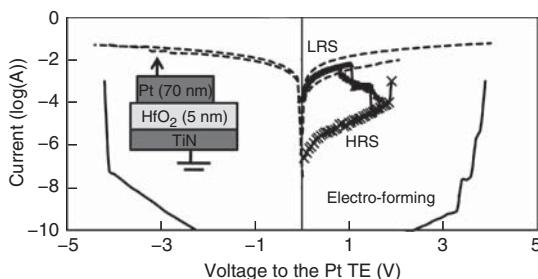


Figure 13.15 Typical switching I - V curves obtained in $\text{TiN}\backslash\text{HfO}_2\backslash\text{Pt}$ cells; effective forming is achievable independently from the voltage polarity (forming and set curves are continuous and crossed I - V traces, respectively); however, a positive voltage sweep

is required for reset (closed-symbol traces); failure to reset is observed systematically for negative polarity and only for a minor population of positive sweeps (dashed curves). (Reproduced with permission from Ref. [104]. © 2010, AIP Publishing LLC.)

were used as anode (see Figure 13.15), while no effective reset switching was observed when TiN was used as anode. These results again pointed to the critical impact of the chemistry at the interface between the oxide and the anode during reset. Moreover, from comparisons between $\text{TiN}\backslash\text{NiO}\backslash\text{Pt}$ and $\text{TiN}\backslash\text{NiO}\backslash\text{Au}$ cells, not only the inert character but also the catalytic properties of the anode material were shown to be important [11]. According to extensive electrode investigations in these systems, reset switching has been analyzed as *mediated by the catalytic properties of the anode material in an electrochemically assisted CF thermal-rupture mechanism* [11, 103].

13.4.2.2 Selection Criteria of Electrode Materials

In view of the catalyst action of the anode, Pt may be viewed as an electrode of choice, not only because of its inert chemical behavior but also because it is a well-known *H-catalyst*, which may favor the formation of more mobile and oxidizing species such as OH^- allowing efficient reset operation. A similar role could be played by Ni anodes in $\text{TiN}\backslash\text{NiO}\backslash\text{Ni}$ cells (see Figure 13.16) [11].

However, alternative anode materials may be preferred for specific reasons:

- The use of electrodes having a large R_{th} parameter improves the heating efficiency, thus resulting in a relatively low V_{reset} . This effect was observed after replacing Pt by Ti [105]. Note that the heating efficiency may also be improved by thinning the electrode [106].
- A serious drawback of Pt is the permeability to oxygen, which has a significant impact during switching. For example, after switching of $\text{Pt}\backslash\text{NiO}\backslash\text{Pt}$ cells, Lee *et al.* observed by secondary ion mass spectroscopy (SIMS) the out-diffusion of oxygen from the NiO layer into the Pt electrode [16]. In fact, due to the heat generated in the forming process, O species created by CF formation may diffuse through Pt electrodes, especially for a low oxygen partial pressure in the ambient atmosphere [104]. As a consequence, lower reset efficiency has been observed for $\text{TiN}\backslash\text{HfO}_2\backslash\text{Pt}$ cells operated in vacuum as compared to the cells

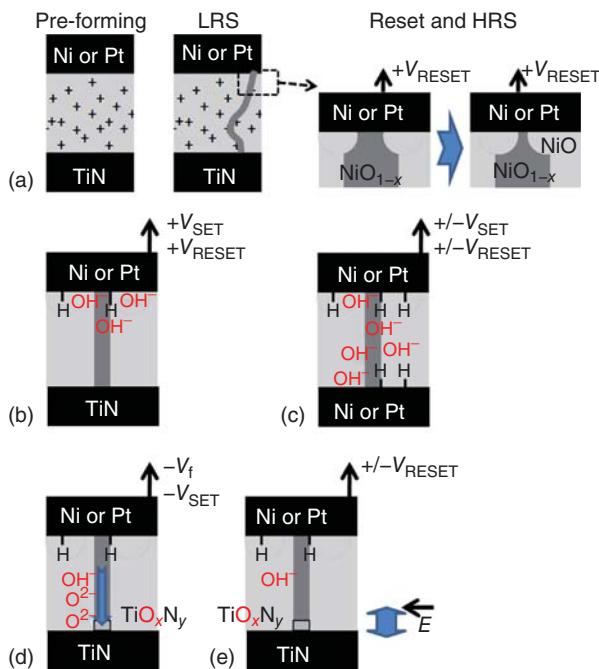


Figure 13.16 Schematics of the switching processes taking place in $\text{TiN}|\text{NiO}|\text{Ni}$ or $\text{TiN}|\text{NiO}|\text{Pt}$ cells; the reset operation is a lateral CF reoxidation at the interface with the Ni or Pt anode (a), assumed to catalytically assisted by the Ni and Pt anodes generating reactive species such as OH^- (b and c);

(d) assumed formation of a parasitic TiO_xN_y layer after negative forming sweep, resulting in field concentration and low amount of oxidizing species during reset (e). (Reproduced with permission from Ref. [11]. © 2013, AIP Publishing LLC.)

operated in air. The reset failures shown in Figure 13.15 (dashed curves) were attributed to this parasitic effect [104].

- If Ni or Cu is used as anode material, the CF is likely to result from the electrochemical injection of the anode metal into the oxide layer. During forming, Cu may indeed oxidize, migrate to the cathode, and then nucleate and grow, which is the operation principle of ECM cells. Since the mobility of Ni is also very high, some reports strongly suggest the formation of Ni-based CF after forming operation in these cells [46, 47, 61]. Even if these cells are usually operated in ECM mode, some may also exhibit functional reset in URS mode (i.e., maintaining the same voltage polarity for forming, set, and reset operations).

13.5 Conclusion

In this chapter, we described the main characteristics of URS, and we presented the main aspects of the current understanding of the switching mechanisms, particularly for NiO-based systems. Out of the various results reported on a

large number of materials, unipolar-switching mechanism is understood as a thermochemical phenomenon, taking place at the nanometer scale through oxidation/reduction processes along conductive filament(s) formed electrically in a metal/oxide/metal cell. Oxide materials and microstructure were shown to significantly affect the structure and shape of the single or multiple CF. In addition, the major role of electrodes in the switching functionality was emphasized. Particularly, there are strong indications that the thermochemical switching is electrochemically assisted at the anode interface in many asymmetric systems composed of catalytic anode materials.

From a device viewpoint, electrical properties are determined by the conductive filament, which in principle makes the device scalable to the filament size. On the other hand, the operation current was shown to be scalable through the reduction of the filament diameter, which in turn should be traded by degraded retention lifetime. Together with these scaling aspects, the unipolar character of switching makes unipolar switching unique and highly attractive for dense integration with a unidirectional diode selector. However, over the past years, bipolar switching has demonstrated greater progress and in particular improved reliability compared to unipolar switching, which is thought to originate at the very microscopic level from an improved switching control obtained when active species are mainly moved by alternate field rather than by temperature. The future of unipolar-switching technology will highly depend on how issues related to switching variability, RSI, and short endurance lifetime will be addressed by the community. The development of appropriate material stacks may bring key solutions in this respect.

References

1. Sawa, A. (2008) Resistive switching in transition metal oxides. *Mater. Today*, **11** (6), 28–36.
2. Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833–840.
3. Jeong, D.S., Thomas, R., Katiyar, R.S., Scott, J.F., Kohlstedt, H., Petraru, A., and Hwang, C.S. (2012) Emerging memories: resistive switching mechanisms and current status. *Rep. Prog. Phys.*, **75**, 076502-1–076502-31.
4. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) Redox-based resistive switching memories – nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21**, 2632–2663.
5. Kim, K.M., Jeong, D.S., and Hwang, C.S. (2011) Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook. *Nanotechnology*, **22**, 254002-1–254002-17.
6. Ielmini, D., Bruchhaus, R., and Waser, R. (2011) Thermochemical resistive switching: materials, mechanisms and scaling projections. *Phase Transitions*, **84** (7), 570–602.
7. Baek, I.G., Lee, M.S., Seo, S., Lee, M.J., Seo, D.H., Suh, D.S., Park, J.C., Park, S.O., Kim, H.S., Yoo, I.K., Chung, U.I., and Moon, J.T. (2004) Highly scalable non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. *IEDM Technical Digest*.
8. Goux, L., Lisoni, J.G., Wang, X.P., Jurczak, M., and Wouters, D.J. (2009) Optimized Ni oxidation in 80-nm contact holes for integration of forming-free and low-power Ni/NiO/Ni

- memory cells. *IEEE Trans. Electron Devices*, **56** (10), 2363–2368.
9. Sato, Y., Tsunoda, K., Kinoshita, K., Noshiro, H., Aoki, M., and Sugiyama, Y. (2008) Sub-100- μ A reset current of nickel oxide resistive memory through control of filamentary conductance by current limit of MOSFET. *IEEE Trans. Electron Devices*, **55** (5), 1185–1191.
 10. Nardi, F., Ielmini, D., Cagli, C., Spiga, S., Fanciulli, M., Goux, L., and Wouters, D.J. (2011) Control of filament size and reduction of reset current below 10uA in NiO resistance switching memories. *Solid State Electron.*, **58**, 42–47.
 11. Goux, L., Degraeve, R., Meerschaut, J., Govoreanu, B., Wouters, D.J., Kubicek, S., and Jurczak, M. (2013) Role of the anode material in the unipolar switching of TiN\NiO\Ni cells. *J. Appl. Phys.*, **113**, 054505-1–054505-9.
 12. Goux, L., Lisoni, J.G., Jurczak, M., Wouters, D.J., Courtade, L., and Muller, C. (2010) Coexistence of the bipolar and unipolar resistive-switching modes in NiO cells made by thermal oxidation of Ni layers. *J. Appl. Phys.*, **107**, 024512-1–024512-7.
 13. Govoreanu, B. *et al.* (2011) 10x10nm² Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. IEDM Technical Digest.
 14. Yo, Y.K., Kang, B.S., Park, Y.D., Lee, M.J., and Park, Y. (2008) Interpretation of nanoscale conducting paths and their control in nickel oxide (NiO) thin films. *Appl. Phys. Lett.*, **92**, 202112-1–202112-3.
 15. Son, Y. and Shin, Y.H. (2008) Direct observation of conducting filaments on resistive switching of NiO thin films. *Appl. Phys. Lett.*, **92**, 222106-1–222106-3.
 16. Lee, M.J., Han, S., Jeon, S.H., Park, B.H., Kang, B.S., Ahn, S.E., Kim, K.H., Lee, C.B., Kim, C.J., Yoo, I.K., Seo, D.H., Li, X.S., Park, J.B., Lee, J.H., and Park, Y. (2009) Electrical manipulation of nanofilaments in transition-metal oxides for resistance-based memory. *Nano Lett.*, **9** (4), 1476–1481.
 17. Yun, J.B., Kim, S., Seo, S., Lee, M.J., Kim, D.C., Ahn, S.E., Park, Y., Kim, J., and Shin, H. (2007) Random and localized resistive switching observation in Pt/NiO/Pt. *Phys. Status Solidi (RRL)*, **1** (6), 280–282.
 18. Deleruyelle, D., Dumas, C., Carmona, M., Muller, C., Spiga, S., and Fanciulli, M. (2011) Direct observation at nanoscale of resistance switching in NiO layers by conductive-atomic force microscopy. *Appl. Phys Express*, **4**, 051101.
 19. Lee, M.H. and Hwang, C.S. (2011) Resistive switching memory: observations with scanning probe microscopy. *Nanoscale*, **3**, 490–502.
 20. Ielmini, D., Spiga, S., Nardi, F., Cagli, C., Lamperti, A., Cianci, E., and Fanciulli, M. (2011) Scaling analysis of submicrometer nickel-oxide-based resistive switching memory devices. *J. Appl. Phys.*, **109**, 034506.
 21. Brivio, S., Tallarida, G., Perego, D., Franz, S., Deleruyelle, D., Muller, C., and Spiga, S. (2012) Low-power resistive switching in Au/NiO/Au nanowire arrays. *Appl. Phys. Lett.*, **101**, 223510-1–223510-4.
 22. Ahn, S.E., Lee, M.J., Park, Y., Kang, B.S., Lee, C.B., Kim, K.H., Seo, S., Suh, D.S., Kim, D.C., Hur, J., Xianyu, W., Stefanovich, G., Yin, H., Yoo, I.K., Lee, J.H., Park, J.B., Baek, I.G., and Park, B.H. (2008) Write current reduction in transition metal oxide based resistance-change memory. *Adv. Mater.*, **20**, 924–928.
 23. Ielmini, D., Nardi, F., and Cagli, C. (2010) Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories. *Appl. Phys. Lett.*, **96**, 053503-1–053503-3.
 24. Chae, S.C., Lee, J.S., Kim, S., Lee, S.B., Chang, S.H., Liu, C., Kahng, B., Shin, H., Kim, D.W., Jung, C.U., Seo, S., Lee, M.J., and Noh, T.W. (2008) Random circuit breaker network model for unipolar resistance switching. *Adv. Mater.*, **20**, 1154–1159.
 25. Kim, D.C., Seo, S., Ahn, S.E., Suh, D.S., Lee, M.J., Park, B.H., Yoo, I.K., Baek, I.G., Kim, H.J., Yim, E.K., Lee, J.E., Park, S.O., Kim, S.H., Chung,

- U.I., Moon, J.T., and Ryu, B.I. (2006) Electrical observations of filamentary conductions for the resistive memory switching in NiO films. *Appl. Phys. Lett.*, **88**, 202102.
26. Goux L. *et al.* (2013) Understanding of the intrinsic characteristics and memory trade-offs of sub- μ A filamentary RRAM operation. Symposium on VLSI Technology.
27. Jung, K., Seo, H., Kim, Y., Im, H., Hong, J., Park, J.W., and Lee, J.K. (2007) Temperature dependence of high- and low-resistance bistable states in polycrystalline NiO films. *Appl. Phys. Lett.*, **90**, 052104.
28. Hiatt, W. and Hickmott, T.W. (1965) Bistable switching in Niobium oxide diodes. *Appl. Phys. Lett.*, **6** (6), 106–108.
29. Hickmott, T.W. (1962) Low-frequency negative resistance in thin anodic oxide films. *J. Appl. Phys.*, **33**, 2669–2682.
30. Simmons, J.G. and Verderber, R.R. (1967) New conduction and reversible memory phenomena in thin insulating films. *Proc. R. Soc. London, Ser. A*, **301**, 77–102.
31. Argall, F. (1968) Switching phenomena in titanium oxide thin films. *Solid-State Electron.*, **11**, 535.
32. Gibbons, J.F. and Beadle, W.E. (1964) Switching properties of thin NiO films. *Solid State Electron.*, **7**, 785.
33. Dearnaley, G., Morgan, D.V., and Stoneham, A.M. (1970) A model for filament growth and switching in amorphous oxide films. *J. Non-Cryst. Solids*, **4**, 593–612.
34. Oxley, D.P. (1977) Electroforming, switching and memory effects in oxide thin films. *Electrocomponent Sci. Technol.*, **3**, 217–224.
35. Pagnia, H. and Sotnik, N. (1988) Bistable switching in electroformed metal-insulator-metal devices. *Phys. Status Solidi*, **108**, 11–65.
36. Ielmini, D., Cagli, C., and Nardi, F. (2011) Physical models of size-dependent nanofilament formation and rupture in NiO resistive switching memories. *Nanotechnology*, **22**, 254022.
37. Choi, B.J., Jeong, D.S., Kim, S.K., Rohde, C., Choi, S., Oh, J.H., Kim, H.J., Hwang, C.S., Szot, K., Waser, R., Reichenberg, B., and Tiedke, S. (2005) Resistive switching mechanism of TiO₂ thin films grown by atomic-layer deposition. *J. Appl. Phys.*, **98**, 033715.
38. Rohde, C., Choi, B.J., Jeong, D.S., Choi, S., Zhao, J.S., and Hwang, C.S. (2005) Identification of a determining parameter for resistive switching of TiO₂ thin films. *Appl. Phys. Lett.*, **86**, 262907.
39. Cao, X., Li, X., Gao, X., Yu, W., Liu, X., Zhang, Y., Chen, L., and Cheng, X. (2009) Forming-free colossal resistive switching effect in rare-earth-oxide Gd₂O₃ films for memristor applications. *J. Appl. Phys.*, **106**, 073723.
40. Zhu, W., Chen, T.P., Liu, Z., Yang, M., Liu, Y., and Fung, S. (2009) Resistive switching in aluminum/anodized aluminum film structure without forming process. *J. Appl. Phys.*, **106**, 093706.
41. Song, J., Inamdar, A.I., Jang, B.J., Jeon, K., Kim, Y., Jung, K., Kim, Y., Im, H., Jung, W., Kim, H., and Hong, J.P. (2010) Effects of ultrathin Al layer insertion on resistive switching performance in an amorphous aluminium oxide resistive memory. *Appl. Phys Express*, **3**, 091101.
42. Jung, K., Kim, Y., Jung, W., Im, H., Park, B., Hong, J., Lee, J., Park, J., and Lee, J.K. (2010) Electrically induced conducting nanochannels in an amorphous resistive switching niobium oxide film. *Appl. Phys. Lett.*, **97**, 233509.
43. Spiga, S., Lamperti, A., Wiemer, C., Perego, M., Cianci, E., Tallarida, G., Lu, H.L., Alia, M., Volpe, F.G., and Fanciulli, M. (2008) Resistance switching in amorphous and crystalline binary oxides grown by electron beam and atomic layer deposition. *Microelectron. Eng.*, **85**, 2414.
44. Wu, X., Zhou, P., Li, J., Chen, L.Y., Lv, H.B., Lin, Y.Y., and Tang, T.A. (2007) Reproducible unipolar resistance switching in stoichiometric ZrO₂ films. *Appl. Phys. Lett.*, **90**, 183507.
45. Spiga, S., Lamperti, A., Cianci, E., Volpe, F.G., and Fanciulli, M. (2009) Transition metal binary oxides for ReRAM applications. *ECS Trans.*, **25** (6), 411–425.
46. Lee, H.Y., Chen, P.S., Wu, T.Y., Wang, C.C., Tzeng, P.J., Lin, C.H., Chen, F.,

- Tsai, M.J., and Lien, C. (2008) Electrical evidence of unstable anodic interface in Ru/HfO_x/TiN unipolar. *Appl. Phys. Lett.*, **92**, 142911.
47. Chen, Y.Y., Pourtois, G., Adelmann, C., Goux, L., Govoreanu, B., Degreave, R., Jurczak, M., Kittl, J.A., Groeseneken, G., and Wouters, D.J. (2012) Insights into Ni-filament formation in unipolar-switching Ni/HfO₂/TiN resistive random access memory device. *Appl. Phys. Lett.*, **100**, 113513.
48. Lv, H.B., Yin, M., Song, Y.L., Fu, F., Tang, L., Zhou, P., Zhao, C.H., Tang, T.A., Chen, B.A., and Lin, Y.Y. (2008) Forming process investigation of Cu_xO memory films. *IEEE Electron Device Lett.*, **29**, 47.
49. Wu, L., Li, X., Gao, X., Zheng, R., Zhang, F., Liu, X., and Wang, Q. (2012) Unipolar resistance switching and abnormal reset behaviors in Pt/CuO/Pt and Cu/CuO/Pt structures. *Solid-State Electron.*, **73**, 11–14.
50. Inoue, I.H., Yasuda, S., Akinaga, H., and Takagi, H. (2008) Nonpolar resistance switching of metal/binary-transition-metal oxides/metal sandwiches: homogeneous/inhomogeneous transition of current distribution. *Phys. Rev. B*, **77**, 035105.
51. Shima, H., Takano, F., Akinaga, H., Tamai, Y., Inoue, I.H., and Takagi, H. (2007) Resistance switching in the metal deficient-type oxides: NiO and CoO. *Appl. Phys. Lett.*, **91**, 012901.
52. Lee, S., Kim, H., Park, J., and Yonga, K. (2010) Coexistence of unipolar and bipolar resistive switching characteristics in ZnO thin films. *J. Appl. Phys.*, **108**, 076101.
53. Yao, J., Sun, Z., Zhong, L., Natelson, D., and Tour, J.M. (2010) Resistive switches and memories from silicon oxide. *Nano Lett.*, **10**, 4105.
54. Wang, Y., Chen, Y.T., Xue, F., Zhou, F., Chang, Y.F., Fowler, B., and Lee, J.C. (2012) Memory switching properties of e-beam evaporated SiO_x on N++-Si substrate. *Appl. Phys. Lett.*, **100**, 083502.
55. Chen, Y.T., Fowler, B., Chang, Y.F., Wang, Y., Xue, F., Zhou, F., and Lee, J.C. (2013) Reduced electroforming voltage and enhanced programming stability in resistive switching of SiO₂ thin films. *ECS Solid State Lett.*, **2** (5), N18–N20.
56. Li, S.L., Li, J., Zhang, Y., Zheng, D.N., and Tsukagoshi, K. (2011) Unipolar resistive switching in high-resistivity Pr_{0.7}Ca_{0.3}MnO₃ junctions. *Appl. Phys. A*, **103**, 21–26.
57. Baek, I.G. *et al.* (2005) Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application. *IEDM Technical Digest*.
58. Peng, H.Y., Li, Y.F., Lin, W.N., Wang, Y.Z., Gao, X.Y., and Wu, T. (2012) Deterministic conversion between memory and threshold resistive switching via tuning the strong electron correlation. *Sci. Rep.*, **2**, 442, 1–6.
59. Lee, H.D., Magyari-Kope, B., and Nishi, Y. (2010) Model of metallic filament formation and rupture in NiO for unipolar switching. *Phys. Rev. B*, **81**, 193202.
60. Park, S., Ahn, H.S., Lee, C.K., Kim, H., Jin, H., Lee, H.S., Seo, S., Yu, J., and Han, S. (2008) Interaction and ordering of vacancy defects in NiO. *Phys. Rev. B*, **77**, 134103.
61. Lee, C.B., Kang, B.S., Lee, M.J., Ahn, S.E., Stefanovich, G., Xianyu, W.X., Kim, K.H., Hur, J.H., Yin, H.X., Park, Y., Yoo, I.K., Park, J.B., and Park, B.H. (2007) Electromigration effect of Ni electrodes on the resistive switching characteristics of NiO thin films. *Appl. Phys. Lett.*, **91**, 082104.
62. Park, G.S., Li, X.S., Kim, D.C., Jung, R.J., Lee, M.J., and Seo, S. (2007) Observation of electric-field induced Ni filament channels in polycrystalline NiO film. *Appl. Phys. Lett.*, **91**, 222103.
63. Seo, S., Lee, M.J., Seo, D.H., Jeoung, E.J., Suh, D.S., Joung, Y.S., Yoo, I.K., Hwang, I.R., Kim, S.H., Byun, I.S., Kim, J.S., Choi, J.S., and Park, B.H. (2004) Reproducible resistance switching in polycrystalline NiO films. *Appl. Phys. Lett.*, **85** (23), 5655.
64. Yoshida, C., Kinoshita, K., Yamasaki, T., and Sugiyama, Y. (2008) Direct observation of oxygen movement during

- resistance switching in NiO/Pt film. *Appl. Phys. Lett.*, **93**, 042106.
65. Park, J.W., Park, J.W., Jung, K., Yang, M.K., and Lee, J.K. (2006) Influence of oxygen content on electrical properties of NiO films grown by rf reactive sputtering for resistive random-access memory applications. *J. Vac. Sci. Technol. B*, **24** (5), 2205–2208.
66. Buh, G.H., Hwang, I., and Park, B.H. (2009) Time-dependent electroforming in NiO resistive switching devices. *Appl. Phys. Lett.*, **95**, 142101.
67. Calka, P., Martinez, E., Lafond, D., Minoret, S., Tirano, S., Detlefs, B., Roy, J., Zegenhagen, J., and Guedj, C. (2011) Origin of resistivity change in NiO thin films studied by hard x-ray photoelectron spectroscopy. *J. Appl. Phys.*, **109**, 124507.
68. O'Keeffe, M. and Moore, W.J. (1961) Diffusion of oxygen in single crystals of nickel oxide. *J. Phys. Chem.*, **65**, 1438.
69. Russo, U., Ielmini, D., Cagli, C., and Lacaita, A.L. (2009) Filament conduction and reset mechanism in NiO-based resistive-switching memory (RRAM) devices. *IEEE Trans. Electron Devices*, **56** (2), 186–192.
70. Russo, U., Ielmini, D., Cagli, C., and Lacaita, A.L. (2009) Self-accelerated thermal dissolution model for reset programming in unipolar resistive-switching memory (RRAM) devices. *IEEE Trans. Electron Devices*, **56** (2), 193–200.
71. Ielmini, D., Cagli, C., and Nardi, F. (2009) Resistance transition in metal oxides induced by electronic threshold switching. *Appl. Phys. Lett.*, **94**, 063511.
72. Ielmini, D. (2008) Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses. *Phys. Rev. B*, **78**, 035308.
73. Bocquet, M., Deleruyelle, D., Muller, C., and Portal, J.M. (2011) Self-consistent physical modeling of set/reset operations in unipolar resistive-switching memories. *Appl. Phys. Lett.*, **98**, 263507.
74. Fantini, A., Wouters, D.J., Degraeve, R., Goux, L., Pantisano, L., Kar, G., Chen, Y.Y., Govoreanu, B., Kittl, J.A., Altimime, L., and Jurczak, M. (2012) Intrinsic switching behavior in HfO₂ RRAM by fast electrical measurements on novel 2R test structures. IMW Workshop.
75. Ielmini, D., Nardi, F., and Cagli, C. (2011) Universal reset characteristics of unipolar and bipolar metal-oxide RRAM. *IEEE Trans. Electron Devices*, **58** (10), 3246–3253.
76. Long, S., Cagli, C., Ielmini, D., Liu, M., and Suñé, J. (2011) Reset statistics of NiO-based resistive switching memories. *IEEE Electron Device Lett.*, **32** (11), 1570–1572.
77. Kinoshita, K., Tsunoda, K., Sato, Y., Noshiro, H., Yagaki, S., Aoki, M., and Sugiyama, Y. (2008) Reduction in the reset current in a resistive random access memory consisting of NiO_x brought about by reducing a parasitic capacitance. *Appl. Phys. Lett.*, **93**, 033506.
78. Goux, L. et al. (2011) Evidences of anodic-oxidation reset mechanism in TiN\NiO\Ni RRAM cells. VLSI Technology Symposium.
79. Kissinger, H.E. (1957) Reaction kinetics in differential thermal analysis. *Anal. Chem.*, **29**, 1702.
80. Karakasidis, T. and Meyer, M. (1997) Grain-boundary diffusion of cation vacancies in nickel oxide: a molecular-dynamics study. *Phys. Rev. B*, **55** (20), 13853.
81. Park, C., Jeon, S.H., Chae, S.C., Han, S., Park, B.H., Seo, S., and Kim, D.W. (2008) Role of structural defects in the unipolar resistive switching characteristics of Pt/NiO/Pt structures. *Appl. Phys. Lett.*, **93**, 042102.
82. Kügeler, C., Weng, R., Schroeder, H., Symanczyk, R., Majewski, P., Ufert, K.D., Waser, R., and Kund, M. (2010) Study on the dynamic resistance switching properties of NiO thin films. *Thin Solid Films*, **518**, 2258–2260.
83. Sato, Y., Kinoshita, K., Aoki, M., and Sugiyama, Y. (2007) Consideration of switching mechanism of binary metal oxide resistive junctions using a thermal reaction model. *Appl. Phys. Lett.*, **90**, 033503.

84. Ielmini, D. (2010) Reset-set instability in unipolar resistive-switching memory. *IEEE Electron Device Lett.*, **31** (6), 552–554.
85. Nardi, F., Cagli, C., Spiga, S., and Ielmini, D. (2011) Reset instability in pulsed-operated unipolar resistive-switching random access memory devices. *IEEE Electron Device Lett.*, **32** (6), 719–721.
86. Kim, D.K., Suh, D.S., and Park, J. (2010) Pulse-programming instabilities of unipolar-type NiOx. *IEEE Electron Device Lett.*, **31** (6), 600–602.
87. Ielmini, D., Nardi, F., Cagli, C., and Lacaita, A.L. (2010) Size-dependent retention time in NiO-based resistive-switching memories. *IEEE Electron Device Lett.*, **31** (4), 353–355.
88. Kinoshita, K., Tamura, T., Aoki, M., Sugiyama, Y., and Tanaka, H. (2006) Bias polarity dependent data retention of resistive random access memory consisting of binary transition metal oxide. *Appl. Phys. Lett.*, **89**, 103509.
89. L. Goux, Y. Y. Chen, L. Pantisano, X. P. Wang, G. Groeseneken, M. Jurczak and D. J. Wouters, On the gradual unipolar and bipolar resistive switching of TiN\HfO₂\Pt memory systems, *Electrochem. Solid-State Lett.*, **13**, 6, G54-G56, 2010.
90. Jeong, D.S., Schroeder, H., Breuer, U., and Waser, R. (2008) Characteristic electroforming behavior in Pt/TiO₂ /Pt resistive switching cells depending on atmosphere. *J. Appl. Phys.*, **104**, 123716.
91. Kwon, D.H., Kim, K.M., Jang, J.H., Jeon, J.M., Lee, M.H., Kim, G.H., Li, X.S., Park, G.S., Lee, B., Han, S., Kim, M., and Hwang, C.S. (2010) Atomic structure of conducting nanofilaments in TiO₂ resistive switching memory. *Nat. Nanotechnol.*, **5**, 148–153.
92. Kim, K.M. and Hwang, C.S. (2009) The conical shape filament growth model in unipolar resistance switching of TiO₂ thin film. *Appl. Phys. Lett.*, **94**, 122109.
93. Kawai, M., Ito, K., and Shimakawa, Y. (2009) Resistance switching in a single-crystalline NiO thin film grown on a Pt0.8Ir0.2 electrode. *Appl. Phys. Lett.*, **95**, 012109.
94. Yoo, I.K., Kang, B.S., Ahn, S.E., Lee, C.B., Lee, M.J., Park, G.S., and Li, X.S. (2010) Fractal dimension of conducting paths in nickel oxide (NiO) thin films during resistance switching. *IEEE Trans. Nanotechnol.*, **9** (2), 131–133.
95. Chang, S.H., Lee, J.S., Chae, S.C., Lee, S.B., Liu, C., Kahng, B., Kim, D.W., and Noh, T.W. (2009) Occurrence of both unipolar memory and threshold resistance switching in a NiO film. *Phys. Rev. Lett.*, **102**, 026801.
96. Liu, C., Chae, S.C., Lee, J.S., Chang, S.H., Lee, S.B., Kim, D.W., Jung, C.U., Seo, S., Ahn, S.E., Kahng, B., and Noh, T.W. (2009) Abnormal resistance switching behaviours of NiO thin films: possible occurrence of both formation and rupturing of conducting channels. *J. Phys. D: Appl. Phys.*, **42**, 015506.
97. Jeong, D.S., Schroeder, H., and Waser, R. (2007) Coexistence of bipolar and unipolar resistive switching behaviors in a Pt/TiO₂/Pt stack. *Electrochem. Solid-State Lett.*, **10** (8), G51–G53.
98. Sun, X., Li, G., Zhang, X., Ding, L., and Zhang, W. (2011) Coexistence of the bipolar and unipolar resistive switching behaviours in Au/SrTiO₃/Pt cells. *J. Phys. D Appl. Phys.*, **44**, 125404.
99. Kim, Y.S., Kim, J.S., Choi, J.S., Hwang, I.R., Hong, S.H., Kang, S.O., and Park, B.H. (2011) Resistive switching behaviors of NiO films with controlled number of conducting filaments. *Appl. Phys. Lett.*, **98**, 192104.
100. Kim, K.M., Choi, B.J., Koo, B.W., Choi, S., Jeong, D.S., and Hwang, C.S. (2006) Resistive switching in Pt/Al₂O₃/TiO₂/Ru stacked structures. *Electrochem. Solid-State Lett.*, **9** (12), G343–G346.
101. Kinoshita, K., Okutani, T., Tanaka, H., Hinoki, T., Yazawa, K., Ohmi, K., and Kishida, S. (2010) Opposite bias polarity dependence of resistive switching in n-type Ga-doped-ZnO and p-type NiO thin films. *Appl. Phys. Lett.*, **96**, 143505.
102. Kim, K.M., Choi, B.J., Song, S.J., Kim, G.H., and Hwang, C.S. (2009) Filamentary resistive switching localized at cathode interface in NiO thin films. *J. Electrochem. Soc.*, **156** (12), G213–G216.

103. Goux, L., Wang, X.P., Chen, Y.Y., Pantisano, L., Jossart, N., Govoreanu, B., Kittl, J.A., Jurczak, M., Altimime, L., and Wouters, D.J. (2011) Roles and effects of TiN and Pt electrodes in resistive-switching HfO₂ systems. *Electrochem. Solid-State Lett.*, **14** (6), H244–H246.
104. Goux, L., Czarnecki, P., Chen, Y.Y., Pantisano, L., Wang, X.P., Degraeve, R., Govoreanu, B., Jurczak, M., Wouters, D.J., and Altimime, L. (2010) Evidences of oxygen-mediated resistive-switching mechanism in TiN\HfO₂\Pt cells. *Appl. Phys. Lett.*, **97**, 243509.
105. Ryu, S.W., Ahn, Y.B., Kim, H.J., and Nishi, Y. (2012) Ti-electrode effects of NiO based resistive switching memory with Ni insertion layer. *Appl. Phys. Lett.*, **100**, 133502.
106. Chang, S.H., Chae, S.C., Lee, S.B., Liu, C., Noh, T.W., Lee, J.S., Kahng, B., Jang, J.H., Kim, M.Y., Kim, D.W., and Jung, C.U. (2008) Effects of heat dissipation on unipolar resistance switching in Pt/NiO/Pt capacitors. *Appl. Phys. Lett.*, **92**, 183507.

14

Modeling the VCM- and ECM-Type Switching Kinetics

Stephan Menzel and Ji-Hyun Hur

14.1

Introduction

This chapter deals with the state-of-the-art modeling of VCM- and ECM-type devices. A suitable ReRAM model needs to cover self-consistently a variety of switching phenomena: (i) current–voltage characteristics, (ii) complementary (resistive) switching behavior, and (iii) nonlinear switching kinetics. The latter means that the switching time decreases exponentially by increasing the applied voltage. As discussed in Chapter 11, the nonlinear switching kinetics is the origin of some universal switching characteristics: (i) $R_{\text{LRS}} \propto I_{\text{SET}}^{-1}$ and (ii) $I_{\text{RESET}} \propto I_{\text{SET}}$. Furthermore, a very high nonlinearity of switching kinetics is required to achieve the goal of 10-year retention under constant read voltage stress and fast switching times as low as 10 ns. Thus, the modeling of the nonlinear switching kinetics is one of the most crucial aspects in developing ReRAM models. Therefore, this chapter focuses on the modeling of the nonlinear switching kinetics in ECM and VCM cells.

Based on a brief description of the microscopic switching model for VCM (Section 14.2) and ECM cells (Section 14.3), the physical processes leading to such highly nonlinear switching kinetics are discussed in Section 14.5. Subsequently, the state-of-the-art in VCM and ECM modeling is presented in Sections 14.6 and 14.7, respectively.

14.2

Microscopic Switching Mechanism of VCM Cells

VCM cells consist of a mixed ionic–electronic conducting (MIEC) insulating (I)-layer sandwiched between two metal electrodes. Typically, one electrode consists of a high work function metal (active electrode), and thus, a Schottky-type interface forms in contact with the I-layer. The counter electrode forms an ohmic interface. But, identical electrodes have been used as well. The resistive switching

relies on the movement of oxygen vacancies (ions) within the I-layer and a succeeding redox reaction in the cation sublattice [1–4]. This can influence the cell resistance in different ways. First, the local conductivity/electron transport mechanism of the material is changed. Second, oxygen vacancies act as mobile donors. If the oxygen-vacancy concentration at a metal/insulator interface is changed, the electrostatic barrier is altered according to the Schottky effect [1, 2, 5]. Depending on the MIEC material, the movement of oxygen vacancies might also lead to phase transformation as in TiO_x -based ReRAMs [6]. If a phase transition occurs, a nucleation step is required before the new phase can grow.

Prior to stable switching, an electroforming step has to be performed as discussed in Chapter 10. During this step, the oxide material is reduced, resulting in a better conducting state, which may be the HRS or LRS. Electroforming is regarded as a complex redox reaction of simultaneously occurring processes [1, 7, 8]. Oxygen is extracted from the insulating layer at the anode, leaving oxygen vacancies behind. This oxidation reaction manifests either in the oxidation of the anode metal or in the release of gaseous oxygen. The corresponding reduction reaction takes place by the local valence change in the cation sublattice. The oxygen vacancies migrate through the insulating layer in the direction of the electric field and accumulate at the cathode interface. The resulting oxygen-deficient region is well-conducting and acts as a virtual cathode. This typically filamentary region grows toward the anode. For HfO_x -based VCM cells, an alternative model is proposed in which oxygen vacancies are formed by breaking the $\text{Hf}-\text{O}$ bonds within the bulk of the oxide layer [9–11]. This process creates a Frenkel defect pair consisting of an interstitial oxygen ion and an oxygen vacancy. In this model, the interstitial oxygen ion is mobile, whereas the oxygen vacancy is assumed to be immobile. In the end, an oxygen-deficient phase is formed. However, it has been shown by theoretical calculations that these types of Frenkel defects are thermodynamically unstable in HfO_2 ; the oxygen ion would relax to its equilibrium position kinetically unhindered within a timescale of phonon frequencies [12]. Moreover, it was shown that it is thermodynamically more favorable to form an oxygen vacancy at an Hf/HfO_2 interface rather than in the bulk [12, 13].

After the forming process, two different proposed switching mechanisms can be discriminated. In the first one, the SET process is considered to be a mere completion of what happens during the electroforming of the filament employing the same microscopic processes. In this case, the filament is oxidized during the RESET operation and a reduction reaction occurs at the cathode. This means that the amount of oxygen vacancies within the film varies during switching by oxygen exchange via the active electrode [10, 14, 15]. The second proposed mechanism relies on the redistribution of oxygen vacancies next to the active electrode during the switching process [1, 2, 16]. This movement alters the interface transport and thus the resistance state. In this case, the total amount of oxygen vacancies stays constant.

14.3

Microscopic Switching Mechanism of ECM Cells

ECM cells consist of one active silver or copper electrode, an ion-conducting insulating layer, and an inert electrode [17, 18]. Thereby, the ion-conducting layer could be a solid electrolyte or a metal oxide. An important variant of ECM switching is the gap-type switch [19], which is described in detail in Chapter 18. The resistive-switching mechanism in ECM cells is based on the electrochemical growth/dissolution of an Ag/Cu filament within the insulating layer (cf. Chapter 17) [17, 18, 20]. The SET process consists of dissolution of the active anodic Ag or Cu electrode, subsequent ion migration, and reduction at the cathode, resulting in the growth of the metallic filament toward the anode. As soon as an electronic contact is achieved, the cell resistance drops to the LRS. The electronic contact can be achieved by the onset of electron tunneling between the filament and the active electrode. By tuning the current limitation during the SET operation, different LRS can be programmed [21–23]. These different resistance states can be attributed to different tunneling gaps between the filament tip and the active electrode [24]. When the current compliance is further increased, the gap can be shortened, establishing a galvanic contact with a lower LRS resistance. Further LRS resistance reduction might result from a filament diameter widening (cf. Chapter 11) [24]. To RESET the cell back to the HRS state, a negative voltage must be applied to the active electrode and the filament is dissolved. Depending on the physical nature of the LRS, the dissolution of the filament takes place in slightly differing ways. If a tunneling gap remains in the LRS, the RESET operation involves the same processes as the SET operation, only the role of the electrodes is interchanged – the inert electrode becomes the anode and the active electrode the cathode. Of course, a nucleation step is not required for the RESET operation. If a galvanic contact is achieved, the RESET switching becomes polarity-independent. Due to the processes involved in the ECM switching, the ECM cell can also be described as a nanobattery. During switching, an electromotive force (EMF) that results from the inherent nonequilibrium states evolves [25]. In fact, nonequilibrium states are also present in VCM cells, but the electronic conductivity is too high to observe a pronounced EMF.

The described growth mode from cathode to anode is the one occurring in conventional ECM cells. But, depending on the ion mobility and the speed of the redox reactions [26], the growth mode changes. The conventional growth mode occurs if the mobility is high and the cations can reach the cathode where they are reduced. If the mobility is low, a reduction can occur within the thin film, which leads to either a reconnection to the active electrode and thus a filamentary growth from the anode to the cathode or a filament growth from the active electrode with discrete nanoclusters [26]. Hereby, the metallic nanoclusters formed within the thin film work as a bipolar electrode, that is, oxidation and reduction occur – potentially with different rates – at opposing sides of this electrode.

14.4

Classification of Simulation Approaches

To model the VCM or ECM switching behavior, different approaches with respect to length and timescale have been used. These different approaches are reviewed in the following sections and the expected results are discussed.

14.4.1

Ab initio and Molecular Dynamics Simulation Models

On the atomic scale, *ab initio*, that is, density functional theory (DFT) and molecular dynamics (MD), simulation methods are used. Due to their high computational cost, only small volumes can be simulated. For *ab initio* simulations, a certain atomic structure (e.g., crystal structure) is considered. In MD methods, the interaction between atoms is described by a force field that can be either deduced from DFT calculation (*ab initio* MD) or empirically chosen. The use of this force field reduces the computational cost such that larger ensembles can be calculated. The accuracy of this method depends greatly on the selection of accurate force fields.

While it is computationally too expensive to simulate the complete switching process using DFT and MD methods, they can be used to address individual problems such as calculation of the electronic structure in ON and OFF state or of migration barriers. Quite often, DFT calculations are simulations of the ground state of a specific system, that is, $T = 0\text{ K}$.

14.4.2

Kinetic Monte Carlo Simulation Models

Kinetic Monte Carlo (KMC) simulation methods offer to link atomistic with continuum methods. In KMC, the occurring physical processes are described by rate equations of the form

$$\Gamma = \Gamma_0 \exp\left(\frac{\Delta W - \beta V}{k_B T}\right), \quad (14.1)$$

where ΔW is the activation energy of the specific process and β indicates how much this barrier is lowered by the local voltage drop V . The prefactor Γ_0 is often determined by the attempt frequency f and is in the range of $10^{12} - 10^{14}\text{ Hz}$. For each process, the occurrence probability Γ is calculated. In each time step, the occurring process is then chosen using a random number. Afterward, the system is updated according to its new configuration. Due to the statistical nature of the KMC method, it can be used to study the variability of resistive-switching or failure mechanism.

14.4.3

Continuum Models

In continuum models, partial differential equations that describe the device behavior in a computational domain, for example, a 3D VCM cell, are solved.

To model the electrical behavior of a ReRAM cell, Poisson equation coupled to the drift–diffusion equations of the charge carriers or a current continuity equation is solved. Local Joule heating is accounted for by solving the heat transfer equation using the electrical power as heat source. To solve the partial differential equations, the computational domain is discretized using finite element method (FEM), finite difference method (FDM), or finite volume method (FVM). While continuum models can be used to simulate the complete device behavior, they rely on the assumption of material parameters. In addition, they bear no information of the actual atomic configurations on the nanoscale. For example, the amount of electronic carriers is described with a concentration that is continuous. If one considers the small volume in which resistive switching takes place, one atom less or more would lead to a huge concentration change.

14.4.4

Compact Models

To enable circuit design, very simple 1D models are required that model the dynamic behavior of the device accurately. These models can be motivated by physical insights gained from experiment and more complex simulation models or by empirically fitting the experimental data using optimized mathematical equations. A compact model can be also used as a first validation of a physical model. It can be easily tested if the basic electrical properties can be covered by the physical model.

14.5

General Considerations of the Physical Origin of the Nonlinear Switching Kinetics

Based on the microscopic mechanism of VCM and ECM cells, different electrochemical and physical processes that can limit the switching speed can be identified: (i) migration of ions, (ii) electron-transfer (redox) reactions taking place at the boundaries, (iii) nucleation of a new phase, or (iv) breaking of M–O bonds. The latter process, however, is, strictly speaking, only valid for strongly covalent bonds as in SiO. For ionic crystals, “bond breaking” has to be regarded rather as pure ion movement. For the sake of convenience, the mathematical equations describing these phenomena are repeated here, and for a detailed description, the reader is referred to Chapters 5 and 9. The motion of ions can be mathematically described by the Mott–Gurney law for ion hopping

$$j_{\text{hop}} = 2zeacf \exp\left(-\frac{\Delta W_{\text{hop}}}{k_B T}\right) \sinh\left(\frac{aze}{2k_B T}E\right), \quad (14.2)$$

where z is the charge number, that is, +2 for an oxygen vacancy; e is the elementary charge; c is the ion concentration; a is the hopping distance; f is the attempt frequency; k_B is the Boltzmann constant; and ΔW_{hop} is the hopping barrier. The second rate-limiting step can be attributed to redox reactions occurring at the metal/MIEC interfaces, for example, oxidation of an oxidizable electrode.

The net current across such an interface due to the involved charge transfer can be described by the Butler–Volmer reaction

$$j_{\text{et}} = j_{0,\text{et}} \left[\exp \left(\frac{(1-\alpha)ze}{k_B T} \Delta\varphi_{\text{et}} \right) - \exp \left(-\frac{\alpha ze}{k_B T} \Delta\varphi_{\text{et}} \right) \right], \quad (14.3)$$

which is driven by the overpotential $\Delta\varphi_{\text{et}}$. If $\Delta\varphi_{\text{et}} > 0$, the oxidation reaction dominates, which is described by the left term in Eq. (14.3). In contrast, the reduction reaction outweighs if $\Delta\varphi_{\text{et}} < 0$. In Eq. (14.3), α denotes the charge-transfer coefficient. The exchange current density $j_{0,\text{et}}$ depends exponentially on the activation energy ΔW_{et}

$$j_{0,\text{et}} \propto \exp \left(-\frac{\Delta W_{\text{et}}}{k_B T} \right). \quad (14.4)$$

If a new phase is created during resistive switching, first, a stable nucleus needs to be formed. To establish a stable nucleus, a critical size is required, which is related to a critical number of atoms N_c . The nucleation time can be calculated according to

$$t_{\text{nuc}} = t_{0,\text{nuc}} \exp \left(\frac{\Delta W_{\text{nuc}}}{k_B T} \right) \exp \left(-\frac{(N_c + \alpha)ze}{k_B T} \Delta\varphi_{\text{nuc}} \right) \quad (14.5)$$

and depends on the nucleation overpotential $\Delta\varphi_{\text{nuc}}$, nucleation prefactor $t_{0,\text{nuc}}$, and nucleation energy ΔW_{nuc} . The fourth reported rate-limiting process is the M–O bond breaking. The corresponding rate law is expressed as

$$G = G_0 \exp \left(-\frac{\Delta W_{\text{bb}} - bE}{k_B T} \right), \quad (14.6)$$

where G_0 is a prefactor; ΔW_{bb} is the activation energy for bond breaking; and b is the bond-breaking coefficient, which depends on the polarizability of the insulating material. Despite their different physical and electrochemical nature, all of these processes are related to an activation barrier and thus obey an Arrhenius-type law. Hence, all processes can be enhanced exponentially when the local temperature increases due to local Joule heating. In addition, the activation barrier can be lowered by an applied electric field/voltage, and thus, an exponential dependence results when the electric field is high enough. Following the analysis by Menzel *et al.* [27], the different processes result in different slopes in a semi-logarithmic switching time $\ln(t_{\text{sw}})$ versus switching voltage (V_{sw}) diagram due to the physical parameters involved. In a pulse experiment, the slope is related to the coefficients in the exponential term of the rate-limiting equation. For voltage acceleration, the slopes read

$$\begin{aligned} m_{\text{hop}} &= -\frac{\alpha ze}{2k_B Tw}, \quad m_{\text{red}} = -\frac{\alpha ze}{k_B T}, \quad m_{\text{ox}} = -\frac{(1-\alpha)ze}{k_B T}, \\ m_{\text{nuc}} &= -\frac{(N_c + \alpha)ze}{k_B T}, \quad \text{and} \quad m_{\text{gen}} = -\frac{b}{k_B Tw}. \end{aligned} \quad (14.7)$$

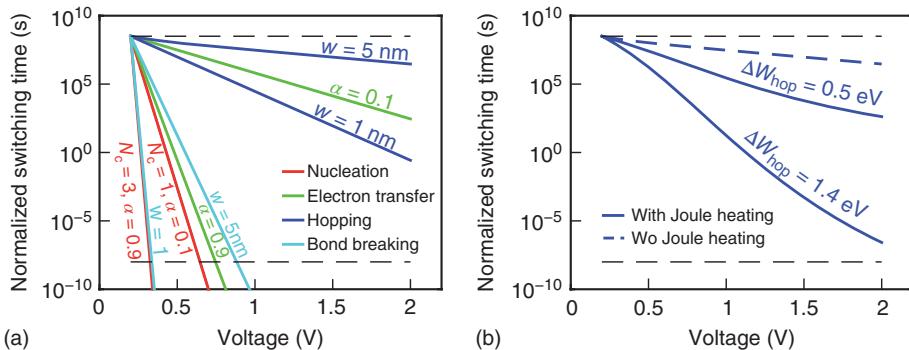


Figure 14.1 (a) Illustration of the nonlinearity in the switching kinetics obtained for electric-field-enhanced nucleation (red), electron-transfer reaction (green), ion migration (blue), and bond breaking (cyan) for different limiting cases at $T = 300\text{ K}$ and $a = 0.3\text{ nm}$. (b) Illustration of

the nonlinearity calculated for ion migration limitation without Joule heating (blue dashed line) and with Joule heating assuming a linear I - V relation with $R = 1\text{ k}\Omega$ and $R_{th} = 7.5 \cdot 10^5\text{ KW}^{-1}$ for different activation energies.

The strength of the electric field depends strongly on the width of the region w where the applied voltage drops according to $E = V/w$. Using physically reasonable parameters, the calculated slopes lie within a certain range as illustrated in Figure 14.1. Thus, the highest nonlinearity would be achieved by bond breaking followed by nucleation, electron transfer, and ion hopping. The slope for bond breaking or nucleation, however, is so steep that it is very likely that this process limits the switching speed in only a very small voltage regime. In contrast, a process with a flat slope limits the switching process very likely at higher voltages.

In addition to the field acceleration, temperature acceleration is also possible due to the Arrhenius type of law. In a filamentary-switching mechanism, the temperature increase can be approximated by

$$T = T_0 + R_{th}P_{el}, \quad (14.8)$$

where T_0 is the ambient temperature, P_{el} is the electrical power dissipated in the VCM cell, and R_{th} is the thermal resistance. The influence of the local Joule heating on the switching kinetics is illustrated in Figure 14.1b. Here, the slope strongly depends on the activation energy of the ion migration and the amount of Joule heating. The higher the activation energy, the higher the resulting nonlinearity. Here, the nonlinearity of the switching process is given not only by the underlying physical process but also by the nonlinearity of the device I - V characteristic [27]. It depends on a particular VCM/ECM cell whether field or temperature acceleration dominates. Furthermore, the ratio between field or temperature acceleration also depends on the considered voltage regime. For low voltages, the dissipated power might be so low that no Joule heating occurs and thus only field acceleration is achieved. At higher voltages, Joule heating might set in, which additionally accelerates the switching process.

14.6

Modeling of VCM Cells

14.6.1

Ab initio Models and MD Models

The *ab initio* models and MD models introduced for VCM-type switching aim at the understanding of individual stages of the electroforming and RS phenomenon: (i) the possible atomic configurations resembling the LRS and HRS state, (ii) the formation energies required for redox reactions, (iii) nucleation of stable phases, and (iv) the calculation of energy barriers for ion-hopping processes.

14.6.1.1 HRS and LRS State Modeling

It is widely accepted that the SET mechanism in VCM cells involves the evolution of oxygen-deficient regions. Its atomistic configuration, however, is not yet known. Different possibilities have been discussed in the literature. One of the most prominent approaches analyzes chains of oxygen vacancies within an insulating oxide matrix using DFT calculations within the local-density approximation (LDA). This method was employed for vacancy chains in rutile TiO_2 [28], cubic HfO_2 [29], and $\alpha\text{-Al}_2\text{O}_3$ [30]. The calculated partial density of states (PDOS) reveals that a continuous electronic filament results, which can resemble the ON state (Figure 14.2). By analysis of the formation energies for such a cohesive chain, it is revealed that the oxygen vacancies need to be in charge state +1 or 0. In contrast, for a single oxygen vacancy, the +2 charge state is energetically favored. With respect to the PDOS, the HRS state can be achieved by simple displacement of two adjacent oxygen vacancies out of the filament. In a further study, the influence of different dopant materials on the filament characteristics was investigated for TiO_2 [32] and HfO_2 [33]. The calculations revealed that the formation energy of a single oxygen vacancy near the dopant is reduced compared to the bulk. Thus,

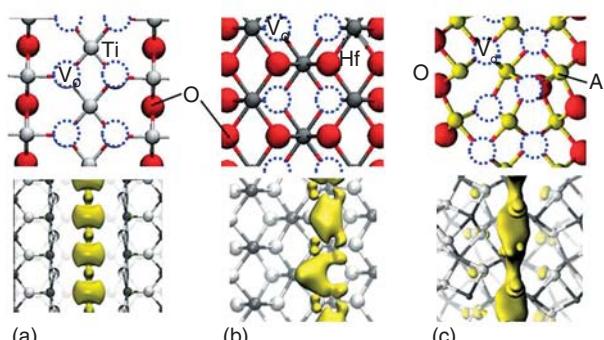


Figure 14.2 Model structures and partial charge density for V_O chain in (a) TiO_2 , (b) HfO_2 , and (c) Al_2O_3 . The formation of conduction path is observed in all three oxides. (Reproduced with permission from [31]. © 2012, IEEE.)

oxygen vacancies tend to gather around a dopant. In addition, the dopant interacts with a nearby oxygen-vacancy filament and can strongly influence its electronic structure and hence its conductivity.

Instead of oxygen-vacancy chains, substoichiometry metal-oxide phases can also resemble the LRS state. Simulations using generalized gradient approximation (GGA) predicted a tetragonal semimetallic Hf_2O_3 structure as the ground state of highly oxygen-deficient hafnia, which undergoes a monoclinic-to-tetragonal phase transition [34]. This tetragonal-phase Hf_2O_3 has been shown to be stable and still exhibits good electric conductivity down to a $1 \times 1 \text{ nm}$ filament cross section in a cubic HfO_2 phase [14]. Also, a zincblende HfO structure can form such a filament [14]. McKenna has shown by *ab initio* calculations that substoichiometric HfO_x in the range of $0.2 < x < 2$ may decompose into $\text{HfO}_{0.2}$ and monoclinic HfO_2 , where the first phase is well-conducting. In TiO_2 , the Magnèli phases $\text{Ti}_y\text{O}_{2y-1}$ with $y=4, 5$ could explain the LRS [6]. It is important to note that an electron-transfer reaction is required in order to describe the resistive-switching process based on the atomistic descriptions of the LRS either as a different phase or as a filament consisting of a chain of neutral oxygen vacancies.

Besides the formation of good-conducting phases within the switching layer, interface-related transport mechanism could also be modulated by changing the number of dopants close to that interface. Bradley *et al.* [35] calculated the metal-oxide band offsets for low, medium, and high oxygen chemical potentials for TiN/HfO_2 and Pt/HfO_2 interfaces (cf. Figure 14.3). In both cases, the conduction band offset increases with increasing oxygen chemical potential, which is equivalent to oxygen interstitials moving toward the interfaces or oxygen

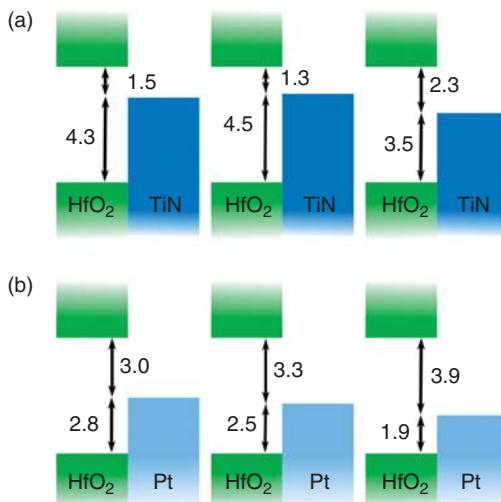


Figure 14.3 Calculated metal– HfO_2 band offsets in electron volts in stable interface structures; left to right: low, intermediate, and high oxygen chemical potential; (a) TiN/HfO_2 and (b) Pt/HfO_2 . (Reproduced with permission from [35]. © 2013, Elsevier.)

vacancies away from the interface. The increase in the band offset is equivalent to a change in the electron transport across this interface. If the offset is larger, the electron injection is decreased, whereas it is increased for a small offset.

14.6.1.2 Electron Transfer

As pointed out, the resistive switching based on the formation of neutral oxygen-vacancy chains or the formation of a substoichiometric phase requires a charge transfer. In a TiO_2 VCM cell, for example, the cohesive energy for an oxygen-vacancy filament is negative if the oxygen vacancies are single positively charged or neutral, whereas it is positive if the oxygen vacancies are doubly positively charged as illustrated in Figure 14.4 [36]. Consequently, the filament is stable for the 0 and +1 charge states and unstable for the +2 charge state. DFT calculations using LDA + U reveal that the HRS state is much more stable by $\sim 4\text{ eV}$ than the LRS state. So, in order to RESET the cell, a positive voltage needs to be applied with respect to the filament. Thus, a neutral vacancy, most likely at the filament top, will be oxidized and will release two electrons to the filament cathode. The V_{o}^{2+} moves

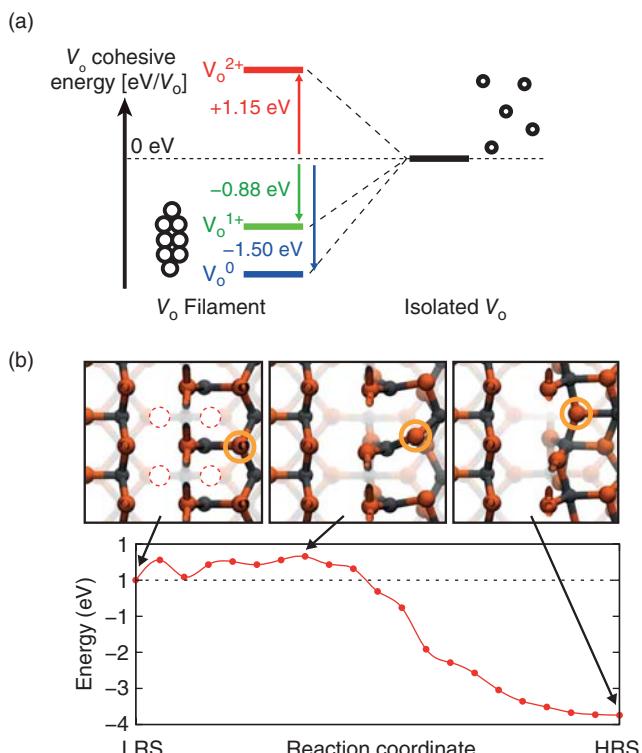


Figure 14.4 (a) Cohesive energies of a V_{o} filament with respect to an isolated V_{o} for different charge states. (b) Energy profile and representative geometries for the

LRS-HRS transition reaction when each V_{o} has a +2 charge state. (Reproduced with permission from [36]. © 2013, IEEE.)

out of the filament. The SET can then be explained by V_0^{2+} moving toward the filament tip and a succeeding reduction at the tip. As an oxidation/reduction reaction occurs at the cathode, a reduction/oxidation reaction will also occur at the anode side. Xue *et al.* [14] proposed that oxygen interstitials are reduced/oxidized at the anode as a counter reaction. However, in this case, Frenkel pairs need to be stable in the bulk of the oxide. O'Hara *et al.* calculated the formation energy for interstitial-oxygen/oxygen-vacancy Frenkel pairs in bulk hafnia [13]. The calculation results revealed that either the interstitial oxygen recombined with the vacancy or the Frenkel pair requires a very high forming energy of $E_{\text{form}} \approx 8.3$ eV. The authors point out that there is no mechanism to provide stability to the Frenkel pair as a complex. For amorphous HfO_2 , it was calculated that the recombination time of a bulk interstitial-oxygen/oxygen-vacancy Frenkel pair is in the subpicosecond range and no activation energy is required for this annihilation process [12]. Rather than forming this pair in the bulk, it is much more energetically favorable to form an “extended” Frenkel defect at the Hf/HfO_2 interface [13]. This means that an oxygen vacancy is created in hafnia at the interface and the released oxygen ion is moved to the Hf metal, probably oxidizing it. To identify the most energetically favorable location of the oxygen interstitials within $\text{Pt}/m\text{-HfO}_2/\text{Pt}$ and $\text{TiN}/m\text{-HfO}_2/\text{Ti}$ samples, Traore *et al.* [37] considered different positions for the removed oxygen and calculated the forming energy of the Frenkel pair using DFT with GGA-PBE (Perdew-Burke-Ernzerhof) functionals. The investigated positions are the HfO_2 layer (as interstitial), the Pt electrode (forming Pt-O bonds), corundum Ti_2O_3 (a possible oxide at the HfO_2/Ti interface), and the Ti electrode (forming Ti-O bonds). Based on the calculated data, the authors conclude that Frenkel pairs are not formed within the bulk in the $\text{Pt}/\text{HfO}_2/\text{Pt}$ stack, whereas the Ti electrode is oxidized in the $\text{TiN}/m\text{-HfO}_2/\text{Ti}$ sample. Further DFT calculations showed that the incorporation of oxygen into TiN or Pt electrodes is energetically unfavorable [35]. Both metals, however, do not show a perfect structure, so oxygen could be incorporated at grain boundaries (GBs) or released at the edge of the electrode in contact with air.

14.6.1.3 Phase Transformations and Nucleation

If a new phase grows during switching or electroforming, a stable nucleus needs to be formed first. McKenna calculated the thermodynamic stability of substochiometric HfO_x for varying oxygen content [38]. They found that it is thermodynamically favorable that HfO_x segregates into a mixture of pure monoclinic HfO_2 and hcp- $\text{HfO}_{0.2}$ in the range of $0.2 < x < 2$. To obtain the nucleation barrier, the formation energy of a metal-rich spherical cluster within a $m\text{-HfO}_2$ phase as illustrated in Figure 14.5a was calculated for different radii r . Figure 14.5b illustrates the resulting change in free energy ΔG for different oxygen content x . For very low oxygen content, the nucleation barriers are too high as the phase would occur. On the other hand, the barrier is the lowest for an almost stoichiometric hafnia film. However, as very few oxygen vacancies are present, the nucleation is probably kinetically hindered. According to the authors, the optimal oxygen content is between $x = 1.50$ and 1.75 , for which the activation barriers are of the order

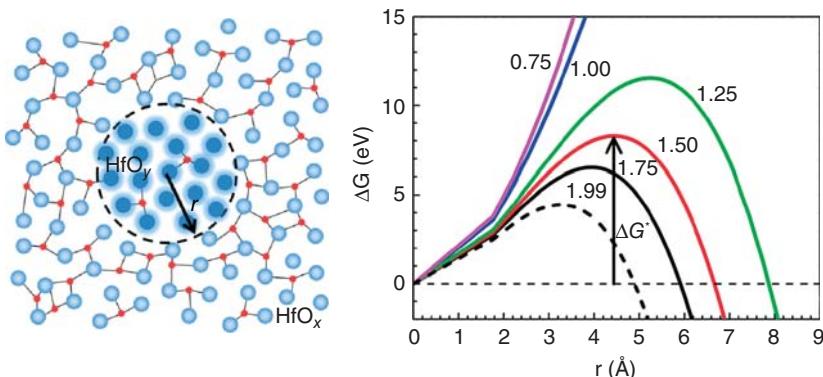


Figure 14.5 (a) The stability of a spherical metallic precipitate embedded in HfO_x relative to a homogeneous HfO_x phase depends on its radius, r_{Hf} . Further growth of the precipitate by outward diffusion of oxygen becomes thermodynamically favorable only beyond a critical radius r . (b) The

change in free energy ΔG on precipitation of the metal-rich cluster of radius r in HfO_x for different composition x . (Reproduced with permission from [38]. © 2014, IOP Publishing Ltd.)

6–8 eV. This still seems to be a very high barrier, but it might be lower at the electrode/hafnia interfaces. Nucleation at an electrode also appears to be favorable if one considers that a charge transfer has to occur.

14.6.1.4 Calculation of Migration Barriers

A very important process during VCM switching is the movement of oxygen within the oxide film. As the formation of oxygen interstitial/oxygen-vacancy pairs is associated with a very high formation energy and these Frenkel pairs are thermodynamically not stable [12, 13], it is most likely that the oxygen transport occurs via the vacancy mechanism. Nevertheless, the theoretical migration barrier for an interstitial oxygen might be lower as for an oxygen vacancy. In *m*-HfO₂, for example, the migration barriers for doubly charged oxygen ions have been calculated to be 0.6, 1.8, and 0.7 eV for the interstitialcy [39], interstitial [39], and vacancy transport mechanism [40]. Also, for amorphous HfO_x, the oxygen ion migration has been described as a vacancy mechanism and the calculated migration barriers were reported to be 0.57–0.66 eV for different substoichiometry [41].

In crystalline materials, GBs or dislocations are thought to be the locations of filamentary switching as oxygen vacancies tend to segregate at the GBs/dislocations. Thus, it is important to study the migration barriers along such defects. In reference [9], the migration path along a (101) twin GB in *m*-HfO₂ was calculated to be 0.65–0.7 eV for a doubly positively charged oxygen vacancy. The calculated migration barriers are close to the reported one for bulk transport. A different result is obtained for the SrO and TiO₂ dislocation cores at a 6° [001] tilt boundary in crystalline SrTiO₃. Here, the oxygen-vacancy migration barrier along the TiO₂

core was calculated to be 3.43 eV, which is much higher than the barrier in the bulk with 0.63 eV or into the dislocation core (0.75 eV). The migration barrier along the SrO core could not be calculated, but the migration barrier into the SrO core is already 1.37 eV. The authors conclude from the simulation study that, if the activation energy of ion migration in the regular lattice is facile, strong perturbation of the lattice, as at a dislocation core or GB, will increase the migration barrier. In contrast, a decrease in the migration barrier is expected for a more open structure if the migration barrier is initially high [42].

Schie *et al.* [43] studied the influence of dopants on the oxygen vacancy migration in SrTiO_3 . Based on empirical potential calculations, formation energies as well as migration energies for an oxygen vacancy in the vicinity of an acceptor dopant were calculated, and it was found that the migration barriers are modulated for positions up to fourth nearest neighbors. The effect of trapping oxygen vacancies near acceptor dopants was shown to be dependent on the type of dopants, its concentration, and the temperature.

14.6.2

Kinetic Monte Carlo Modeling

KMC methods have been employed to simulate the electroforming and resistive switching in HfO_2 thin films in 3D [44, 45] and 2D [46, 47]. Until now, however, the capability to simulate the VCM switching kinetics has not been demonstrated. All of these models rely on an oxygen-ion generation/recombination mechanism first proposed by Gao *et al.* [11]. In this model, the oxygen-vacancy-rich region defines the conduction mechanism and the oxygen vacancies are assumed to be immobile. The oxygen-vacancy-rich region is formed (forming/SET) by creation of oxygen-interstitial/oxygen-vacancy Frenkel pairs within the bulk followed by the outward diffusion of oxygen interstitials. The RESET is modeled as a recombination of these Frenkel pairs. With respect to the simulation results obtained by *ab initio* calculations that have been presented in the previous section, some discrepancies can be found. First of all, oxygen vacancies should be able to move under the applied fields and local temperature during forming/switching. Secondly, the *ab initio* simulations revealed that the formation energy for Frenkel pairs within the bulk is very high, and they are not stable [12, 13]. Thus, the simulation results obtained with these KMC models should be considered with care.

Based on the described microscopic form, Vandelli *et al.* simulated the electroforming process in HfO_2 VCM cells [44]. The electron transport across the preformed polycrystalline HfO_2 layer is assumed to occur preferentially along GB by multiphonon trap-assisted tunneling (TAT) via oxygen-vacancy defects. For the simulation, several GBs are assumed, and all of them are modeled as filaments with 1 nm diameter. Due to the TAT electron capture/emission process, the power P is released locally into the lattice. This power dissipation is used to calculate the 3D temperature profile T by solving the Fourier's heat flow equation

$$P(x, y, z) = k_{\text{th}} \nabla^2 T(x, y, z), \quad (14.9)$$

where k_{th} is the thermal conductivity of the hafnia film. The generation rate of new oxygen-vacancy defects, which is associated with the creation of oxygen vacancy/interstitial Frenkel pairs by Hf–O bond breaking within the bulk film, is calculated according to Eq. (14.6) to

$$G_F(x, y, z) = G_0 \exp \left(-\frac{\Delta W_{\text{bb}} - bE(x, y, z)}{k_B T(x, y, z)} \right). \quad (14.10)$$

The electrical field E is obtained by solving the Poisson equation accounting for the charge and electron occupation of the defects. Further movement of oxygen ions and/or vacancies is not considered in this model. In addition, recombination of the Frenkel pair is not considered, which should occur without activation energy within the timeframe of less than picoseconds [12]. This model has been used to simulate the electroforming process in TiN/HfO₂/TiN structure under constant voltage stress of 3.3 V. The time evolution of the transient current in one GB is depicted in Figure 14.6a. The overall current is carried by the current flowing through several GBs, which all degrade [44]. But, just for one GB, an abrupt current increase is observed (red dashed line) corresponding to the formation of a highly conductive filament. The progressive degradation of the leakage path has been analyzed in detail. Three different stages have been identified. Figure 14.6b illustrates the defect, power dissipation, and temperature distribution for the points in time A–D. In the initial stage (A,B), defects are generated only by the applied electric field as the current is too low to result in a temperature increase. In stage II (C), Joule heating becomes more prominent, accelerating the process of defect generation, and finally, a thermal runaway occurs (D). Several simulations were performed in order to study the influence of the ambient temperature on the time to breakdown (BD). As illustrated in Figure 14.6c, the simulation results are in good agreement with the experimental data with respect to the temperature dependence as well as the statistics of time to BD.

The described model was further modified by Butcher *et al.* [45, 48] in order to study the final processes during electroforming and the subsequent resistive switching. To model the electric current, the simulation domain is divided into small bins, each with a specific electrical conductivity that depends on the concentration of oxygen vacancies. The potential distribution is calculated using the current continuity equation, and Fourier's heat flow equation is solved to obtain the temperature distribution. As in the previous model, the generation rate for oxygen vacancy/interstitial Frenkel pairs is calculated according to

$$G_F(x, y, z) = G_0 \exp \left(-\frac{\Delta W_{\text{bb}} - bE_{\text{ext}}}{k_B T} \right). \quad (14.11)$$

The activation energy ΔW_{bb} is assumed to be inhomogeneous: it is 2.8 eV in the GB region and 6.2 eV in the bulk of the oxide film. In contrast to the model designed by Vandelli *et al.*, here also the diffusion of oxygen interstitials with rate

$$D = D_0 \exp \left(-\frac{\Delta W_{\text{hop}} - e^{\frac{a}{2}} [E_{\text{ext}} + E_{\text{CR}}]}{k_B T} \right) \quad (14.12)$$

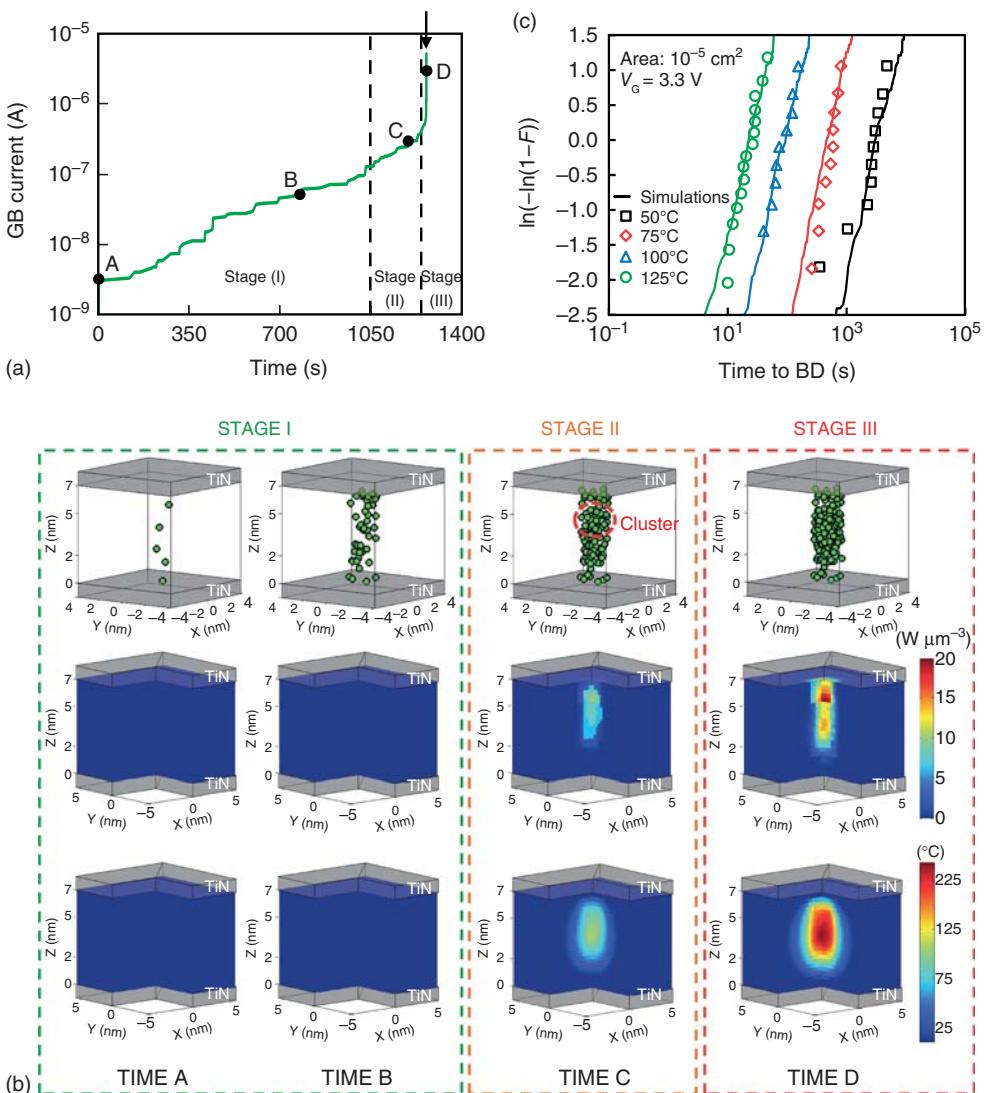


Figure 14.6 (a) Evolution of the current flowing across the GB determining the BD as simulated for a 7-nm-thick TiN/HfO₂/TiN stack during constant voltage stress at $V_G = 3.3$ V. (b) Three-dimensional maps of (first row) defects, (second row) power dissipation, and (third row) temperature along the GB that is converted into the BD spot as extracted at different times (A, B, C, and

D) of the simulated current I_{GB} transient in (a). The green spheres represent the defects created by stress. Stages I–III refer to the three subsequent degradation stages as in (a). (c) Experimental (symbols) and simulated (lines) time to BD distributions on a 7-nm TiN/HfO₂/TiN subjected to a 3.3 V CVS at different temperatures. (Reproduced with permission from [44]. © 2013, IEEE.)

and the recombination of the Frenkel pairs with rate

$$R = R_0 \exp\left(-\frac{\Delta W_R}{k_B T}\right) \quad (14.13)$$

are considered. Still, the oxygen vacancies are assumed to be immobile. In Eq. (14.12), E_{CR} denotes the local field that is determined by the oxygen ions and vacancy charge states are included. The activation energies for diffusion and recombination are 0.7 and 0.5 eV, respectively. While the recombination is considered in this model, the activation energy is chosen to be very high as compared to the fact that this process should occur kinetically unhindered with a speed faster than picoseconds (cf. *ab initio* section). The simulation model has been employed to simulate the resistance distribution obtained for different electroforming conditions, that is, different ambient temperature and current compliance. The simulated resistance distributions reveal that the higher the ambient temperature/current compliance, the lower is the final resistance state (cf. Figure 14.7a). In addition, the simulation data agrees very well with the experimental data. The simulated electroformed filament structures were used as input for further simulations of the resistive switching, that is, first RESET and subsequent SET operation. During RESET, a gap evolves between the bottom electrode and the filament. A complete I - V characteristic for a pulsed RESET/SET operation is shown in Figure 14.7b. Thus, the model is able to simulate the electroforming process as well as the SET and RESET operation. The simulated I - V characteristic, however, does not match the experimental behavior very well. The current decrease due to Joule heating in the LRS is not observed in experiment. In addition, SET and RESET seem to occur more abruptly in experiment than in simulation.

14.6.3

Continuum Modeling

By using the continuum models, several groups have investigated the dynamics of the VCM switching. In all of these models, the oxygen-vacancy transport has been regarded as the physical process that determines the switching kinetics.

Noman *et al.* [49] used a drift–diffusion model to simulate the operating window for resistive devices based on homogenous ion motion similarly to the approach of Strukov *et al.* [50]. Particularly, it was investigated if such a device can fulfill both: 10-year retention and switching faster than 100 ns. The switching model is based on the modulation of a Schottky barrier at one metal/oxide interface due to the movement of charged oxygen vacancies that act as donors. As the concentration of donors is increased/decreased at this interface, the barrier height decreases/increases due to the Schottky effect and the current transport across the interface is increased/decreased (SET/RESET operation). The second interface is assumed to be ohmic. To model the current transport across the metal-insulator-metal structure, the Poisson equation

$$\nabla \cdot (-\epsilon_r \epsilon_0 \nabla \varphi) = e(p - n + N_D) \quad (14.14)$$

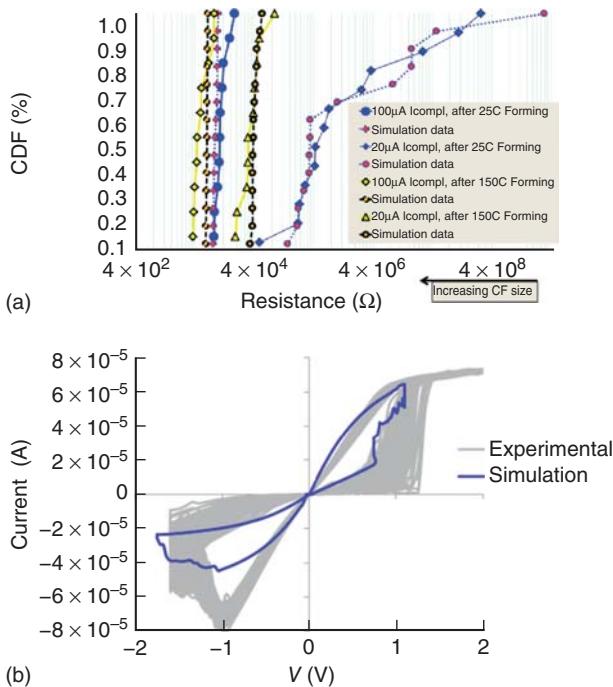


Figure 14.7 (a) Comparison between experimental and simulated resistances. Constant voltage forming was performed and simulated at two different current compliance levels as well as two different ambient temperatures as specified in the legend. The read voltage is 0.1 V.

(b) Simulated pulse SET/RESET cycling (blue color) in TiN/HfO₂/TiN cell. Pulse width = 100 ns, rise/fall = 10 ns, $V = -1.75$ for RESET, $V = 1.1$ V for SET. Experimental data are shown in the background for comparison (gray color). (Reproduced with permission from [45]. © 2013, IEEE.)

is solved for the potential φ coupled to the continuity equations for electrons n

$$\frac{\partial n}{\partial t} = \frac{1}{e} \nabla \cdot (e\mu_n n E + eD_n \nabla n) \quad (14.15)$$

and holes p

$$\frac{\partial p}{\partial t} = \frac{1}{e} \nabla \cdot (e\mu_p p E - eD_p \nabla p), \quad (14.16)$$

where ϵ_r denotes the relative permittivity of the oxide, ϵ_0 is the free-space permittivity, and μ_n/μ_p and D_n/D_p are the electron/hole mobility and diffusion constant, respectively. The donor concentration N_D resembles the oxygen-vacancy concentration with charge state +1. For the injection of electrons into the insulator, thermionic emissions over the barrier and a parallel channel for electron tunneling are considered in order to calculate the current flow through the device. The time dependence of the oxygen-vacancy motion is modeled using

$$\frac{\partial N_D}{\partial t} = -\frac{1}{e} \nabla \cdot J_{N_D} \quad (14.17)$$

with the vacancy flux J_{N_D} being calculated according to

$$J_{N_D} = \frac{2}{a} D_0 N_D \sinh\left(\frac{eaE}{2k_B T}\right) \left[1 - \frac{N_D}{N_{D,\max}}\right] - D_0 \frac{dN_D}{dx} \cosh\left(\frac{eaE}{2k_B T}\right). \quad (14.18)$$

In Eq. (14.18), the first term describes the ion drift and the second one the ion diffusion. Hereby, the nonlinear drift dependence on high electric field is taken into account. In addition, a maximum concentration $N_{D,\max}$ is introduced, which is related to the maximum oxygen site concentration in the crystal. The diffusion constant D_0 is given by

$$D_0 = \frac{1}{2} a^2 f_0 \exp\left(-\frac{\Delta W_{\text{hop}}}{k_B T}\right) \quad (14.19)$$

and depends on the attempt frequency f_0 . Joule heating was not considered in the simulation model. By using this equation system, first, an equilibrium-vacancy distribution under zero bias was calculated (cf. solid line in Figure 14.8). By applying ± 5 V 100 ns programming pulses with respect to the Schottky interface M1, the ON and OFF-state distributions are obtained. The corresponding potential profiles shown as inset in Figure 14.8a reveal that the shape of the Schottky barrier at electrode M1 is strongly modulated and thus the device resistance is changed. Starting with the ON and OFF-state distributions, the equilibration of the vacancy distribution is simulated for different oxygen-vacancy hopping barriers under zero bias. Both resistance states relax to the equilibrium state over time as illustrated in Figure 14.8b. The relaxation time, however, depends on the activation energy for the hopping transport. According to the simulation results, a 10-year retention can be achieved within this model if the activation energy is higher than 1.02 eV. In a further analysis, Norman *et al.* calculated the oxygen-vacancy velocity for different ambient temperatures and activation energies (cf. Figure 14.8c). Based on the calculation results, the authors conclude that there are no conditions where 10-year retention and 100 ns writing time can be fulfilled at the same time. This result is actually restricted by the assumptions that temperatures higher than 1250 K or electric fields higher than 1 MV cm^{-1} are not feasible without destroying the device and only the oxygen-vacancy motion limits the switching speed. In fact, at low voltage, other processes might determine the switching kinetics, thus enabling to achieve more orders of nonlinearity. In a similar analysis of the drift velocity by Strukov and Williams [51], the authors concluded that it is possible to have year-scale retention and 10 ns switching for activation energies larger than 1.2 eV. In this study, the authors used the local Lorentz field to calculate the field enhancement of the ionic drift. This proposition, however, is incorrect as the electric field experienced by the moving ion spatially fluctuates and is position-dependent as pointed out by Meuffels and Schroeder [52]. Consequently, only the mean field can be used to calculate the ion enhancement.

In the previously described models, Joule heating was just considered indirectly by assuming a different temperature. With this approach, it cannot be identified whether electric field or temperature is the origin of the nonlinearity in a VCM cell. To answer this question, Menzel *et al.* used an electrothermal FEM simulation

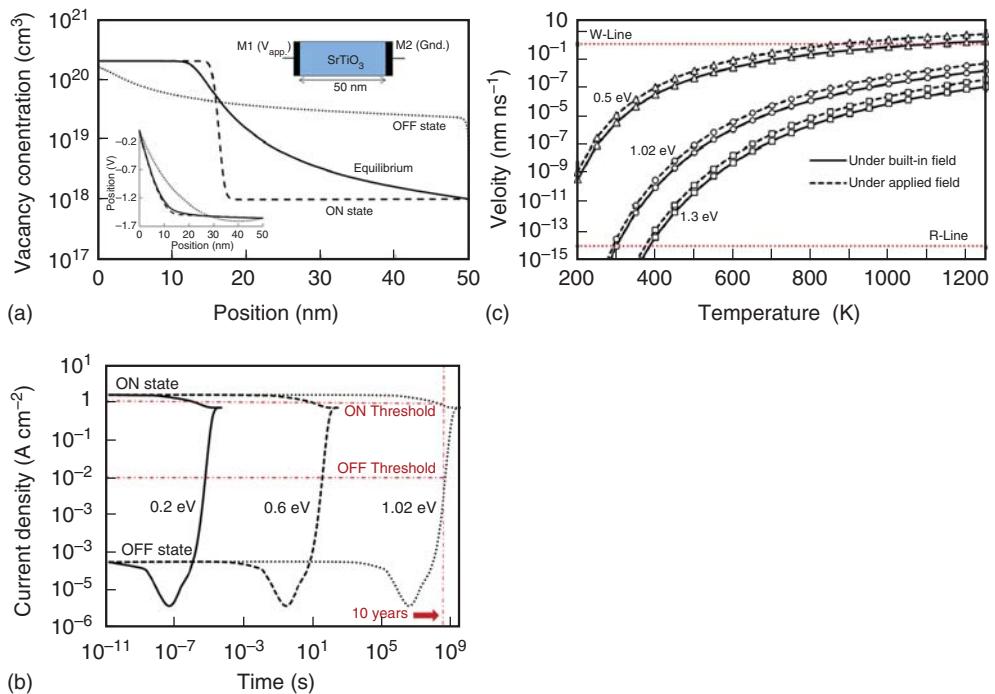


Figure 14.8 (a) Simulated plot of vacancy profile after application of $\pm 5\text{ V}$ 100 ns programming pulses. The “OFF” (“ON”) state corresponds to a situation when tunneling current is reduced (increased). Lower inset shows corresponding potential profiles, and upper inset shows the device structure. (b) Simulated current density measured at -1 V as a function of time for three different activation energies for vacancy migration. Current is measured periodically as the vacancy

profile evolves from the programmed to equilibrium state. To achieve 10 years of retention, an activation energy of 1.02 eV or larger is necessary. (c) Vacancy velocity as a function of temperature under macroscopically observable electric field. Built-in field is 0.31 MV cm^{-1} ($1.55\text{ V}/50\text{ nm}$), and applied field is 1 MV cm^{-1} . The applied field is generated by applying 5 V bias over the 50 nm thick SrTiO_3 layer. (Reproduced with permission from [49]. © 2011, Springer-Verlag.)

model to explain the switching kinetics of a strontium-titanate-based VCM cell [53]. As a limiting process, only the oxygen-vacancy drift has been considered. It was assumed that the oxygen vacancies need to move a distance L in order to trigger the SET process. By using this criterion and Eq. (14.2), the switching time t_{SET} can be calculated as

$$t_{\text{SET}} = \frac{L}{v_0 \sinh\left(\frac{zea}{2k_B T} E\right)} \exp\left(\frac{E_A}{k_B T}\right). \quad (14.20)$$

The electric field E and the local temperature T were obtained by solving the conductivity equation and the heat-transfer equation for the investigated device structure (cf. Figure 14.9a). In this electrothermal model, the nonlinearity of the

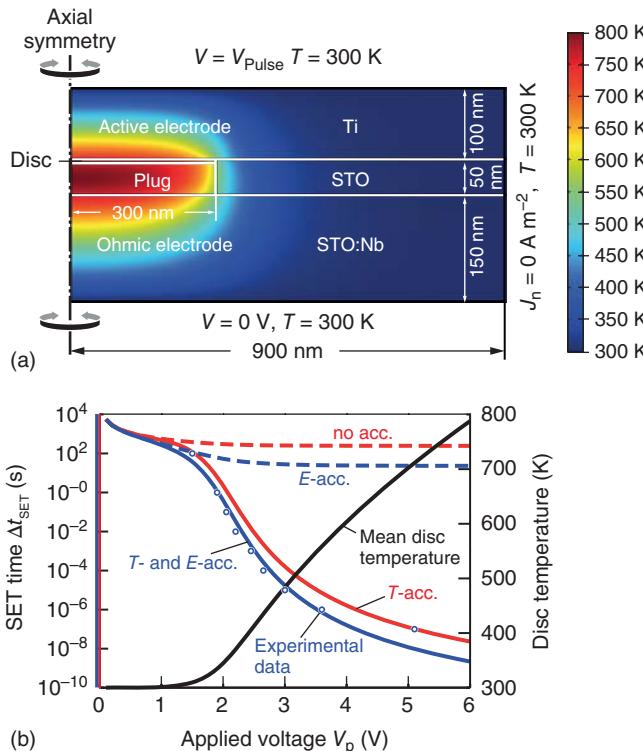


Figure 14.9 (a) Cross section of the simulated cell. The temperature distribution for $V_p = 5 \text{ V}$ is shown in color. (b) Calculated SET time depending on the applied voltage for the four different cases in red and blue compared to the experimental data (blue circles). The black line corresponds

to the mean disc temperature used for the calculations. The data point at 100 ns must be regarded with caution because it may be affected by the RC time of the setup. (Reproduced with permission from [53]. © 2011, Wiley-VCH.)

HRS I - V characteristic was modeled by a filament consisting of a small disc close to the active electrode, which resembles the nonlinear barrier transport, and a well-conducting plug region. By using this simulation approach and by comparison to the experiment, the authors could identify that the origin of the nonlinear switching kinetics is the temperature increase induced by Joule heating rather than the nonlinear drift at high electric fields. While the local Joule heating accelerates the switching speed by approximately 9 orders of magnitude, the electric field can only account for 2 orders of magnitude as illustrated in Figure 14.9b.

The investigated device in that study was rather big; thus, further simulations were performed in order to investigate the device scaling behavior [5]. It was shown that also for very small devices, fast switching speed is expected. Moreover, the switching energy decreases down to less than picojoules, while the currents can be maintained at $<100 \mu\text{A}$. As Joule heating plays an important role, the device design should also be optimized with respect to the thermal properties.

The first fully dynamic axisymmetric 2D model that is capable of simulating the I - V characteristics and the switching kinetics of a VCM cell was developed by Larantis *et al.* [54]. In this model, a TiN/HfO_x/TiN VCM cell was considered. Here, the local conductivity is a function of the dopant concentration N_D , which can be oxygen vacancies. Barrier transport or space charges in the bulk were not considered in the model. The dopant movement is modeled using the drift–diffusion equation

$$\frac{\partial N_D}{\partial t} = \nabla(D\nabla N_D - \mu EN_D). \quad (14.21)$$

The electric field and the local temperature are calculated solving the continuity equation

$$\nabla\sigma\nabla\varphi = 0 \quad (14.22)$$

coupled with Fourier's heating law

$$-\nabla k_{\text{th}}\nabla T = \sigma|\nabla\varphi|^2. \quad (14.23)$$

The initial dopant concentration was assumed to be homogenous in a filamentary region, thus describing the LRS. By applying a positive voltage, the positive dopants at the hottest spot within the filament start to migrate within the electric field. This leads to an opening of the filament close to the anode as illustrated in Figure 14.10a. By reversing the polarity, the gap is closed again. The corresponding I - V characteristic is shown in Figure 14.10b and c for different stop voltages during the RESET process compared to the experimental data of the TiN/HfO_x/TiN device. The simulation model shows a very good agreement with the experimental data. The gradual decrease during the RESET process can be explained by the gradual gap opening. For a certain voltage V_{stop} , the local electric field and temperature largely decrease such that no further gap opening occurs. In order to open the gap even more, a higher voltage needs to be applied. The stop voltage also influences the following SET voltage. If the gap increases, the local electric field and temperature driving the SET transition are decreased. Thus, a higher voltage is required.

The simulation model was used to simulate the RESET switching dynamics of the TiN/HfO_x/TiN VCM cells. Figure 14.11 shows the simulated resistance evolution during RESET for different pulse voltages compared to the experimental data. For voltages higher than 0.7 V, a very good agreement between simulation and experiment has been achieved. Here, the switching speed varies over 3 orders of magnitude upon a voltage change of 0.5 V. The simulation data was further analyzed by plotting the RESET switching time versus the inverse maximum temperature along the filament. From the resulting Arrhenius plot, activation energy of 1 eV is extracted, which is exactly the used activation energy for the ion migration transport. Thus, it can be concluded that the local temperature determines the RESET switching speed.

This model has also been used later by the authors to simulate complementary switching, that is, complementary resistive switching in a single cell [55]. Lu and

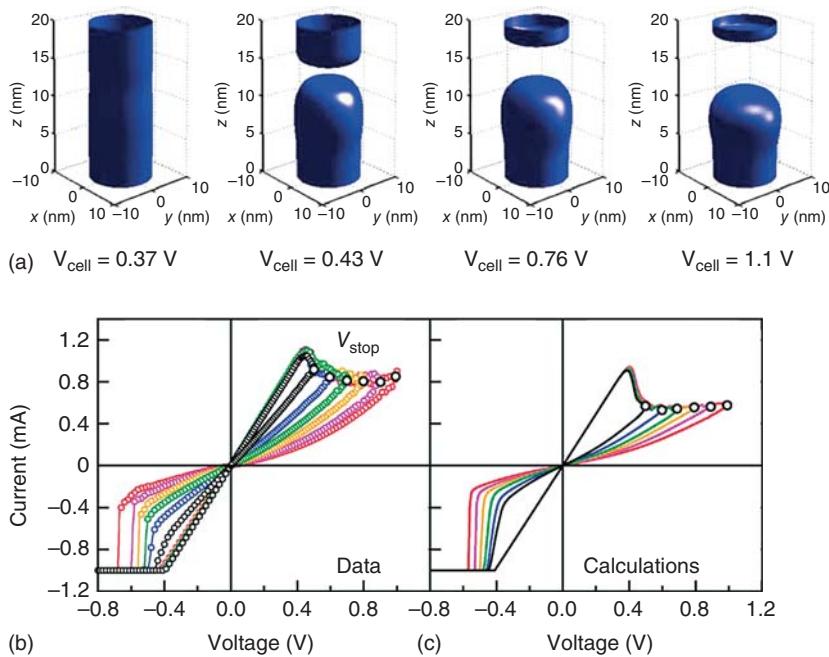


Figure 14.10 (a) Three-dimensional contour plot of doping density, corresponding to points having $N_D = 6 \cdot 10^{20} \text{ cm}^{-3}$, illustrating the evolution of the conductive filament shape during the gradual RESET transition, providing an insight into the length of the depleted gap Δ . (b) Measured and (c) calculated I - V curves, showing RESET ($V > 0$) and SET ($V < 0$) transitions, obtained by applying

triangular voltage sweeps. First, the RESET sweep, applied to an initial ON state with $R = 400 \Omega$, is interrupted at V_{stop} for preparing a HRS state with variable R . Then, the SET sweep is applied, showing that V_{SET} increases with V_{stop} , hence with R . (Reproduced with permission from [54], © 2012, IEEE.)

coworkers adapted this model to simulate the bipolar and complementary switching in a TaO_x -based VCM cell [56]. In this model, Joule heating within the electrodes, nonlinear dopant drift at high electric fields, and thermodiffusion are also included. In both models, the initial model designed by Nardi *et al.* and the model adapted by Lu and coworkers, the electron transport is modeled by combining Ohm's law with the current continuity equation, and the resistance change is described by a pure change in the bulk electronic conductivity. The effect of contact potentials and electron transport through the metal/oxide interfaces (Schottky barrier transport) is neglected, although it is widely assumed to play an important role in the switching process. In 2015, Marchewka and coworkers implemented a physics-based VCM model that comprises drift, diffusion, and thermodiffusion of electronic and ionic charge carriers and includes the effect of a contact potential and related barrier formation as well as Joule heating effects [57, 58]. This model could accurately reproduce the gradual RESET transition during pulse operation. Based on the simulation results, the gradual nature of

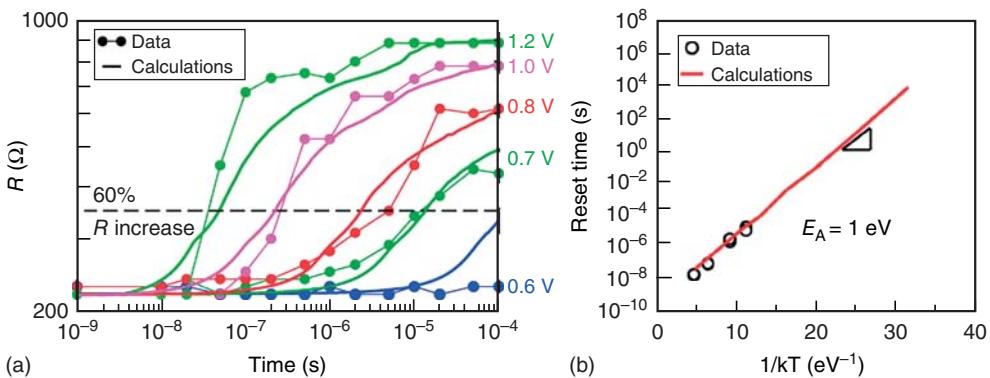


Figure 14.11 (a) Measured and calculated R as a function of total biasing time during a sequence of rectangular RESET pulses for increasing pulse amplitude V_{app} from 0.6 to 1.2 V (solid lines). The dashed line illustrates the threshold resistance, corresponding to a 60% increase with respect to the initial value around 220Ω to define RESET time t_{RESET} . (b) Measured and calculated t_{RESET}

as a function of $1/k_B T$. The RESET time was evaluated from the RESET transition in (a). T in the Arrhenius plot was evaluated from the simulation at the maximum temperature along the conductive filament at V_{RESET} . The data is extrapolated to lower T . (Reproduced with permission from [54]. © 2012, IEEE.)

the RESET could be linked to the temperature-activated oxygen-vacancy motion with the drift and diffusion processes approaching an equilibrium condition, combined with a moderate sensitivity of the current response to the induced contact barrier changes [58]. Furthermore, the model was applied to explain the transition from complementary switching to pure bipolar switching in a Ta/TaO_x/Pt stack by changing the Ta layer thickness [59]. Based on this model, the transition occurs when the asymmetry between the two metal-insulator electron barriers is increased [57].

14.6.4 Compact Modeling

The first approach to describe resistive switching was by Strukov *et al.* in 2008 [60]. The presented “memristor” model describes the change in resistance as a linear function of the total device current – not the ionic current – as described in Chapter 2. This model predicts an inverse proportionality of the switching time to the applied voltage, that is, $t_{sw} \propto 1/V_{app}$ [61]. This contradicts the experimental data that shows an exponential relationship. In addition, the linear memristor model also fails to model the behavior of two antiparallelly connected bipolar ReRAMs, that is, a complementary resistive switch (CRS) [61]. Later, the same group introduced a nonlinear driving force for the resistance change, which was fitted to the experimental data [62]. Here, the resistance change was attributed to a variation of a tunneling gap. With this model, the dynamic nonlinear switching behavior of a Pt/TiO_x/Pt device could be reproduced. Later, it was demonstrated

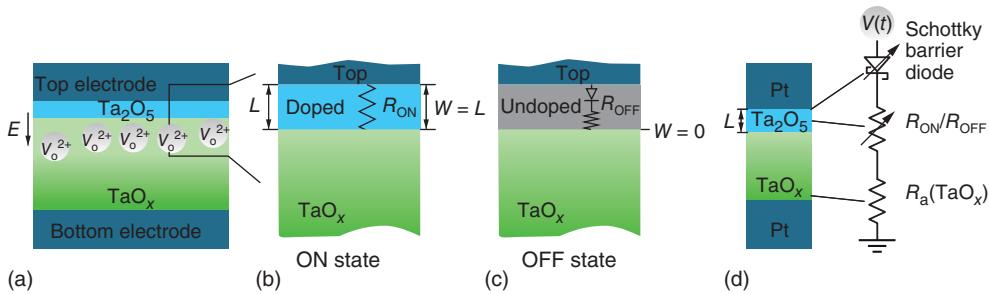


Figure 14.12 Conceptual schematic of bipolar resistive switching. (a) Overall schematic. (b) On-state assumption, $W=L$ doped region in ohmic state. (c) Off-state assumption, no doped layer $W=0$ with Schottky

barrier in the top interface. (d) Conceptual circuit model with variable Schottky barrier, variable resistor, and base-layer resistance. (Reproduced with permission from [63]. © 2010, The American Physical Society.)

that CRS behavior can be reproduced by this model as well [61]. However, the mathematical description of the driving force was fitted to experiment and contains no information about the physical mechanisms involved. The I - V relation is symmetric with respect to positive or negative voltages, as the resistance is described by a symmetric tunneling gap. Thus, it fails to describe ReRAMs with an asymmetric I - V relation.

Hur *et al.* addressed this asymmetry by self-consistently solving the coupled equations of oxygen migration, formation of Schottky barrier between the electrode and the oxide layer, and electronic conduction equation for both metallic conduction (LRS) and thermionic emission over the Schottky barrier at HRS [63]. In this paper, they suggested that the asymmetry of I - V curve at HRS is caused by the Schottky barrier, which shows diode-like rectifying characteristic, formed at the interface between the electrode and the semiconducting filament. The conceptual illustration of this model is shown in Figure 14.12.

The length of the oxide layer and thus its resistance $R(t)$ are modulated according to the applied current. Then, the current in the HRS can be expressed from the Schottky barrier conduction relation as follows:

$$I(t) = \begin{cases} I_0 \left[\exp \left\{ \frac{e}{\eta k_B T} [V(t) - I(t)R(t)] \right\} - 1 \right] & \text{for } V(t) > 0 \\ I_0 & \text{for } V(t) < 0 \end{cases} \quad (14.24)$$

Here, the current prefactor is given by

$$I_0 = AA^*T^2 \exp(-e\phi_B/k_B T), \quad (14.25)$$

where A is the electrode area; A^* is the Richardson constant; T is the temperature, which is 300 K in all cases; ϕ_B is the effective Schottky barrier height; and V is the applied bias. The ideality factor η in Eq. (14.24) is given by

$$\eta = 1 + \left(\frac{\delta}{\epsilon_i} \right) \frac{(\epsilon_s/w_D) + eD_{its}}{1 + (\delta/\epsilon_i)eD_{itm}}, \quad (14.26)$$

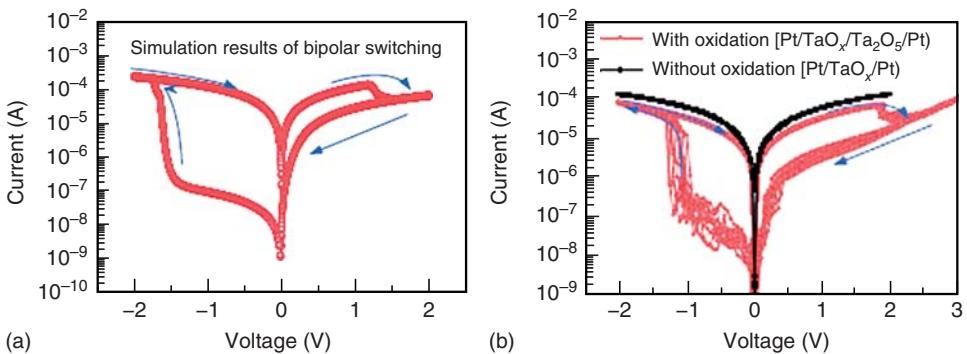


Figure 14.13 (a) Results of bipolar resistive switching simulation. (b) Typical experimental results of double layered bipolar ReRAMs showing asymmetry at HRS. This

results from the Pt/TaO_x/Pt (black) and Pt/Ta₂O₅/TaO_x/Pt (red) films. (Reproduced with permission from [63]. © 2010, The American Physical Society.)

where δ is the metal–semiconductor interfacial layer thickness, ϵ_i is the dielectric constant of the insulator, ϵ_s is the dielectric constant of the semiconductor, w_D is the depletion layer thickness, D_{its} and D_{itm} are the interfacial trap densities in the semiconductor and insulator, respectively. The Schottky barrier height is modified by the applied voltage and the change in the oxide layer thickness by

$$\phi_B \approx \phi_{B0} - \left(\frac{e^3 n_s \phi_{B0}}{8\pi^2 \epsilon_s^3} \right)^{1/4} \left(1 - \frac{V}{4\phi_{B0}} \right), \quad (14.27)$$

where n_s is the base-layer charge density and ϵ_s is the dielectric constant of the thin oxide layer. With this effective Schottky barrier and equivalent circuit model (Figure 14.12d), the resulting HRS currents can be obtained by substituting Eq. (14.27) to Eqs. (14.24) and (14.25). The simulated I - V curve exhibits good qualitative and quantitative agreement with the experiments (Figure 14.13).

The previous model was extended by the authors to describe multilevel switching [64]. In this model, the ion migration process is described using more realistic equations of motion and related Schottky barrier formation. The multiplicity of resistance levels comes from the “level” of oxidation of the filament, which can hardly be fully oxidized under normal condition. For the preforming state, the resistance is much higher than that of the HRS. This fact tells that even in HRS, the oxidation level of the filament is far below its initial state. This gives the possibility for additional oxidation accompanied by a resistance increase. The level of oxidation in Ref. [64] is parameterized with the total length of the filament $L(t)$ correlated with the high-resistance volume length $l(t)$. This multilevel phenomenon is observed in almost all the oxide-based ReRAMs with different detailed characteristics as discussed in Chapter 11. The ionic equation of motion differs from the previously published compact models as it takes into account the dependence of

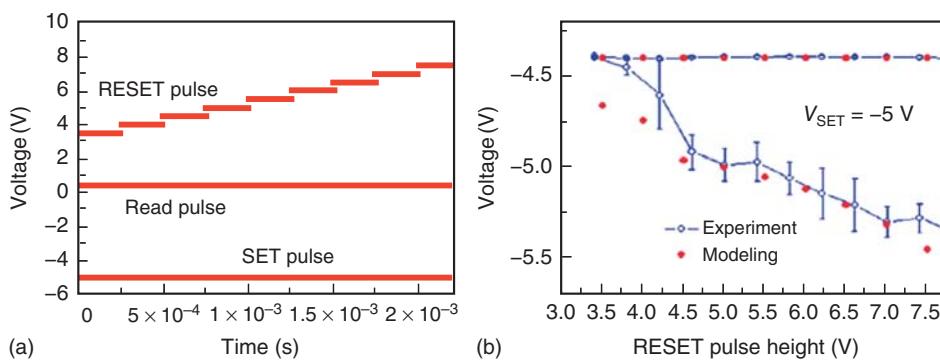


Figure 14.14 Experimental and modeling results of pulse operation of tantalum oxide ReRAM. (a) Pulse signals with respect to time. RESET pulse amplitudes are increased from 3.5 to 7.5 V for every $250 \mu\text{s}$, SET voltage is -5 V . For SET/RESET pulses, 500 ns

duration is used and rise/fall time of the pulses is set to 0. (b) Read operations are performed right after SET and RESET pulses are applied. (Reproduced with permission from [64] © 2012, IOP Publishing Ltd.)

the ionic mobility on the oxygen-vacancy density as

$$\mu_O = [\mu_H (1 - \zeta(t)) + \mu_L] \cdot \exp \left[-\frac{\left(\Delta W_{\text{hop}} - \frac{1}{2} zeaE \right)}{k_B T} \right], \quad (14.28)$$

where $\zeta(t) = l(t)/L(t)$ denotes the filament fraction that is oxidized, $\mu_H (1 - \zeta(t))$ is the oxygen ion mobility in the insulating volume depending on the insulator volume, μ_L is the oxygen ion mobility in the HRS, ΔW_{hop} is the ion migration barrier, and $z = -2$. By considering the degree of oxidation, the Schottky barrier height relation becomes more complex and can be represented as follows:

$$\phi_B(t) \approx \phi_{B0}(t) - \sqrt{\frac{eE_m}{4\pi\epsilon}} = \phi_{Bi} \sqrt{\frac{L(t)}{L_0} \zeta(t)} - \sqrt{\frac{eE_m}{4\pi\epsilon}}. \quad (14.29)$$

By coupling Eqs. (14.28) and (14.29), the resulting resistance levels are obtained in terms of the presented parameters according to

$$\begin{aligned} R(t) &= \frac{\eta k_B T}{eI(t)} \log \left(\frac{I(t)}{I_0(t)} + 1 \right) + R_{\text{ins}} \zeta(t) + R_m (1 - \zeta(t)) + R_0 \\ &\approx \begin{cases} R_m + R_0 & \text{for LRS.} \\ \frac{\eta k_B T}{eI(t)} \log \left(\frac{I(t)}{I_0(t)} + 1 \right) + R_{\text{ins}} + R_0 & \text{for multilevel HRS.} \end{cases} \end{aligned} \quad (14.30)$$

The corresponding simulation results shown in Figure 14.14 show that the extended model can reproduce the experimentally observed multilevel switching data of Ta/TaO_x/Pt cells. With increasing the RESET pulse amplitude, the filament becomes more oxidized, and consequently, a higher HRS resistance is achieved.

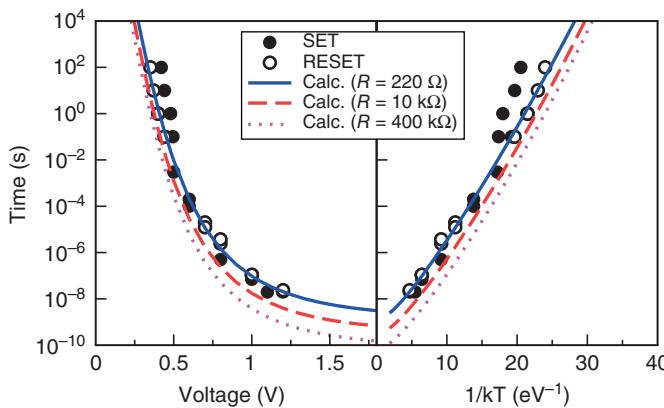


Figure 14.15 Measured and calculated SET/RESET times as a function of voltage (left) and as a function of $1/k_B T$, where T is the local temperature due to Joule heating (right). SET times are almost equal to RESET times, except for a small overvoltage effect

at long times possibly due to nucleation. The Arrhenius plot indicates an almost linear dependence on $1/k_B T$, thus supporting the temperature-activated model for migration. (Reproduced with permission from [65]. © 2012, IEEE.)

In the previous models, Joule heating has not been considered. The nonlinearity of the switching process in this case is mainly related to the electric-field-enhanced migration and the nonlinearity of the I - V relation. The latter aspect could be regarded as an implicit temperature modeling since the nonlinearity of the I - V relation highly affects the nonlinearity of the switching process when Joule heating occurs (cf. Section 14.5). Ielmini *et al.* developed a compact model that is based on the lateral growth of a conducting filament due to thermally and electric-field-enhanced ion migration [65]. The change in the filament diameter ϕ is calculated according to

$$\frac{d\phi}{dt} = \pm A_0 \exp\left(-\frac{\Delta W_{\text{hop}} - e\beta V}{k_B T}\right), \quad (14.31)$$

where A_0 denotes a rate prefactor and β is a barrier lowering factor. The temperature within the filament is approximated using

$$T = T_0 + \frac{V^2}{8\rho_{\text{el}}k_{\text{th}}}, \quad (14.32)$$

where T_0 is the ambient temperature, k_{th} is the filament thermal conductivity, and ρ_{el} is the filament resistivity. The latter is assumed to be diameter-dependent, according to Fuchs–Sondheimer theory. The filament resistance is assumed to be linearly dependent on the voltage. A more detailed description of the model is given in Chapter 11. By using this model, the SET and RESET switching kinetics could be reproduced over many orders of magnitude as illustrated in Figure 14.15.

At very fast switching times, the kinetics exhibit a flattening of the slope, that is, the nonlinearity decreases. This is in agreement with the results obtained for temperature acceleration, as discussed in Section 14.5.

Similar models have been proposed, for example, by Bocquet *et al.* [15] and Huang *et al.* [66]. In the latter model, vertical filament growth and lateral growth of the diameter based on ion migration are considered. The model proposed by Bocquet describes the change in filament diameter by an electron-transfer reaction instead of ion migration. The mathematical form, however, is more or less identical, and thus, similar results are obtained.

A statistical compact model was introduced by Degraeve *et al.* [67, 68]. In this model, the filament is represented by a top reservoir and a bottom reservoir connected by a constriction with variable width. Therefore, it is also called “hourglass model.” Single oxygen vacancy can move to/from the top and bottom reservoirs to/from the constriction. Which of these four processes happens is decided by a KMC algorithm. The amount of oxygen vacancies within the constriction defines the resistance state of the device, which is modeled using a quantum point contact model. Thus, only discrete I - V curves are valid. This model reproduces the variability in resistive switching of HfO_x -based VCM cells quite well. One of the major results is that the model predicts a “balance” during RESET transition. This means that there is only one stable state for a particular voltage during RESET. Hence, the system would relax to a certain balance resistance independent of the former resistance state, that is, higher or lower resistance.

14.7 Modeling of ECM Cells

14.7.1

Ab initio Models and MD Models

Based on the ECM SET operation principle, there are four different processes involved: oxidation and ion injection at the anode, ion migration, nucleation, and reduction. The charge state of the cations is not always obvious. For an Ag ion, the charge state is +1, whereas for Cu ions, the charge state could be +1 or +2 depending on the specific system. In the Cu_2S system, the charge state is +1 due to its bond nature. In contrast, the charge state in the other materials SiO_2 , Ta_2O_5 , a-Si, and Al_2O_3 cannot be easily foreseen. It depends, for example, on the ease of injecting a specific kind of ions from the active electrode into the switching layer, that is, the oxidation reaction. Furthermore, the mobility of the charged species within the switching layer is also important. Cyclic voltammetry experiments in the preforming stage of $\text{Cu}/\text{SiO}_2/\text{Pt}$ cells revealed that Cu^{1+} and Cu^{2+} ions can be injected but the dominating species is Cu^{2+} [69]. By using soft X-ray absorption spectroscopy, the bonding characteristics of Cu^{z+} ions in SiO_2 have been analyzed by Cho *et al.* [70]. This analysis reveals that Cu^{2+} ions form weaker bonds than Cu^{1+} , suggesting that Cu^{2+} ions are more mobile and play the dominating role in the resistive switching in Cu/SiO_2 ECM systems.

If Cu is used as active anode material, it is not clear which is the preferred charge state of the ions, +1 or +2. Sankaran *et al.* investigated the ion injection

and ion diffusion in CuTe/ α -Al₂O₃/Si ECM cells using DFT calculation and MD simulations based on empirical potentials [71]. Here, the Cu_xTe_{1-x} (0.5 < x < 0.7) electrode serves as ion source if a positive potential is applied. To determine the most stable ionic oxidation state in amorphous Al₂O₃, the injection energies for Cu and Te ions with charge states +2 to -2 were calculated. Figure 14.16 shows the calculated energies for different shifts of the Fermi level within the Al₂O₃. For an Al₂O₃ band gap of 5.1 eV, only the Cu²⁺, Cu⁺, and Te²⁺ ions are thermodynamically favored for injection. The same result was obtained for an Al₂O₃ band gap of 6.9 eV. Due to the stoichiometry of the CuTe electrode, the authors conclude that the Cu⁺ is favored over the Cu²⁺ species. In a second step, Sankaran *et al.* calculated the diffusion constant at different temperatures using MD simulations. Based on the simulation results, the diffusion coefficient and the activation energy could be extracted. While the activation energies for Cu⁺ (0.9 eV) and Te²⁺ (0.85 eV) are quite similar, the diffusion coefficient of Cu⁺ is approximately 2 orders of magnitude higher, that is, $5.8 \cdot 10^{-4}$ cm² s⁻¹ versus $3.4 \cdot 10^{-6}$ cm² s⁻¹. In addition, the injection of Cu⁺ ions is thermodynamically favored over the injection of Te²⁺ ions. This suggests that the copper ions are responsible for the switching in the CuTe/ α -Al₂O₃/Si stack. It was also found that the Cu⁺ ions tend to jump between low atomic density regions (cavities) within the α -Al₂O₃ [71]. This observation has been confirmed with MD simulations of Cu⁺ diffusion in the presence of aluminum or oxygen vacancies. While the activation energy stays almost constant, an increasing amount of vacancies increases the diffusion coefficient [71].

14.7.2

KMC Modeling

The forming and SET operations in ECM cells have been simulated by Pan *et al.* using a 2D KMC model [72, 73]. Figure 14.17a illustrates all the physical processes that have been considered in the model. Note that the nucleation process is not taken into account. Each of the processes has a different activation energy and obeys a rate law according to Eq. (14.1). Here, the activation energy is modulated within the electric field according to $-ae\Delta V$ and $(1 - \alpha)e\Delta V$ for forward and reverse transition, respectively. The local potential drop ΔV is calculated by solving the Poisson equation. During filamentary growth, the calculated current is ionic, but as soon as the filament bridges the insulator, the current is calculated using a resistor model for the filament. Electron tunneling between filament tip and counter electrode is not considered. This simulation model has been used to simulate the forming SET/operation. The resulting I - V characteristics are shown in Figure 14.17b along with the filament evolution. As nucleation is neglected, several filaments start to grow but only one bridges the insulator (2). As a voltage is still applied, a filament overgrowth occurs (3).

In a further simulation study, Pan *et al.* investigated the forming dynamics [73]. Two different scenarios were discussed: pulse forming and forming with voltage sweeps. In the pulse forming study, three different regimes could be identified (cf. Figure 14.18a), which can be loosely explained by the Butler–Volmer reaction. In

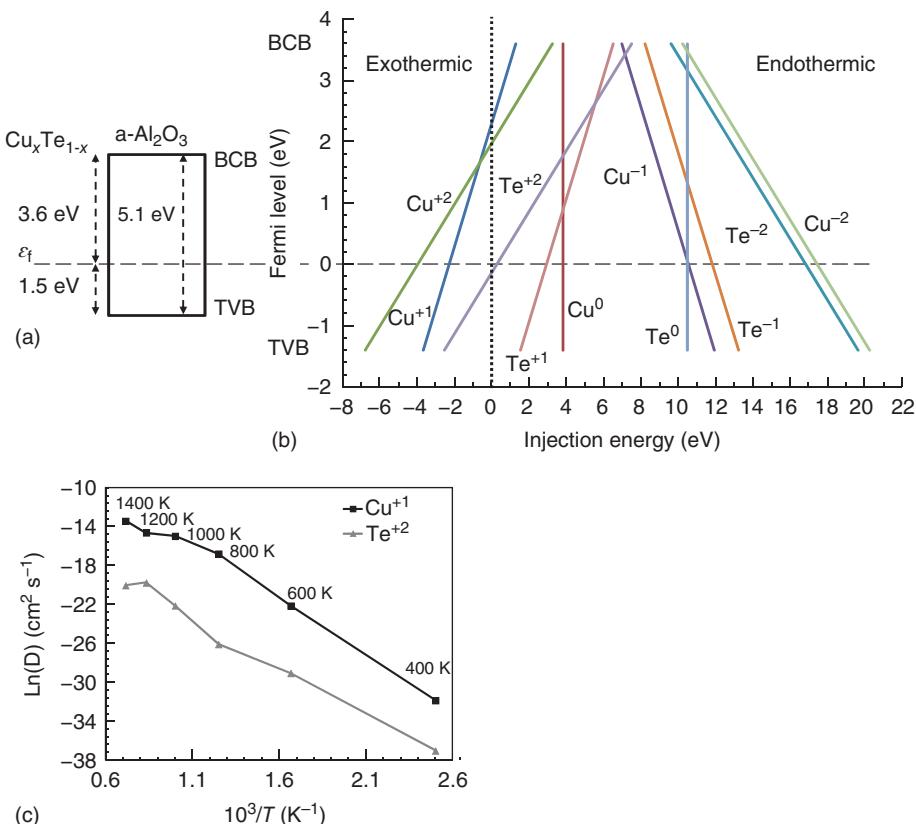


Figure 14.16 (a) Flat band diagram of $\text{Cu}_x\text{Te}_{1-x}$ aligned with $\alpha\text{-Al}_2\text{O}_3$ at the Fermi level with a valence band offset of 1.5 eV. (b) The evolution of the energies of injection of Cu and Te ions in $\alpha\text{-Al}_2\text{O}_3$ with respect to the position of the Fermi level in the band gap of 5.1 eV of $\alpha\text{-Al}_2\text{O}_3$ is delimited by the bottom of the conduction band (BCB) and the top of the valence

band (TVB). The dotted line separates the exothermic regime to the endothermic one. (c) Temperature dependence of the diffusion coefficient of Te^{2+} interstitials in $\alpha\text{-Al}_2\text{O}_3$ (gray lines) and compared with one of Cu^{1+} (black line). (Reproduced with permission from [71]. © 2012, The Electrochemical Society.)

regime (1) and (2), the surface overpotential is small and, thus, the cation flux is linearly dependent on the surface potential. In contrast, a high surface overpotential leads to an exponential increase in the cation flux. In addition, the amount of atoms that build up the filament is smaller for higher voltages, as the absorbed ions are immediately crystallized. During the forming process with voltage sweeps, only the regimes (2) and (3) could be identified as illustrated in Figure 14.18b. While the general trend in both simulation studies coincides with the experimental data of a $\text{Cu}/\text{SiO}_2/\text{Pt}$ cell, the absolute values deviate. Especially, for the pulsed forming, the experimental data shows 3 orders of magnitude change in forming

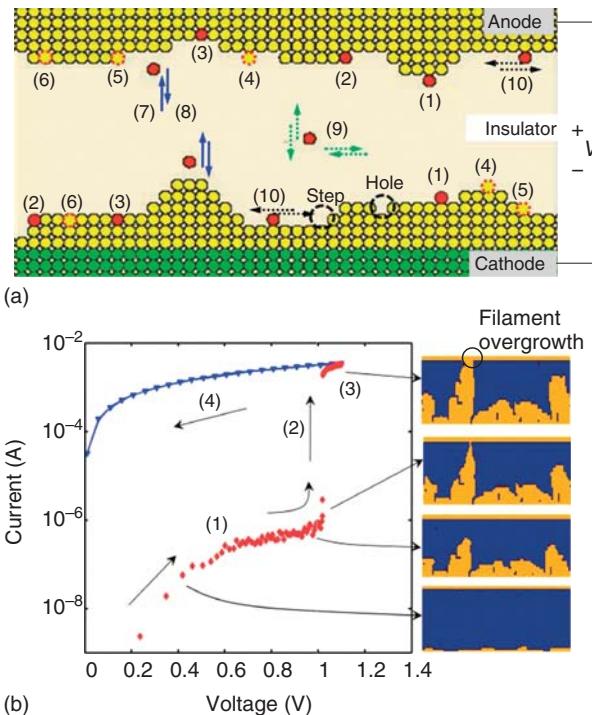


Figure 14.17 (a) Processes included in the KMC simulation. Oxidation at (1) adatom site (0.65 eV), (2) step site (0.7 eV), and (3) hole site (0.75 eV). Reduction at (4) adatom site (0.35 eV), (5) step site (0.3 eV), and (6) hole site (0.25 eV). (7) Adsorption (0.15 eV). (8) Desorption (0.3 eV). (9) Bulk diffusion (0.15 eV). (10) Surface diffusion (0.2 eV). (b) The KMC-simulated I - V characteristics

and related filament shape. The voltage sweep rate is 3 V s^{-1} . Region (1) is ionic current while (3) and (4) are ohmic current. The resistivity of metal filament is set to $5 \times 10^{-8} \Omega \text{m}$. The external system resistance is taken as $R_{ext} = 100 \Omega$. ((a) Reproduced with permission from [73]. © 2011, IEEE. (b) Reproduced with permission from [72]. © 2010, IEEE.)

time by varying the voltage by about 0.3 V. The simulation data only shows an acceleration of 2 orders of magnitude when the voltage is increased by 1.5 V. Also, the overall forming time is greater in the simulation.

Menzel *et al.* developed a refined 2D KMC model that also includes electron tunneling between the growing filament and the counter electrode as well as the influence of the mechanical stress that evolves during filamentary growth [75]. The former feature allowed for the simulation of multilevel switching over many orders of magnitude. The simulation results revealed that the mechanical stress influences the shape of the growing filament. If the Young's modulus of the host insulator matrix is very low, a dendritic growth occurs. In contrast, a wire-like filament evolves if the Young's modulus is high. Furthermore, the diameter of this filament is determined by the exact value of the Young's modulus [75].

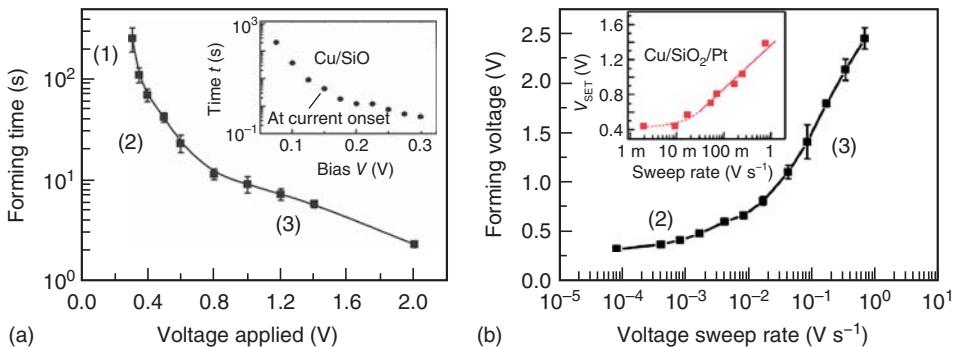


Figure 14.18 (a) KMC-simulated forming time t_{form} versus applied voltage. The inset shows the experimental results on Cu/SiO₂ [22]. (b) Simulated V_{form} versus voltage

sweep rate. The inset shows the experimental results on Cu/SiO₂/Pt system [74]. (Reproduced with permission from [73]. © 2011, IEEE.)

14.7.3

Continuum Modeling

Several groups developed continuum models for the simulation of SET and RESET operation in ECM cells [76–78]. A first 2D and 2D axisymmetric model was presented by Menzel *et al.* in 2009 [76]. To model the SET operation, a level set method was used, where the interface between the growing filament and the insulator is described by a transition function. This interface moved according to the current calculated using the continuity equation. In this model, only linear drift was considered. It could be shown that one filament “wins” during the SET operation due to a growth mechanism driven by a self-accelerated field.

In 2012, Lin *et al.* presented a more detailed 2D axisymmetric model also including the redox reactions occurring at the boundaries and nonlinear ionic transport [77]. The SET and RESET operations were modeled using the following equation system. The potential distribution was calculated solving the Poisson equation

$$\nabla \cdot (\epsilon \nabla \varphi) = -\rho. \quad (14.33)$$

The electrode/electrolyte interfaces were modeled using the Helmholtz layer (HL) model. Within the HL layer, the potential drops linearly (cf. Figure 14.19a and b) and is characterized by a plate capacitor C_h . Thus, Robin boundary conditions were used for the interfaces 3, 4, and 5. The complete set of boundary conditions reads

$$\left. \frac{\partial \varphi}{\partial r} \right|_{1,2} = 0, \quad \left. \frac{\partial \varphi}{\partial n} + \frac{C_h}{\epsilon} \varphi \right|_3 = \frac{C_h}{\epsilon} \varphi_h, \quad \left. \frac{\partial \varphi}{\partial n} + \frac{C_h}{\epsilon} \varphi \right|_{4,5} = 0. \quad (14.34)$$

The ionic transport within the electrolyte is modeled using the drift–diffusion equations for the silver ions c_{Ag^+}

$$\frac{\partial c_{\text{Ag}^+}}{\partial t} = \nabla \cdot \left(D \nabla c_{\text{Ag}^+} - \nu c_{\text{Ag}^+} \right), \quad (14.35)$$

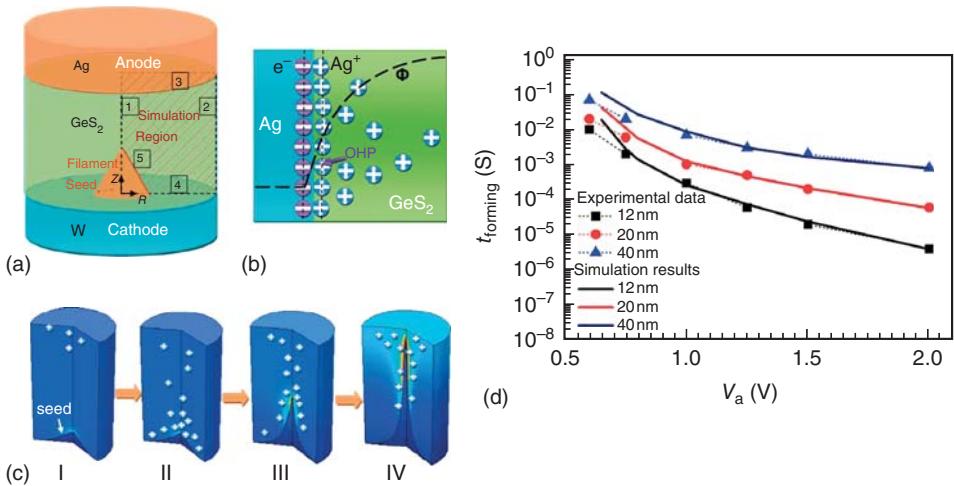


Figure 14.19 The schematic view of the simulated ECM cell. (a) The simulation region of the axisymmetric device is two-dimensional as indicated by the shadow region. The anode and cathode metals are Ag and W. The solid electrolyte layer is GeS₂. The cylindrical coordinate is centered at a cone-shaped Ag filament seed. (b) Electrostatic potential φ across the electric double layer (space charge accumulated at the filament-electrolyte interface according to the Stern model). OHP is short for the outer Helmholtz plane. (c) Schematically shows the simulated filament growth during

forming process: (I) Ag ions transport from the anode toward the cathode; (II) Ag filament starts growing when Ag ions accumulate on the seed surface; (III) self-accelerated growth of the Ag filament toward the anode; and (IV) a complete filament has formed through the electrolyte. The color in each figure shows the electric field intensity. (d) Forming time versus applied voltage V_{app} for Ag/GeS₂/W memory cell with different thicknesses. Solid lines and solid points represent the simulation results and experimental data [79], respectively. (Reproduced with permission from [77]. © 2012, IEEE.)

where v resembles the nonlinear ionic drift according to the Mott–Gurney law for ion hopping. In the HL, the Ag ions are reduced and the time-dependent change in the silver atom concentration c_{Ag} calculates to

$$\frac{\partial c_{\text{Ag}}}{\partial t} = \frac{k_B T}{h} \exp\left(-\frac{\Delta W_r}{k_B T}\right) c_{\text{Ag}^+} \exp\left(\frac{\alpha z e \Delta \varphi_h}{k_B T}\right). \quad (14.36)$$

where the reactions is driven by the voltage drop over the HL $\Delta \varphi_h$. If the silver concentration in a specific volume exceeds the bulk silver density, the filament boundary moves. As nucleation is not considered, an initial silver seed is assumed in the simulations. The transient evolution of the filamentary growth is depicted in Figure 14.19c. During SET operation a needle-shaped filament evolves. The simulation model and parameters were calibrated in order to match the experimental data of an Ag/GeS₂/W ECM cell. In this context, the obtained hopping distance of 2.3 nm appears to be too high. The simulation results for a pulsed forming operation are shown in Figure 14.19d compared to the experimental data. The achieved agreement is very good for high applied voltages but deviates at low voltages. In

this low voltage regime, the reaction kinetics are dominated by the ion transport, whereas the reaction kinetics limits the switching speed at high voltages. The model has also been used to study the RESET operation [77]. The possibility of simulating the *I-V* characteristics using this model has not been demonstrated yet.

In the previous models, a galvanic contact is always achieved in the ON state, resulting in a very low device resistance. In these cases, Joule heating needs to be considered as well. In addition, it has been reported that the RESET operation can be polarity-independent in this low resistance regime [80]. Menzel *et al.* introduced a 2D axisymmetric RESET model to explain the nature of this polarity-independency [80], which has been described in Chapter 11. Due to the high temperature along the filament, oxidation and reduction occur along the filament. Hence, the RESET operation can be understood as a thermally assisted self-dissolution of the conducting filament, and consequently, it is polarity-independent due to the thermal nature.

14.7.4

Compact Modeling

The first ECM compact models have been published in 2009 [76, 81]. Russo *et al.* described the final stages of the SET switching with a radial growth model [81]. In this model, the change in the filament diameter is determined by a nonlinear field-dependent Arrhenius-type law. This dependence was physically motivated by the Mott–Gurney law for ion hopping or by the Butler–Volmer equation. The model also included Joule heating due to the high currents involved. By using this model, the multilevel switching in a small low-resistance regime could be explained. Menzel *et al.* presented a vertical filament growth model for the simulation of the SET switching kinetics [76]. To account for the nonlinearity, the charge-transfer reactions occurring at both electrode/insulator interfaces were modeled using the Butler–Volmer equation. Later, Yu and Wong developed an empirical compact model for simulating the *I-V* characteristics, the SET and RESET switching dynamics, and the multilevel switching [82]. During the SET operation, first, a vertical filament growth is considered followed by a lateral growth similarly to the model designed by Russo *et al.* [81]. Correspondingly, the RESET starts with a lateral dissolution followed by a vertical one. The change in the filament height and diameter is driven by the ionic current according to the Mott–Gurney law for ion hopping. The local temperature change was estimated based on the dissipated power. The nonlinear switching kinetics are explained by nonlinear ionic drift. However, to fit the experimental data, an unphysical hopping distance $\geq 40\text{ nm}$ had to be considered. The multilevel switching was empirically obtained by setting the initial filament radius of the filament to certain value that corresponded to the experimental resistance data for the used current compliance.

The first self-consistent ECM model, which was able to simulate the multilevel switching, the switching kinetics, and the *I-V* characteristics with a single equation system, was presented by Menzel *et al.* [24]. In this model, the different resistance states are represented by the tunneling gap between the

growing filament and active electrode as described in detail in Chapter 11. The nonlinear switching kinetics were attributed to the charge transfer occurring at the metal/insulator boundaries according to the Butler–Volmer equation, while the ionic transport was assumed to be linear. Later, the authors developed this model further in order to include all possible physical processes limiting the switching speed: nucleation, charge transfer, and nonlinear ion drift [83]. This model is described here briefly as it represents the state of the art. A more detailed description is given in Chapter 11. The model considers a cylindrical filament, which grows in an insulating matrix as illustrated in Figure 14.20a along with the equivalent circuit diagram. It distinguishes between an ionic current path, which is responsible for the filamentary growth, and the electron tunneling between the filament tip and active electrode. The ionic drift within the insulating layer is modeled using the Mott–Gurney law for ion hopping, whereas the electron-transfer reactions (oxidation/reduction) occurring at the active electrode/insulator and filament/insulator interfaces are modeled using the Butler–Volmer equation. In the equivalent circuit diagram, these processes are described by the voltage controlled current sources η_{hop} , η_{ac} , and η_{fil} , respectively. The electron tunneling current is mathematically described by the linear Simmons equation for a symmetrical barrier [83]. The growth of the filament and, thus, the change in the tunneling gap x are described by Faraday’s law according to:

$$\frac{dx}{dt} = -\frac{M_{\text{Cu}}}{ze\rho_{m,Cu}} J_{\text{BV}} (\Delta\varphi_{\text{fil}}). \quad (14.37)$$

and depend on the ionic current and correspondingly on the overpotential η_{fil} , which is calculated using the equivalent circuit diagram shown in Figure 14.20a. The filamentary growth starts after the formation of stable nucleus. This nucleation time is calculated according to Eq. (14.5). By using this model, Menzel *et al.* were able to reproduce the switching kinetics of an Ag/AgI/Pt ECM cell as shown in Figure 14.20b. Three different slopes are visible. As discussed in Section 14.5, this indicates that different physical processes are limiting the switching speed in the corresponding voltage regimes. By analyzing the simulation data, the limiting processes in the three different regimes could be identified. In regime I (low voltages), the switching kinetics are determined by the nucleation speed. The electron-transfer reaction limits the switching speed in regime II. In contrast, at intermediate and high voltages, the switching speed is limited by the electron-transfer reaction and the ion-hopping transport.

The previous model was further extended by Menzel *et al.* to model the switching variability [84]. The filament in this model consists of individual atoms, and a layer-by-layer growth was assumed. In this way, the resistance state of the whole ECM device can be described by the number of atoms that build up the filament. Thus, the state variable is the number of atoms. In order to study the variability, a KMC approach was used. Here, two different processes are considered, which are derived from the Butler–Volmer equation. The reduction rate

$$\frac{1}{\tau_{\text{red}}} = \Gamma_{\text{red}} \exp \left(-\frac{\alpha z e}{k_B T} \Delta\varphi_{\text{fil}} \right) \quad (14.38)$$

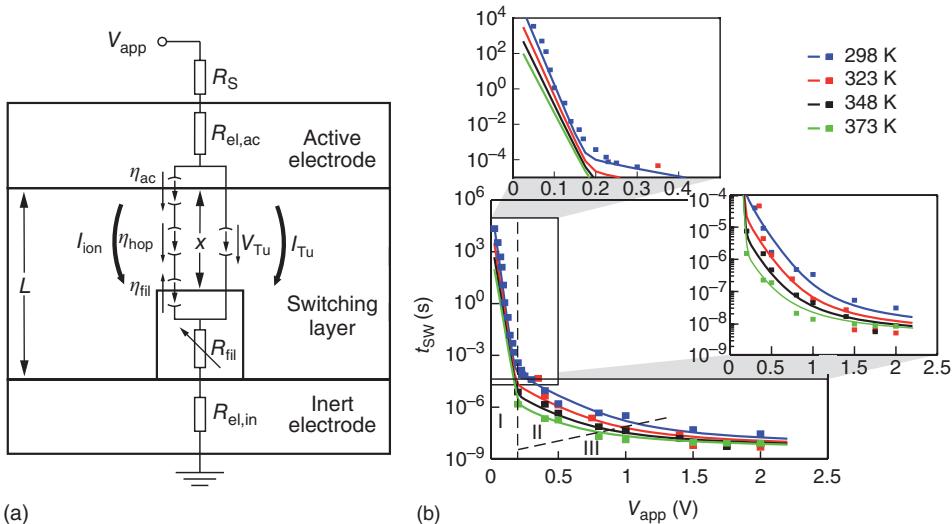


Figure 14.20 (a) Schematic of the switching model with an equivalent circuit diagram. A switching layer of thickness L is sandwiched between the active top electrode and the inert bottom electrode. A cylindrical filament grows within the electrolyte film and modulates the tunneling gap x between the filament and the active electrode. In the switching layer, both ionic and electronic current paths are present. (b) Pulsed SET

switching kinetics of the AgI-based ECM cell for different ambient temperatures $T = 298\text{ K}$ (blue), 323 K (red), 348 K (black), and 373 K (light green). The simulated data are displayed using solid lines and the experimental data using squares. I, II, III mark the nucleation-limited, the electron-transfer-limited, and the mixed-control regime, respectively. (Reproduced from [83] by permission of the PCCP Owner Societies.)

describes the attachment of one atom to the filament, and the oxidation rate

$$\frac{1}{\tau_{ox}} = \Gamma_{red} \exp \left(\frac{(1-\alpha)ze}{k_B T} \Delta\varphi_{fil} \right) \quad (14.39)$$

resembles the loss of one atom. One of these two processes is chosen randomly, and the number of atoms in the filament changes. The time step is then updated using a random number. Similarly to the previous model, a nucleation time is randomly determined prior to the simulation of the filamentary growth. For this, an exponential distribution function of the nucleation time is assumed with a mean value according to Eq. (14.5). Figure 14.21 shows the simulated switching kinetics using this model. For each voltage amplitude, 30 simulations were performed. The simulation results are in good agreement with the experimental data of Ag/AgI/Pt cell with respect to the mean values and the Weibull statistics. The simulation results are also consistent with the simulation results using the previously described deterministic model [83].

To model the effect of the EMF on ECM switching, Valov *et al.* [25] extended the model of Menzel *et al.* [24] by introducing a “nano”battery into the ionic current path as depicted in Figure 14.22a. The EMF voltage of this battery is modeled

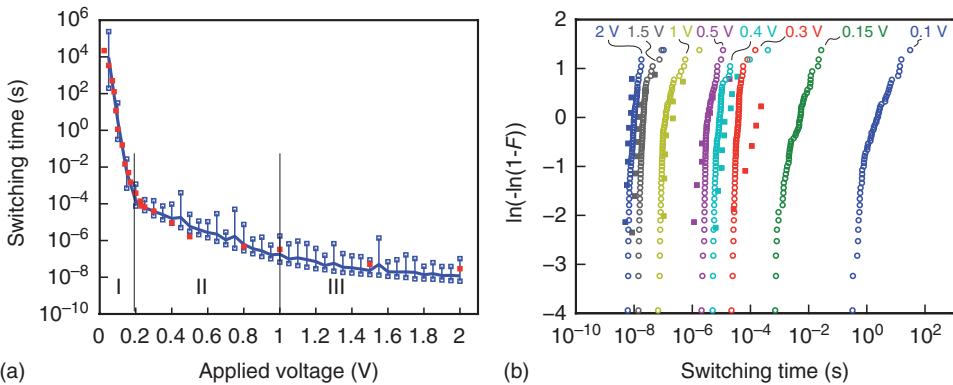


Figure 14.21 (a) Simulated switching kinetics using the 1D KMC model designed by Menzel *et al.* [84]. The blue line denotes the mean switching time, the open blue squares denote the maximum and minimum time, and the red squares denote the experimental data of an AgI-based ECM

cell. (b) Weibull statistics of the SET switching kinetic for different voltage amplitudes (open circles) compared to the experimental data of an AgI-based ECM cell (squares). (Reproduced with permission from [84]. © 2014, IEEE.)

according to

$$V_{\text{emf}}(c_{\text{ion}}) = V_0 + \frac{k_B T}{2e} \ln \left(\frac{c_{\text{ion}}}{c_0} \right), \quad (14.40)$$

and depends on the local ion concentration c_{ion} with respect to a reference concentration c_0 and an equilibrium voltage V_0 . The ion concentration acts as a second state variable besides the tunneling gap x . The change in both state variables linearly depends on the ionic current as

$$\begin{bmatrix} \dot{x} \\ \dot{c}_{\text{ion}} \end{bmatrix} = \begin{bmatrix} -K_1 \cdot I_{\text{ion}}(V, c_{\text{ion}}) \\ -K_2 \cdot I_{\text{ion}}(V, c_{\text{ion}}) \end{bmatrix} \quad \text{with} \quad \begin{array}{l} 0 \leq x \leq d \\ c_{\min} \leq c_{\text{ion}} \leq c_{\max} \end{array}, \quad (14.41)$$

where K_1 and K_2 are rate constants. The simulated I - V characteristics exhibit a non-zero-crossing as shown in Figure 14.22b, which reproduces the observed experimental behavior [25]. The nonzero-crossing is a consequence of adding the EMF to the model. This model has also been used to explain different modes of endurance failure [85]. If the RESET operation is incomplete, the switching voltages decrease with increasing number of cycles, finally leading to a “stuck in ON state” failure as illustrated in Figure 14.22c. In contrast, the ON state becomes unstable for EMF voltages in the range of the RESET voltage (Figure 14.22d).

14.8

Summary and Outlook

In this chapter, we reviewed the state-of-the-art modeling of ECM and VCM devices.

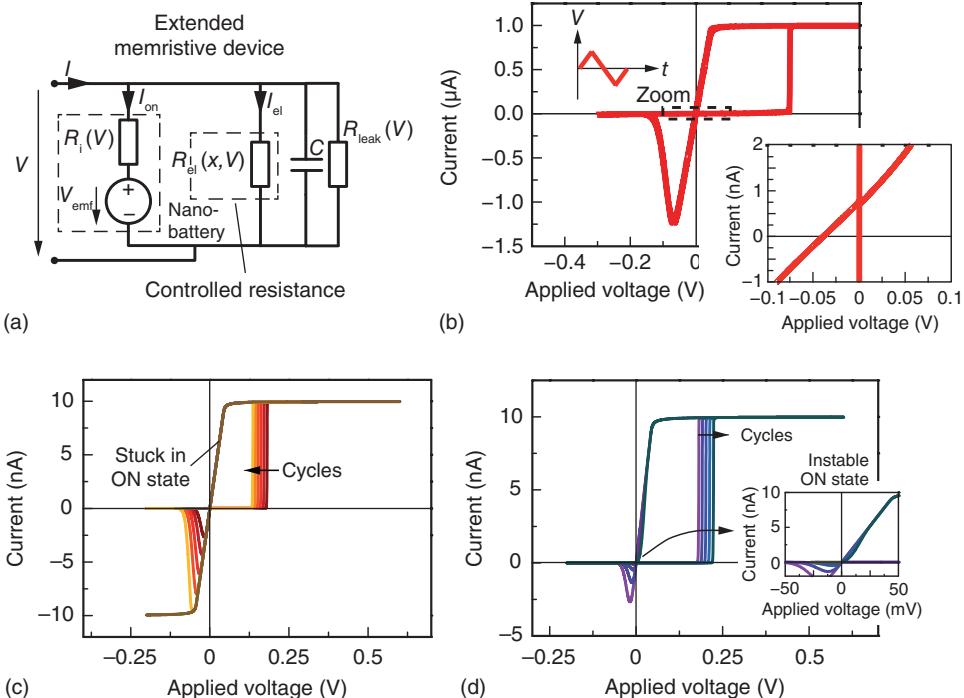


Figure 14.22 (a) Equivalent circuit of the extended memristive element. The ionic current is defined by the nanobattery, which controls the state-dependent resistor representing the electronic current path. The capacitance of the device is neglected as its influence is not significant. The partial electronic conductivity in the electrolyte induces a state-independent resistance R_{leak} due to a leakage current in parallel. (b) Simulated I - V characteristic of the extended memristor

model. The zoom shows the nonzero-crossing behavior. (c) The cell is stuck in ON state due to incomplete RESET operation. (d) Instable ON states are observed for emf voltages in the range of the RESET voltage. (Panels (a) and (b) are reproduced with permission from [25]. © 2013, Macmillan Publishers Limited. Panels (c) and (d) are reproduced with permission from [85]. © 2014, IEEE.)

The switching mechanism in ECM cells is quite well understood, and very predictive models are available. The switching kinetics of ECM cells are limited by different mechanisms in different voltage regimes: (i) nucleation, (ii) electron-transfer reactions, and (iii) ion migration. For VCM devices, the detailed mechanism is still under debate. Basic VCM models are available, but more fundamental research is required to improve the modeling. Particularly, the complex interaction between the atomic configuration of defects, electronic conductivity, Joule heating, and a subsequent configuration change are some of the major challenges in VCM modeling. In contrast to ECM cells, the highly nonlinear switching kinetics are achieved by both electric-field and temperature enhancement due to Joule heating.

Acknowledgment

SM would like to acknowledge the support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

1. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21** (25–26), 2632–2663.
2. Yang, J.J., Pickett, M.D., Li, X., Ohlberg, D.A.A., Stewart, D.R., and Williams, R.S. (2008) *Nat. Nanotechnol.*, **3** (7), 429.
3. Nian, Y.B., Strozier, J., Wu, N.J., Chen, X., and Ignatiev, A. (2007) *Phys. Rev. Lett.*, **98** (14), 146403/1–146403/4.
4. Yang, J.J., Strukov, D.B., and Stewart, D.R. (2013) *Nat. Nanotechnol.*, **8** (1), 13–24.
5. Waser, R., Bruchhaus, R., and Menzel, S. (2012) in *Nanoelectronics and Information Technology*, 3rd edn (ed R. Waser), Wiley-VCH Verlag GmbH, pp. 683–710.
6. Kwon, D.-H., Kim, K.M., Jang, J.H., Jeon, J.M., Lee, M.H., Kim, G.H., Li, X.-S., Park, G.-S., Lee, B., Han, S., Kim, M., and Hwang, C.S. (2010) *Nat. Nanotechnol.*, **5** (2), 148–153.
7. Yang, J.J., Miao, F., Pickett, M.D., Ohlberg, D.A.A., Stewart, D.R., Lau, C.N., and Williams, R.S. (2009) *Nanotechnology*, **20** (21), 215201.
8. Szot, K., Speier, W., Bihlmayer, G., and Waser, R. (2006) *Nat. Mater.*, **5** (4), 312–320.
9. Bersuker, G., Gilmer, D.C., Veksler, D., Kirsch, P., Vandelli, L., Padovani, A., Larcher, L., McKenna, K., Shluger, A., Iglesias, V., Porti, M., and Nafria, M. (2011) *J. Appl. Phys.*, **110** (12), 124518.
10. Yu, S., Guan, X., and Wong, H. (2012) *IEEE Trans. Electron Devices*, **59** (4), 1183–1188.
11. Gao, B., Yu, S., Xu, N., Liu, L.F., Sun, B., Liu, X.Y., Han, R.Q., Kang, J.F., Yu, B., and Wang, Y.Y. (2008) IEEE International Electron Devices Meeting 2008, Technical Digest, pp. 563–566.
12. Clima, S., Govoreanu, B., Jurczak, M., and Pourtois, G. (2014) *Microelectron. Eng.*, **120**, 13–18.
13. OHara, A., Bersuker, G., and Demkov, A.A. (2014) *J. Appl. Phys.*, **115** (18), 183703.
14. Xue, K., Traore, B., Blaise, P., Fonseca, L., Vianello, E., Molas, G., De Salvo, B., Ghibaudo, G., Magyari-Kope, B., and Nishi, Y. (2014) *IEEE Trans. Electron Devices*, **61**, 1394–1402.
15. Bocquet, M., Deleruyelle, D., Aziza, H., Muller, C., Portal, J.M., Cabout, T., and Jalaguier, E. (2014) *IEEE Trans. Electron Devices*, **61** (3), 674–681.
16. Lee, S., Lee, J.S., Park, J.-B., Kyoung, Y.K., Lee, M.-J., and Noh, T.W. (2014) *APL Mater.*, **2**, 066103.
17. Waser, R. and Aono, M. (2007) *Nat. Mater.*, **6** (11), 833–840.
18. Valov, I. and Kozicki, M.N. (2013) *J. Phys. D: Appl. Phys.*, **46**, 074005.
19. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005) *Nature*, **433** (6), 47–50.
20. Valov, I. (2014) *ChemElectroChem*, **1**, 26–36.
21. Kund, M., Beitel, G., Pinnow, C.U., Roehr, T., Schumann, J., Symanczyk, R., Ufert, K.D., and Mueller, G. (2005) Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20nm. IEEE International Electron Devices Meeting, 2005, IEDM Technical Digest.
22. Bernard, Y., Renard, V.T., Gonon, P., and Joussemaue, V. (2011) *Microelectron. Eng.*, **88** (5), 814–816.
23. Schindler, C. (2009) Resistive switching in electrochemical metallization memory cells. PhD thesis, RWTH, Aachen.
24. Menzel, S., Böttger, U., and Waser, R. (2012) *J. Appl. Phys.*, **111** (1), 014501/1–014501/5.
25. Valov, I., Linn, E., Tappertzhofen, S., Schmelzer, S., van den Hurk, J., Lentz, F., and Waser, R. (2013) *Nat. Commun.*, **4**, 1771.

26. Yang, Y., Gao, P., Li, L., Pan, X., Tappertzhofen, S., Choi, S., Waser, R., Valov, I., and Lu, W.D. (2014) *Nat. Commun.*, **5**, 4232/1–4232/9.
27. Menzel, S., Salina, M., Böttger, U., and Wimmer, M. (2015) *Adv. Funct. Mater.* DOI: 10.1002/adfm.201500825
28. Kamiya, K., Yang, M.Y., Park, S.G., Magyari-Köpe, B., Nishi, Y., Niwa, M., and Shiraishi, K. (2012) *Appl. Phys. Lett.*, **100** (7), 73502/1.
29. Kamiya, K., Yang, M., Nagata, T., Park, S., Magyari-Köpe, B., Chikyow, T., Yamada, K., Niwa, M., Nishi, Y., and Shiraishi, K. (2013) *Phys. Rev. B – Condens. Matter Mater. Phys.*, **87**, 155201
30. Yang, M.Y., Kamiya, K., Magyari-Köpe, B., Momida, H., Ohno, T., Niwa, M., Nishi, Y., and Shiraishi, K. (2013) *Jpn. J. Appl. Phys.*, **52**, 04CD11.
31. Kamiya, K., Yang, M.Y., Magyari-Köpe, B., Niwa, M., Nishi, Y., and Shiraishi, K. (2012) Proceedings of the 2012 IEEE International Electron Devices Meeting (IEDM), pp. 20.2.1–20.2.4.
32. Zhao, L., Park, S., Magyari-Köpe, B., and Nishi, Y. (2013) *Appl. Phys. Lett.*, **102** (8), 83506/1.
33. Zhao, L., Ryu, S.-W., Hazeghi, A., Duncan, D., Magyari-Köpe, B., and Nishi, Y. (2013) Proceedings of the 2013 Symposium on VLSI Technology.
34. Xue, K.-H., Blaise, P., Fonseca, L.R., and Nishi, Y. (2013) *Phys. Rev. Lett.*, **110** (6), 065502 (5 pp.).
35. Bradley, S.R., McKenna, K.P., and Shluger, A.L. (2013) *Microelectron. Eng.*, **109**, 346–350.
36. Kamiya, K., Yang, M.Y., Magyari-Köpe, B., Niwa, M., Nishi, Y., and Shiraishi, K. (2013) *IEEE Trans. Electron Devices*, **60**, 3400–3406.
37. Traore, B., Xue, K.-H., Vianello, E., Molas, G., Blaise, P., De Salvo, B., Padovani, A., Pirrotta, O., Larcher, L., Fonseca, L.R.C., and Nishi, Y. (2013) Proceedings of the 2013 IEEE International Reliability Physics Symposium (IRPS), pp. 5E.2.1–5E.2.6.
38. McKenna, K.P. (2014) *Modell. Simul. Mater. Sci. Eng.*, **22**, 025001.
39. Foster, A., Shluger, A., and Nieminen, R. (2002) *Phys. Rev. Lett.*, **89** (22), 225901/1.
40. Capron, N., Broqvist, P., and Pasquarello, A. (2007) *Appl. Phys. Lett.*, **91**, 2905.
41. Clima, S., Chen, Y.Y., Degraeve, R., Mees, M., Sankaran, K., Govoreanu, B., Jurczak, M., De Gendt, S., and Pourtois, G. (2012) *Appl. Phys. Lett.*, **100**, 133102–133102-4.
42. Metlenko, V., Ramadan, A., Gunkel, F., Du, H., Schraknepper, H., Hoffmann-Eifert, S., Dittmann, R., Waser, R., and De Souza, R. (2014) *Nanoscale*, **6** (21), 12864–12876.
43. Schie, M., Waser, R., and De Souza, R.A. (2014) *J. Phys. Chem. C*, **118** (28), 15185–15192.
44. Vandelli, L., Padovani, A., Larcher, L., and Bersuker, G. (2013) *IEEE Trans. Electron Devices*, **60** (5), 1754–1762.
45. Butcher, B., Bersuker, G., Gilmer, D., Larcher, L., Padovani, A., Vandelli, L., Geer, R., and Kirsch, P. (2013) Proceedings of the 2013 IEEE International, Electron Devices Meeting (IEDM), pp. 22.2.1–22.2.4.
46. Guan, X., Yu, S., and Wong, H. (2012) *IEEE Trans. Electron Devices*, **59** (4), 1172–1182.
47. Huang, P., Gao, B., Chen, B., Zhang, F.F., Liu, L.F., Du, G., Kang, J.F., and Liu, X.Y. (2012) Proceedings of the 2012 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 312–315.
48. Butcher, B., Bersuker, G., Vandelli, L., Padovani, A., Larcher, L., Kalantarian, A., Geer, R., and Gilmer, D.C. (2013) Proceedings of the 2013 5th IEEE International Memory Workshop (IMW), pp. 52–55.
49. Noman, M., Jiang, W., Salvador, P.A., Skowronski, M., and Bain, J.A. (2011) *Appl. Phys. A: Mater. Sci. Process.*, **102** (4), 877–883.
50. Strukov, D.B., Borghetti, J.L., and Williams, R.S. (2009) *Small*, **5** (9), 1058–1063.
51. Strukov, D.B. and Williams, R.S. (2009) *Appl. Phys. A-Mater. Sci. Process.*, **94** (3), 515–519.
52. Meuffels, P. and Schroeder, H. (2011) *Appl. Phys. Mater. Sci. Process.*, **105** (1), 65–67.
53. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser,

- R. (2011) *Adv. Funct. Mater.*, **21** (23), 4487–4492.
54. Larentis, S., Nardi, F., Balatti, S., Gilmer, D.C., and Ielmini, D. (2012) *IEEE Trans. Electron Devices*, **59** (9), 2468–2475.
55. Nardi, F., Balatti, S., Larentis, S., Gilmer, D.C., and Ielmini, D. (2013) *IEEE Trans. Electron Devices*, **60** (1), 70–77.
56. Kim, S., Choi, S., and Lu, W. (2014) *ACS Nano*, **8**, 2369–2376.
57. Marchewka, A., Waser, R., and Menzel, S. (2015) 2015 IEEE International Conference On Simulation of Semiconductor Processes and Devices (SISPAD), Washington, DC, 2015, pp. 297–300.
58. Marchewka, A., Rösgen, B., Skaja, K., Du, H., Jia, C.-L., Mayer, J., Rana, V., Waser, R. & Menzel, S. (2015) *Adv. Electron. Mater.* (accepted).
59. Schönhals, A., Wouters, D.J., Marchewka, A., Breuer, T., Skaja, K., Rana, V., Menzel, S., and Waser, R. (2015) Proceedings of the 2015 7th IEEE International Memory Workshop (IMW), pp. 73–76.
60. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) *Nature*, **453** (7191), 80–83.
61. Linn, E., Siemon, A., Waser, R., and Menzel, S. (2014) *IEEE Trans. Circuits Syst. I Regul. Pap. (TCAS-I)*, **61** (8), 2402–2410.
62. Pickett, M.D., Strukov, D.B., Borghetti, J.L., Yang, J.J., Snider, G.S., Stewart, D.R., and Williams, R.S. (2009) *J. Appl. Phys.*, **106** (7), 074508.
63. Hur, J.H., Lee, M.-J., Lee, C.B., Kim, Y.-B., and Kim, C.-J. (2010) *Phys. Rev. B*, **82**, 155321.
64. Hur, J.H., Kim, K.M., Chang, M., Lee, S.R., Lee, D., Lee, C.B., Lee, M.J., Kim, Y.B., Kim, C.J., and Chung, U.I. (2012) *Nanotechnology*, **23** (22), 225702/1–225702/5.
65. Ielmini, D., Nardi, F., and Balatti, S. (2012) *IEEE Trans. Electron Devices*, **59**, 2049–2055.
66. Huang, P., Wang, Y., Li, H., Gao, B., Chen, B., Zhang, F., Zeng, L., Du, G., Kang, J., and Liu, X. (2014) *IEEE Trans. Nanotechnol.*, **13**, 1127–1132.
67. Degraeve, R., Fantini, A., Clima, S., Govoreanu, B., Goux, L., Chen, Y.Y., Wouters, D.J., Roussel, Ph., Kar, G.S., Pourtois, G., Cosemans, S., Kittl, J.A., Groeseneken, G., Jurczak, M., and Altimime, L. (2012) Proceedings of the 2012 Symposium on VLSI Technology, p. 75–76.
68. Degraeve, R., Fantini, A., Raghavan, N., Chen, Y.Y., Goux, L., Clima, S., Cosemans, S., Govoreanu, B., Wouters, D.J., Roussel, Ph., Kar, G.S., Groeseneken, G., and Jurczak, M. (2013) Proceedings of the 2013 Symposium on VLSI Technology, pp. 98–99.
69. Tappertzhofen, S., Mündlein, H., Valov, I., and Waser, R. (2012) *Nanoscale*, **4** (10), 3040–3043.
70. Cho, D.Y., Tappertzhofen, S., Waser, R., and Valov, I. (2013) *Nanoscale*, **5** (5), 1781–1784.
71. Sankaran, K., Goux, L., Clima, S., Mees, M., Kittl, J., Jurczak, M., Altimime, L., Rignanese, G.-M., and Pourtois, G. (2012) *ECS Trans.*, **45** (3), 317–330.
72. Pan, F. and Subramanian, V. (2010) Proceedings of the 2010 IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 19–22.
73. Pan, F., Yin, S., and Subramanian, V. (2011) *IEEE Electron Device Lett.*, **32** (7), 949–951.
74. Schindler, C., Staikov, G., and Waser, R. (2009) *Appl. Phys. Lett.*, **94** (7), 072109/1–072109/3.
75. Menzel, S., Kaupmann, P., and Waser, R. (2015) *Nanoscale*, **7**, 12673.
76. Menzel, S., Klopstra, B., Kügeler, C., Böttger, U., Staikov, G., and Waser, R. (2009) *Mater. Res. Soc. Symp. Proc.*, **1160**, 101–106.
77. Lin, S., Zhao, L., Zhang, J., Wu, H., Wang, Y., Qian, H., and Yu, Z. (2012) Proceedings of the 2012 IEEE International Electron Devices Meeting (IEDM). pp. 26.3.1–26.3.4
78. Dorion, P., Cueto, O., Reyboz, M., Vianello, E., Barbe, J.C., Grigoriu, A., and Maday, Y. (2013) Proceedings of the 2013 IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 340–343.
79. Jameson, J.R., Gilbert, N., Koushan, F., Saenz, J., Wang, J., Hollmer, S., and Kozicki, M.N. (2011) *Appl. Phys. Lett.*, **99** (6), 063506.

80. Menzel, S., Adler, N., van den Hurk, J., Tappertzhofen, S., Valov, I., and Waser, R. (2013) Proceedings of the 2013 5th IEEE International Memory Workshop (IMW), pp. 92–95.
81. Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A.L., and Kozicki, M.N. (2009) *IEEE Trans. Electron Devices*, **56** (5), 1040–1047.
82. Yu, S. and Wong, H.-S. (2011) *IEEE Trans. Electron Devices*, **58** (5), 1352–1360.
83. Menzel, S., Tappertzhofen, S., Waser, R., and Valov, I. (2013) *Phys. Chem. Chem. Phys.*, **15** (18), 6945–6952.
84. Menzel, S., Wolf, B., Tappertzhofen, S., Valov, I., Böttger, U., and Waser, R. (2014) Proceedings of the 2014 6th IEEE International Memory Workshop (IMW), pp. 42–45.
85. Tappertzhofen, S., Linn, E., Böttger, U., Waser, R., and Valov, I. (2014) *IEEE Electron Device Lett.*, **35** (2), 208–210.

15

Valence Change Observed by Nanospectroscopy and Spectromicroscopy

Christian Lenser, Regina Dittmann, and John Paul Strachan

The resistance change in transition metal (TM) oxide redox-based resistive random access memory (ReRAM) is supposed to be caused by ion migration (typically mobile donors such as oxygen vacancies or cation interstitials) accompanied by a valence change in the metal ion. One obstacle for the elucidation of valence changes is that the net changes in the valence state induced during electroforming and switching are very small and occur primarily at the electrode interface or within nanoscale filaments. In this chapter, an overview of different spectroscopic approaches to detect valence changes in ReRAM materials, ranging from surface sensitive to bulk sensitive techniques, is given, providing different degrees of lateral and vertical spatial resolution. We present the state-of-the-art investigations of electrode–oxide interfaces and their valence changes upon electrical treatments in interface-type switching materials such as $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$. Moreover, we show that spatially resolved spectroscopic techniques provide valuable insights into filament formation and the corresponding structural and chemical changes in different types of filamentary-switching oxides such as SrTiO_3 (STO), TiO_2 , Ta_2O_5 , and HfO_2 .

15.1

Introduction

The physical properties of TM oxides are, to a large extent, governed by the oxygen stoichiometry and the associated formation of point defects (vacancies or interstitials) and ordered defect phases (e.g., Ruddlesden–Popper phases, Magnéli phases). Removal of a negatively charged oxygen ion from the oxide lattice creates a donor-type point defect (an oxygen vacancy or a cation interstitial), which carries a positive charge relative to the lattice and is usually compensated for by the introduction of electrons to maintain charge neutrality. These electrons are typically localized on the TM ions neighboring the donor-type defect, changing the valence state of that ion through the addition of an electron into an unoccupied state. The electrical conductivity that results from the additional charge carriers can be interpreted as local doping through point defects (self-doping). Since point

defects in ionic materials (see Chapter 4) carry a charge, they can move under the influence of an electric field (see Chapter 5) and alter the spatial distribution of the dopants. The valence change in the TM ion can, therefore, be seen as evidence of induced electrical conductivity, giving rise to the definition of “valence change memories” (VCMs).

Since these processes occur at the atomic scale, the direct investigation of the valence change demands that certain challenges are overcome, and methods for the direct observation of the valence change remain scarce. However, it is of key relevance to gain information of the changes occurring in the materials during electrical treatments since it is needed as input for material design and modeling of switching devices. Whether the investigated phenomenon is spatially confined in the lateral dimension (e.g., a conducting filament) or in the vertical direction (e.g., interface changes), the volume in which the valence change takes place is generally small and suitable techniques need to be employed to isolate the active volume from the passive environment. The spatial resolution challenge typically requires the probing lengths of electrons, X-rays, scanning probes, or a combination of these.

Another major challenge is the ratio of reduced metal ions to the fully oxidized ions. For example, substantial electronic conductivity in (STO) can be observed at electron concentrations of approximately 10^{18} cm^{-3} [1, 2], corresponding to a fraction of $\text{Ti}^{3+}/\text{Ti}^{4+}$ ions of roughly $10^{18}/10^{22} = 10^{-4}$. It is clear from this estimate that the detection of the reduced ionic species can be experimentally challenging due to a low total signal.

Investigation of the volume between the contacting electrodes poses a severe obstacle that has to be tackled. Various studies solve this issue by (i) preparation of cross-sectional slices of the device, with area selection typically guided by outer morphological changes observed; (ii) prefabrication of devices on thin windows, which are mainly transparent to high energy electrons or X-rays; or (iii) elimination of the top electrode altogether and instead utilizing a scanning probe (scanning tunneling microscopy (STM) or conductive atomic force microscopy (CAFM)) to induce the desired material changes relevant to resistance switching, which then allows the study of the switched area by, for example, surface techniques.

Fortunately, the advent of advanced characterization techniques in the laboratory, such as the development of aberration-corrected electron microscopes, and the expansion and improvement of synchrotron radiation facilities over the past decades have provided many informative insights into valence change systems.

Direct material studies related to resistance switching are relatively uncommon, and this is due to the need for an experimental method that simultaneously provides (i) high spatial resolution, (ii) the ability to probe material properties between electrodes, and (iii) the ability to detect a signal related, in some manner, to the valence change in the material. While most laboratory techniques are not able to overcome all three challenges, the following sections survey a number of methods that, in principle, can and discuss insightful work performed on several material systems including STO TiO_2 , Ta_2O_5 , HfO_2 , and $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$.

15.2

Methods and Techniques

In this section, an overview of the techniques that are most commonly used for the state-of-the-art investigations of valence change systems is presented. As shown in Figure 15.1, a coarse distinction of spectroscopy techniques can be made according to the excitation source and the detected signal, both of which can be either photons or electrons. As a rule of thumb, photons (particularly, in comparison to electrons) have a small scattering cross section with solid matter and, therefore, provide bulk sensitive information due to an attenuation length of up to several micrometers, while the typical mean free path for electrons in solids is on the order of a few nanometers due to their large scattering cross section.

The most commonly used technique based on the detection of photons is **X-ray absorption spectroscopy (XAS)**. This technique excites core-level electrons with either hard ($h\nu > 2 \text{ keV}$) or soft X-ray radiation, inducing electronic transitions and thus probing the electronic structure and local bonding environment of the absorbing atom. While the energy position of the absorption edge serves to uniquely identify the absorbing chemical species, the symmetry and strength of the crystal field can be related to the oxygen coordination of cations in oxides. Detection can be realized either in transmission mode or via a secondary signal emitted by a relaxation process, such as fluorescent X-rays, secondary electrons (total electron yield, TEY), or Auger electrons (Auger electron yield, AEY).

Specifically, in the hard X-ray regime, a specialized technique sensitive to the bonding environment and local symmetry of the absorber is **X-ray absorption fine structure spectroscopy (XAFS)**. XAFS can be subdivided into **X-ray absorption near-edge spectroscopy (XANES)** close to the absorption edge

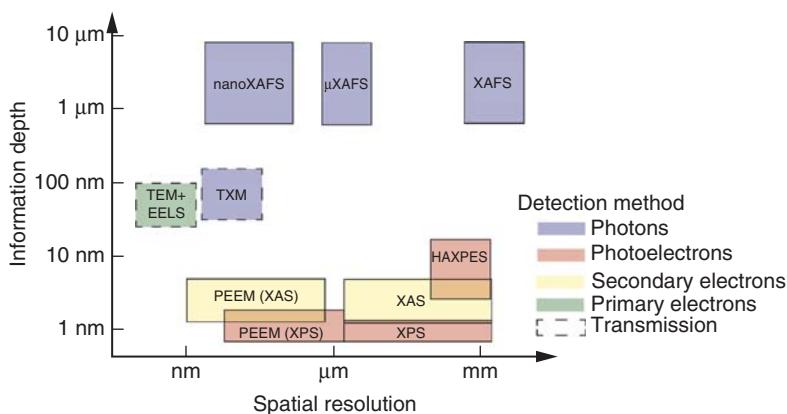


Figure 15.1 Schematic overview of the most common spectroscopy techniques used in VCM systems. The color of the boxes indicates the detection method, while the size of the box is related to the

dimensions accessible by each technique. The information depth of transmission techniques is determined by the permissible sample thickness.

($E - E_F \lesssim 30\text{--}50\text{ eV}$, where $E - E_F = 0$ marks the position of the absorption edge) and **extended X-ray absorption fine structure (EXAFS)** for energies $E - E_F \gtrsim 30\text{--}50\text{ eV}$. The former serves to examine local distortions in the absorber site, such as neighboring point defects, while the latter can provide information about the coordination number, bond lengths, and bond angles.

XAFS techniques can be combined with focusing optics to provide spatial resolution on the micro- or nanoscale (**μ -XAFS** and **nano-XAFS**, respectively).

Providing nanometer spatial resolution and chemical sensitivity at the same time, **(scanning) transmission X-ray microscopy ((S)TXM)** is a modern technique available at synchrotron radiation facilities that has been successfully used to analyze VCM stacks. A monochromatic X-ray beam is transmitted through a thin sample, where either the beam itself is focused and scanned across the sample (**STXM**) or the detection optics provide spatial resolution (**TXM**). Thin ($\sim 100\text{ nm}$) samples are a prerequisite since the inelastic mean free path of soft X-rays ($\lesssim 1\text{ keV}$) is short. Advantages are high spatial resolution, chemical sensitivity with a good energy resolution, and even subnanosecond time resolution using a fast point detector such as an avalanche photodiode. Despite the experimental challenges imposed by limitations in the sample thickness ($<1\text{ }\mu\text{m}$) and the necessity for the transmission geometry, some studies have shown the benefits of this technique [3–5].

Since the detection of electrons provides inherent surface sensitivity due to their short mean free path, a very popular surface analysis technique is **X-ray photoelectron spectroscopy (XPS)**, also known as **electron spectroscopy for chemical analysis (ESCA)**. This technique is based on the photoelectric effect where a sample is irradiated with X-rays of a characteristic wavelength, while an electron analyzer is used to detect the energy and momentum of the photoelectrons. The binding energy of the photoelectron E_{BE} is element specific, and the kinetic energy $E_{kin} = h\nu - E_{BE}$ determines the mean free path in the solid.

With the use of a standard laboratory X-ray source (Al K α radiation), emitted photoelectrons typically have kinetic energies from a few tens of electron volts to a few hundreds of electron volts, which can be associated with a mean free path of a few nanometers at most. Note that the mean free path is element-specific, but a general dependence of the inelastic mean free path electron on the kinetic energy can be given by the energy dependence of the scattering cross section $\delta \sim E^{0.5\text{--}0.75}$ [6, 7].

A fundamental challenge for the investigation of resistive-switching metal–insulator–metal (MIM) structures is the presence of the top electrode, which often consists of a heavy noble metal and therefore acts a strong absorber of the emitted photoelectrons due to the low mean free path. One way to mitigate this problem is to increase the photon energy of the excitation, a technique that is called **hard X-ray photoelectron spectroscopy (HAXPES)**. When combined with thin top electrodes in the range of a few nanometers, it is possible to probe buried interfaces and buried layers in a nondestructive way. For specific sample designs where either the top electrode can be removed or the electric treatment performed with a temporarily contacted electrode, such as a scanning probe tip,

HAXPES can also be used to study the bottom interfaces and deeper regions of a switched sample. Recently, advanced characterization techniques have been employed to investigate the switching in bilayer structures *in operando* using HAXPES. The formation of Ti suboxides was found during the electroforming of a Ti/HfO₂/TiN stack using electrical biasing, and further changes in the oxidation state of the Ti electrode were detected during switching [8, 9]. In a similar manner, valence changes could be observed for a Pt/TaO_x system using HAXPES [10] and in NiO_x using a “mobile” electrode [11].

In order to investigate the localized phenomena, the technique of choice is **photoemission electron microscopy (PEEM)**, which combines a typical XPS or XAS experiment with the spatial resolution provided by an electron imaging optics. With a spatial resolution down to 10 nm and good chemical sensitivity, PEEM is ideally suited to provide information on the lateral distribution of chemical states in an oxide sample. Information can be obtained in a very surface sensitive mode while detecting photoelectrons (XPS mode) or with a slightly increased information depth using slow secondary electrons (XAS mode). The latter variant typically offers better spatial resolution due to the higher electron counts. The low sampling depth in either case can be circumvented either by top electrode delamination [12, 13] or by the utilization of hard X-rays as the excitation source [14, 15].

The highest spatial resolution is offered by aberration-corrected **transmission electron microscopes (TEMs)**, which can provide atomic resolution in the ideal case. With the combination of **energy-dispersive X-ray spectroscopy (EDX)** and **electron energy loss spectroscopy (EELS)**, structural, elemental, and chemical analyses can, in principle, be performed on the subnanometer scale. The main drawback of this technique is the destructive sample preparation routine, which generally involves the preparation of a thin electron-transparent lamella via a focused ion beam (FIB) technique. During the preparation, surface damage due to implantation of Ga⁺ ions and removal of material can change the material properties dramatically. In addition, selecting the appropriate location for the FIB cut can be challenging since the switching location is not always obvious. Sample holders that provide the possibility of *in situ* experiments in the microscope via electrical connections offer a powerful way to circumvent this problem. Although not covered in this review, such techniques have been employed successfully in studies of unipolar thermochemical memories (TCMs) and bipolar electrochemical metallization (ECM) memories [16–18].

Advanced spectroscopy and spectromicroscopy techniques can provide important contributions to the field of resistive-switching memories and memristive devices. First and foremost, they provide the possibility to directly verify theoretical models and predictions in a real system and, ultimately, provide the crucially needed material information to improve and advance the existing models.

It is important for the experimentalist to choose the characterization technique appropriate for the investigated system. All of the techniques listed present a compromise between different aspects such as spatial resolution, chemical sensitivity, energy resolution, probing depth, and destructive or nondestructive sample preparation routes. Especially, destructive sample preparation routes – such as

XPS depth profiling by Ar⁺ sputtering – can introduce artifacts into the analysis due to preferential sputtering of oxygen and a concomitant reduction of the oxide through the sputtering process [19]. Furthermore, the necessity of sample cleaning methods (such as heating in ultra high vacuum or sputter cleaning) for surface sensitive techniques can lead to changes in the chemical state of the oxide sample and, therefore, have to be chosen and controlled very carefully [20].

15.3

Interface Phenomena

Several important phenomena in VCM-type resistive-switching devices are related to the chemistry at the metal/oxide interface. For the purpose of device characterization by spectroscopy, it is convenient to define such phenomena as interface-related, which can be examined with area-averaging techniques, that is, without the use of specialized focusing or magnifying optics. This definition serves to distinguish these from localized switching phenomena, which are associated with the formation of conducting filaments.

This section discusses the formation of oxide layers between a reactive metal electrode and an insulating or conducting oxide, the electrochemical manipulation of such oxide layers through electrical stimuli, and the impact of device performance.

15.3.1

Reactive Metal Layers on Insulating Oxides

Reactive metal electrodes deposited on insulating oxides can induce electrical conductivity through a redox-reaction involving oxygen transfer from the oxide to the metal. This phenomenon is exemplarily demonstrated in Figure 15.2a for sputtered Ti layers deposited on top of epitaxial, insulating Fe-doped STO (Fe:STO) films of 20 nm thickness. The initial resistance R_0 – as read out by a low voltage signal below the switching threshold – decreases abruptly between Ti layers of 5 and 10 nm thickness, marking a transition from an insulating system that requires electroforming in order to enable resistive switching properties to a conducting system without the need for electroforming [21].

Using HAXPES, it was possible to identify the underlying mechanism of the increased conductivity, which turned out to be a redox-reaction between the electrode and insulator $\text{Ti} + \text{SrTiO}_3 \leftrightarrow \text{TiO}_x + \text{SrTiO}_{3-x}$. Figure 15.2b shows the photoemission spectra obtained using 4.2 keV X-rays for excitation. The escape depth of the photoelectrons in Ti increases to >10 nm when the electron kinetic energy is greater than 3.5 keV, enabling the examination of the buried Ti/STO interface.

Through least-squares fitting techniques using Lorentzian–Gaussian functions, the shoulder at the low binding energy side of the Ti 2p emission doublet can be attributed to the existence of multiple Ti oxidation states in the interface region of a 4 nm thin Ti layer on STO. Although there is no way to distinguish between Ti

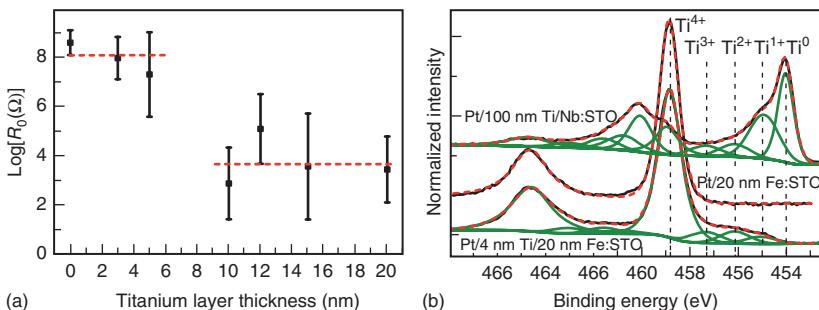


Figure 15.2 (a) Thickness dependence of the resistance of Ti/20 nm Fe:STO interfaces with varying Ti thickness. The resistance drops by several orders of magnitude when 5 nm Ti are exceeded (red lines are a guide to the eye). (b) Hard X-ray photoemission

spectroscopy (HAXPES) data of Ti layers of various thicknesses on STO. All layers are capped by a Pt layer to prevent surface oxidation. (Reprinted with permission from [21]. © 2012, AIP Publishing LLC.)

ions in the oxide and the electrode, the absence of a metallic Ti^0 is a firm indicator of oxygen removal from the oxide and transfer into the electrode.

The deposition of a reactive Ti electrode can, therefore, be used to induce significant conductivity in the band-gap insulator $SrTiO_3$, presenting a viable way of controlling the redox state of the oxide through simple process engineering. This phenomenon is the basis for many state-of-the-art VCM systems in existence to date [3, 22, 23].

15.3.2

Formation of a Blocking Layer on Conducting Oxides

For all VCM-type systems, the interface between the active electrode and the oxide layer plays an important role in device performance, for example, as a source of nonlinear charge carrier transport. For selected systems, however, the interface provides the functionality of the device through reversible oxidation and reduction of a nonstoichiometric oxide layer at the interface. This so-called interface-type switching is most often encountered between conducting oxides, such as doped manganites, and high-oxygen-affinity electrode metals such as Ti, Ta, and Al [24–27]. Detailed photoemission studies during the deposition of Al on $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO) have demonstrated both the oxidation of the growing Al layer and the reduction of the underlying PCMO, as shown in Figure 15.3a,b [28].

The oxygen transfer from the oxide to the metal layer can be understood in terms of thermodynamic driving forces, where the Gibbs free energy change ΔG of the reaction $Al + Pr_{0.7}Ca_{0.3}MnO_3 \leftrightarrow AlO_x + Pr_{0.7}Ca_{0.3}MnO_{3-x}$ is negative and drives the reaction forward. Since the process takes place at room temperature, the diffusion length of oxygen is very short and the oxidation is restricted to a very thin layer at the interface between the oxide and the electrode. This

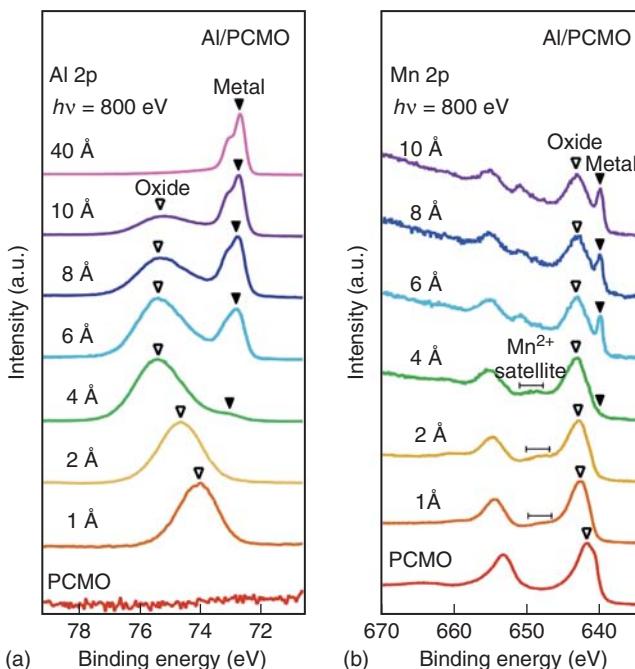


Figure 15.3 (a, b): Al 2p and Mn 2p XPS spectra, respectively, taken after the subsequent deposition of thin Al layers on a PCMO thin film. The entire experiment was

performed without breaking vacuum to prevent surface oxidation from the ambient air. (Reprinted with permission from [28]. © 2010, AIP Publishing LLC.)

interfacial oxide layer plays an important role in the electrical transport across the interface, increasing the resistivity by the formation of an insulating oxide layer [26, 29].

15.3.3

Electrically Induced Redox Reactions at the Interface

For memristive systems based on conducting oxides (such as manganites or cobaltites) and reactive metal electrodes, electrical conduction is often homogeneous across the entire device area [25, 30]. Such devices generally exhibit low electrical resistance after fabrication, which can be increased by an appropriate voltage treatment with the positive bias applied to the top electrode. By using cross-sectional TEM complemented by EELS and HAXPES, it has been shown that this electroforming treatment increases the thickness of the interfacial oxide layer significantly [29, 31]. While TEM can observe the thickness increase directly, the nondestructive examination by HAXPES shows a spectral weight transfer from low to high oxidation states of the Ti. Due to the limited probing depth of photoemission, this amounts to an increase in the thickness of the TiO_x layer.

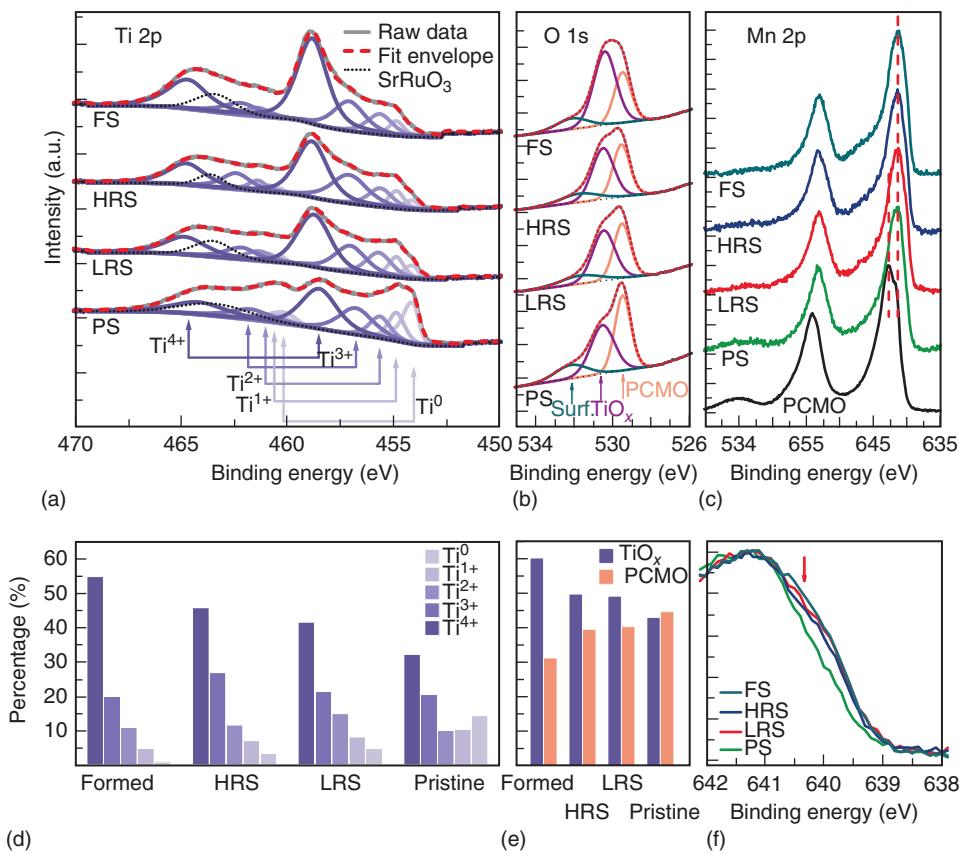


Figure 15.4 (a–c) Ti 2p/O 1s/Mn 2p XPS spectra collected on the four regions, sorted from low resistance (pristine) to high resistance (formed). The spectrum named PCMO refers to the unpatterned reference sample. (d, e) Area percentages of the constituent

components of the Ti 2p and O 1s envelope, respectively. (f) Close-up of the Mn 2p_{3/2} edge. (Reproduced from [32] © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.)

Figure 15.4 demonstrates in detail how the application of an external electrical stimulus can reversibly change the valence state in a Ti/PCMO junction. The pristine state (PS) of the junction was measured without any electrical treatment, while a junction in the formed state (FS) was examined after the application of a positive voltage sweep to the Ti top electrode that induced a nonvolatile change in resistance. Low-resistance state (LRS) and high-resistance state (HRS) refer to states obtained by subsequent voltage treatments with negative and positive bias, respectively. The valence change in the Ti top electrode is readily observed from the Ti 2p HAXPES spectra shown in Figure 15.4a. The low binding energy component associated to metallic Ti (denoted Ti⁰) decreases with increasing resistance, while the high binding energy component corresponding to fully

oxidized Ti (marked Ti^{4+}) increases ($\text{PS} \rightarrow \text{LRS} \rightarrow \text{HRS} \rightarrow \text{FS}$), see Figure 15.4d for the area percentages. The O 1s emission line shown in Figure 15.4b directly demonstrates the oxygen transfer from the PCMO to the TiO_x layer (area percentages shown in Figure 15.4e)), while the Mn 2p line shows slight changes in the line shape associated with an increase in the ratio of $\text{Mn}^{3+}/\text{Mn}^{4+}$ ions when moving from low to high resistance. The observation of oxidation of the Ti layer and reduction of the PCMO is direct evidence that the resistance change is caused by a redox reaction at the interface [32].

15.4

Localized Redox Reactions in Transition Metal Oxides

Localized switching effects are often referred to as *filamentary* switching since the valence change is localized to a limited area, rather than homogeneous across the metal/oxide interface, and often with dimensions well below 1 μm . For the purpose of spectroscopy, these phenomena require the use of spatially resolved techniques to locate the switched volume. This section provides an overview of those studies and their contribution to the understanding of VCM.

15.4.1

Single Crystalline Model System – Doped SrTiO_3

Several studies on acceptor-doped SrTiO_3 have shown that after electrical or chemical reduction, the presence of oxygen vacancies in the first coordination shell of TM dopants on the B-site can be identified by K-edge XANES of the dopant, while the Ti K-edge yields no evidence of the same [33–35]. This fact has led to the utilization of TM dopants as “tracers” for the valence change in SrTiO_3 , a phenomenon that is based on the bond enthalpy.

Of the TMs in the third period of the periodic table, the bond enthalpy of the $\text{Ti}-\text{O}$ bond is the highest, that is, the $\text{Ti}-\text{O}$ bond is the most stable [35]. Any oxygen ions removed from the system are therefore preferentially removed from nextneighbor sites around TM dopants. Due to the overall lower concentration of dopants and the preferred reduction of dopant ions, a small change in the oxygen-vacancy concentration leads to a relatively large change in the fraction of dopant ions [TM] that have an oxygen vacancy associated with them $[\text{TM} - V_{\text{O}}^{**}]/[\text{TM}]$. The association causes the local deformation of the crystal lattice around the dopant ion and reduces the site symmetry, which can be detected by appropriate spectroscopic techniques. The technique of choice is X-ray absorption near-edge structure (XANES) spectroscopy, which, as described in Section 15.2, is very sensitive to changes in the local bonding configuration around the absorbing ion and allows the separation of contributions from B-site ions in unperturbed ($\text{B}-\text{O}_6$) and distorted ($\text{B}-\text{O}_5\text{V}$) environments. *Ab initio* calculations performed for $\text{Fe}^{3+}-\text{O}_6$ and $\text{Fe}^{3+}-\text{O}_5\text{V}_{\text{O}}^{**}$ centers demonstrate that the valence state of the

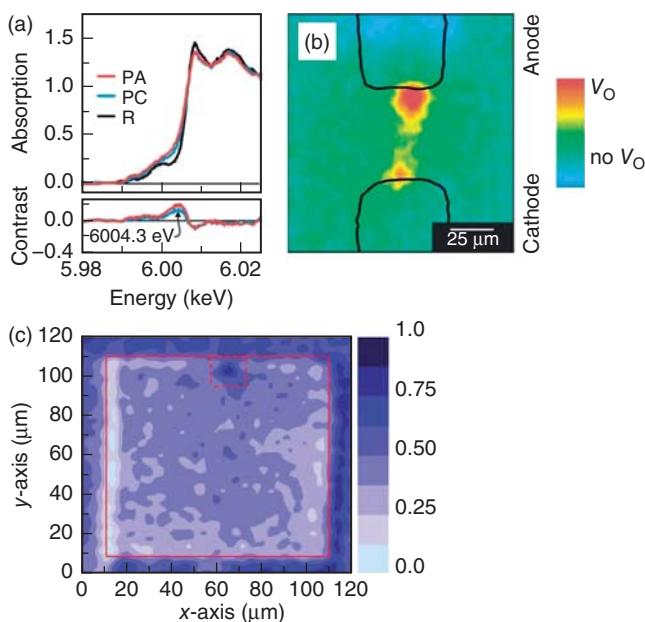


Figure 15.5 (a) Cr K-edge XANES spectra of Cr-doped SrTiO_3 single crystals taken near anode (PA), cathode (PC), and a reference (R). Plotting the spatially resolved data for the excitation energy of 6004.3 eV yields a false color map of the oxygen-vacancy distribution in (b). (Reprinted with permission from [33, 37]. Copyright © 2007 WILEY-VCH Verlag GmbH & Co. KGaA,

Weinheim.) (c) Fe K-edge fluorescence map of a $100 \times 100 \mu\text{m}^2$ MIM device (MIM = Pt /Fe-doped SrTiO_3 epi-film/Nb-doped SrTiO_3 substrate) recorded at 7122 eV excitation energy, indicating the spatial distribution in analogy to (b). The location of the conducting filament is marked with a dashed, red square. (Reprinted with permission from [37]. © 2012, AIP Publishing LLC.)

absorbing ion is much less important for the absorption characteristics in the XANES region than the presence of the oxygen vacancy [34].

The first studies that pioneered the field of spectroscopy on memristive systems were performed on a single crystalline model system consisting of Cr-doped SrTiO_3 with a lateral electrode array [33, 36]. As shown in Figure 15.5a,b, it could be demonstrated that a conducting channel had formed between two electrodes during electrical treatment and that this channel contained a large number of oxygen vacancies. The presence of an oxygen vacancy in the first coordination shell of Cr alters the near-edge structure of the Cr K absorption edge, which can be spatially resolved by scanning the sample under a microfocused X-ray beam.

A similar approach was used on MIM structures based on epitaxial Fe-doped STO thin films (Figure 15.5c), confirming the prevalent theory that a single conductive filament forms during electroforming, which can be reversibly switched close to the anode [37]. As compared to the single crystal situation, the spectral change at the filament location is more pronounced, which might be an indication that the much larger electric fields involved in switching the thin film

induce additional structural changes or that switching is preferentially located in highly defective areas of the thin film [38].

15.4.2

Localized Structural and Compositional Changes in TiO_2

A well-studied VCM system is TiO_2 , which can exhibit both unipolar and bipolar switching typically dictated by the film properties, electrode metal choice, and the current compliance utilized [39]. The unipolar-type switching in the commonly used $\text{Pt}/\text{TiO}_2/\text{Pt}$ stack was investigated with high-resolution TEM (HRTEM) and *in situ* transport measurements [40], while the same system exhibiting bipolar switching has been studied by X-ray nanospectroscopy and TEM-based electron diffraction [4]. In the first work by Kwon *et al.*, a large-area ($\sim 300 \mu\text{m}$ diameter) device was cross-sectioned by FIB after electroforming and resistance switching. Careful study of these cross-sectioned lamella revealed the presence of substoichiometric nanocrystalline $\text{Ti}_n\text{O}_{2n-1}$, particularly Ti_4O_7 , otherwise known as *Magnéli phases*, as shown in Figure 15.6. The Magnéli nanocrystals either spanned or partially spanned the thickness of the oxide layer, and in association with these substoichiometry regions, the evolution of oxygen was indirectly observed by the development of “blown-off” regions of the top electrode. Such gas eruption features have been frequently observed in the TiO_2 system as well as in others. While this technique – involving cross-sectional slices – made it difficult to confirm that the discovered Magnéli nanocrystals were responsible for the switching, the team conducted temperature-dependent transport measurements of intact devices in the SET state, demonstrating the characteristic metal–semiconductor transition (at 130 K) of Ti_4O_7 . *In situ* switching using a STM tip further demonstrated that the reset switching of a filament eliminated the Ti_4O_7 Magnéli phase in favor of an anatase phase.

X-ray studies of $\text{Pt}/\text{TiO}_2/\text{Pt}$ also revealed that resistance switching is associated with the development of a suboxide region in the device, as shown in Figure 15.7. The main differences in this study are that they applied to the bipolar switching mode (roughly 100 \times less current during switching), and instead of cross-sectioning, the device was fabricated atop a thin silicon nitride membrane to give a top view perspective of the MIM crossbar structure. By using STXM, a chemical mapping was performed across the area of the device, showing the amorphous as-grown TiO_2 film (Figure 15.7a, green area), a region including the junction area, which underwent a phase transformation to anatase (red), and a 300–500 nm diameter area (blue) showing a substoichiometric composition TiO_{2-x} . The chemical reduction of TiO_2 is signaled in the X-ray absorption spectra shown in Figure 15.7b by the presence of Ti^{3+} (peak at 456 eV). Ti_4O_7 , in particular, possesses an equal fraction of Ti^{4+} and Ti^{3+} , and TEM-based electron diffraction showed that the substoichiometric region is the Magnéli phase Ti_4O_7 [4].

The formation of the Magnéli phases in unipolar and bipolar TiO_2 devices is notable for the emergence of long-range ordering, given sufficiently high concentrations of vacancies and a thermodynamic driving force, such as that encountered

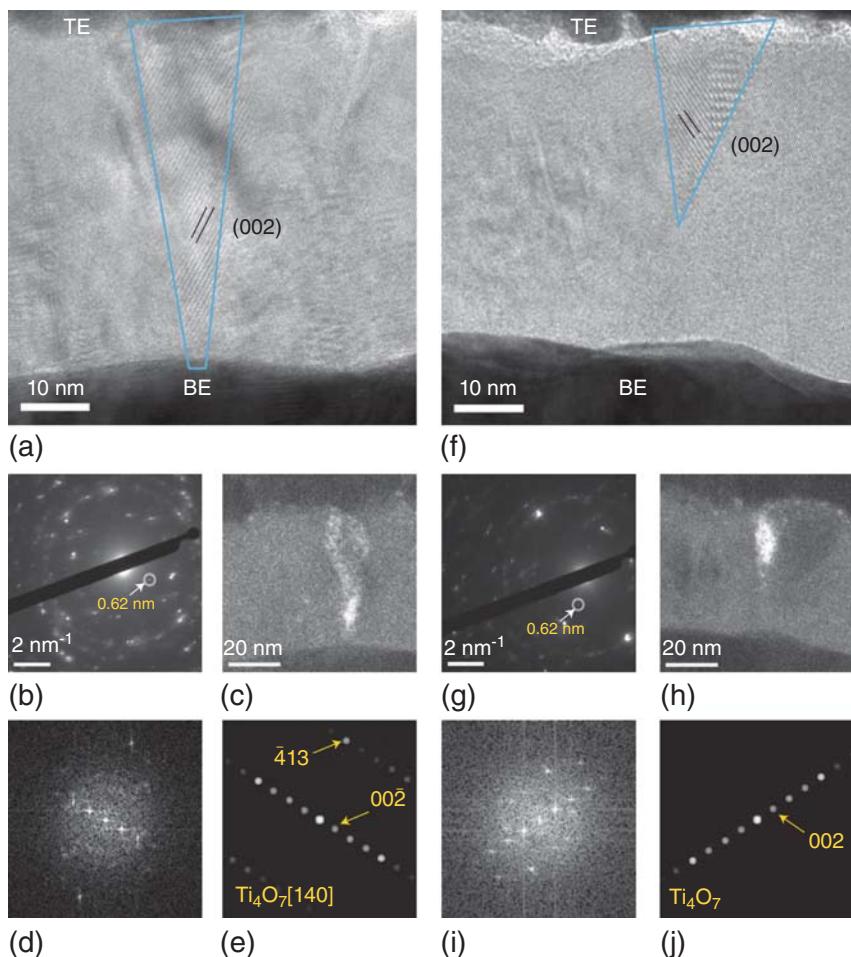


Figure 15.6 HRTEM study of Pt/TiO₂/Pt resistive-switching device after electrical switching and cross-sectioning. Ti₄O₇ nanocrystals are observed, which either

span the entire thickness (a–e) or show an incomplete progress (f–j). (Reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [40], © 2010.)

in the V–O, Ti–O, W–O, and Mo–O systems. With respect to resistance switching, it is not conclusively known if the additional thermodynamic stability in the resulting vacancy ordering compared to an amorphous arrangement results in enhanced switching behavior such as reduced voltage or higher repeatability. Due to the large Joule heating generated in localized switching, the structural transformation may instead be a result of an annealing effect. In both studies, the development of polycrystalline phases from initially disordered films was observed, including in areas outside of the direct switching channels. Additionally, the long-range ordering is not always observed in bipolar TiO₂ devices [41].

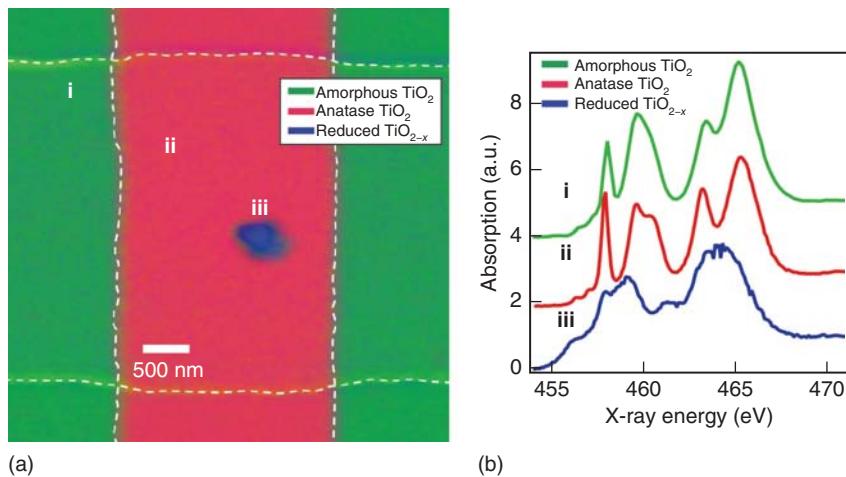


Figure 15.7 Scanning transmission X-ray microscopy (STXM) of a Pt/TiO₂/Pt bipolar resistive-switching device after electroforming and switching. Top view (a) of crossbar with associated mapping for the three different identified phases of titanium oxide. The as-grown blanket TiO₂ film was amorphous but developed a nanocrystalline

anatase phase within and partly extending beyond the junction region. A 300–500 nm diameter area exhibited a substoichiometry in the presence of Ti³⁺. (b) Corresponding X-ray absorption spectroscopy (XAS) of the three identified phases. (Reproduced from [4]. © 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.)

Conversely, modifying the stack layers to include one with high concentration of disordered vacancies, a-TiO_{2-x}, in series with a thin stoichiometric TiO₂ results in improved resistance switching without electroforming [42].

15.4.3

Compositional Changes in Ta₂O₅ and HfO₂

Two important binary oxide materials of increased interest since 2008 are Ta₂O₅ and HfO₂, which have demonstrated the highest cycle endurances (up to 10¹² cycles) in any VCM system [22, 43–45]. In contrast to the Ti–O system, for these two systems, there are no known ordered phases of vacancies, and in fact, only one thermodynamically stable oxide phase exists, namely Ta₂O₅ and HfO₂, with all other suboxides being metastable. The switching oxide films are typically polycrystalline or amorphous and deliberately substoichiometric either by growth or by chemical reaction with an electrode such as Ti, Ta, or Hf instead of noble metals in order to reduce the voltages required for switching and to improve cyclability. Because of this, and since the SET states of both systems are typically low resistance (nearly metallic) and therefore not described by the terminal oxide phase, the actual switching material is often referred to as TaO_x and HfO_x.

The TaO_x system has been studied by both TEM-EELS and nano-XAFS [46–48]. Figure 15.8a–c shows the results of cross-sectional analysis of a device

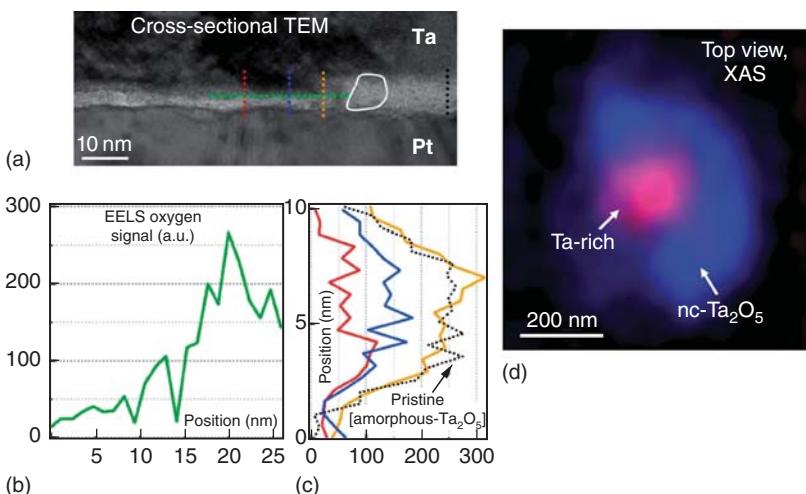


Figure 15.8 TEM and XAS of the conducting channel region of two different Ta_2O_5 resistive-switching devices. (a) TEM of a device region identified as controlling the transport. (b) Horizontal EELS line profile across the channel region, green line in (a), showing a substantial drop in the measured oxygen intensity. (c) Corresponding

EELS vertical line profiles in several positions outside and within the conducting region. (d) Separate device probed by hard X-ray absorption spectroscopy and microscopy showing a Ta-rich region surrounded by a structurally altered phase. (Reproduced from [46]. © 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.)

region determined to control the resistance switching based on CAFM analysis. HRTEM revealed that the conducting channel remained amorphous, while surrounding regions underwent partial crystallization to the terminal oxide Ta_2O_5 , due to Joule heating. However, the O/Ta ratio decreased substantially within the conducting channel, as shown in Figure 15.8b,c. These observations were supported by X-ray absorption mapping of a separate device tuned to the Ta L_3 edge (9.89 keV). By utilizing a zone plate for X-ray focusing and raster scanning the sample junction area, this study revealed a Ta-rich phase (~ 200 nm diameter) surrounded by nanocrystalline Ta_2O_5 . More direct *in situ* TEM studies of $\text{Pt}/\text{SiO}_2/\text{Ta}_2\text{O}_5/\text{TaO}_{2-x}/\text{Pt}$ devices were performed in both the SET and RESET states revealing the generation of multiple percolating paths through the SiO_2 layer having the composition TaO_{1-x} [47]. Filaments with a diameter of only a few nanometers were also discovered in the a- TaO_{2-x} layer, as shown in Figure 15.9. This study is consistent with the prior study [47] as to the material nature of the conducting paths in the Ta–O resistive-switching system.

Thus far, the Hf–O system displays some of the physical characteristics observed in Ta–O, in addition to the high endurance and relatively low-resistance device states. Scanning probe studies of polycrystalline HfO_2 , along with first-principles calculations, show an increased stability of oxygen vacancies at grain boundary sites and a correspondingly easier voltage-driven breakdown [49, 50]. Calka *et al.* performed a localized electroforming/switching study of

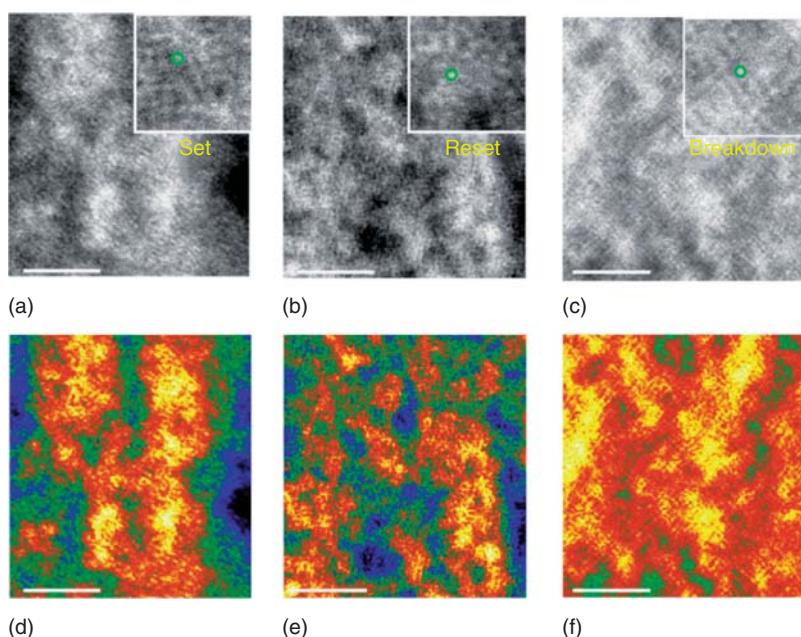


Figure 15.9 *In situ* STEM high-angle annular dark field (HAADF) investigation of Pt/SiO₂(1.5 nm)/Ta₂O₅(10 nm)/TaO_{2-x}(30 nm)/Pt resistance switching highlighting changes to the a-TaO_{2-x} film layer. Bright area: Ta-rich phase (conducting paths), dark area: nonconducting clusters. (d–f) Pseudo-color maps converted from the raw images

of (a–c), respectively. Yellow: Ta-rich phase (conducting paths), blue: nonconducting clusters. Bright and yellow regions (in a–f) are regarded as conductive percolation paths. Scale bar, 3 nm (a–f). (Reprinted by permission from Macmillan Publishers Ltd: Nature Communications [47], © 2013.)

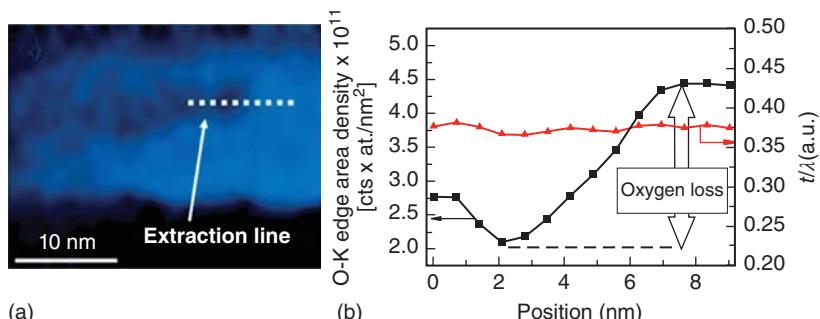


Figure 15.10 TEM and EELS study of HfO₂/TiN, electroformed by a CAFM tip and subsequently cross-sectioned. (a) EELS map derived from the oxygen K-edge highlighting a highly oxygen-deficient region. (b) Line profile of the oxygen signal (black)

across the deficient region and relative thickness (red) of the lamella across the same region. (Reproduced from [51] with permission. © IOP Publishing. All rights reserved.)

HfO_2 polycrystalline films grown on TiN using a diamond-coated Si tip at a fixed location and subsequently performed cross-sectional TEM-EELS analysis of this region [51]. By using scanning TEM (STEM) EELS of the O K-edge, a region exhibiting $\sim 55\%$ reduction in oxygen signal was found, as shown in Figure 15.10. HRTEM of this region revealed local amorphization of the oxygen-deficient channel, with a surrounding nanocrystalline HfO_2 . These direct material studies of the Ta–O and Hf–O systems, particularly in contrast to those of Ti–O, highlight some of the key differences in structural and electronic properties intrinsic to different binary oxide systems.

15.5

Conclusions

While research into the mechanisms of VCM-type resistance switching can be traced back all the way to the 1960s, the outstanding advances in recent decades in the instrumentation and related computational abilities have brought new perspectives to the field. Knowledge of the microscopic properties is a critical prerequisite for both complete understanding and future engineering of these devices. We have reviewed in this chapter some of the recent work taking advantage of modern and powerful ways to look into matter and shed additional light on this topic.

Acknowledgment

CL and RD would like to acknowledge the support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

1. Moos, R., Meneskou, W., and Härdtl, K.H. (1995) *Appl. Phys. A: Mater. Sci. Process.*, **61** (4), 389–395.
2. Son, J., Moetakef, P., Jalan, B., Bierwagen, O., Wright, N.J., Engel-Herbert, R., and Stemmer, S. (2010) *Nat. Mater.*, **9** (6), 482–484.
3. Strachan, J.P., Yang, J.J., Montoro, L.A., Ospina, C.A., Ramirez, A.J., Kilcoyne, A.L.D., Medeiros-Ribeiro, G., and Williams, R.S. (2013) *Beilstein J. Nanotechnol.*, **4**, 467–473.
4. Strachan, J.P., Pickett, M.D., Yang, J.J., Aloni, S., Kilcoyne, A.L.D., Medeiros-Ribeiro, G., and Williams, R.S. (2010) *Adv. Mater.*, **22** (32), 3573–3577.
5. Koehl, A., Wasmund, H., Herpers, A., Guttmann, P., Werner, S., Henzler, K., Du, H., Mayer, J., Waser, R., and Dittmann, R. (2013) *APL Mater.*, **1**, 042102.
6. Tanuma, S., Powell, C.J., and Penn, D.R. (2011) *Surf. Interface Anal.*, **43** (3), 689–713.
7. Fadley, C. (2005) *Nucl. Instrum. Methods Phys. Res., Sect. A*, **547** (1), 24–41.
8. Sowinska, M., Bertaude, T., Walczyk, D., Thiess, S., Schubert, M.A., Lukosius, M., Drube, W., Walczyk, C., and Schroeder, T. (2012) *Appl. Phys. Lett.*, **100**, 233509.
9. Bertaude, T., Sowinska, M., Walczyk, D., Thiess, S., Gloskovskii, A., Walczyk, C.,

- and Schroeder, T. (2012) *Appl. Phys. Lett.*, **101** (14), 143501/1–143501/5.
10. Wei, Z., Kanzawa, Y., Arita, K., Katoh, Y., Kawai, K., Muraoka, S., Mitani, S., Fujii, S., Katayama, K., Iijima, M., Mikawa, T., Ninomiya, T., Miyanaga, R., Kawashima, Y., Tsuji, K., Himeno, A., Okada, T., Azuma, R., Shimakawa, K., Sugaya, H., Takagi, T., Yasuhara, R., Horiba, H., Kumigashira, H., and Oshima, M. (2008) Highly Reliable TaO_x ReRAM and Direct Evidence of Redox Reaction Mechanism IEEE Technical Digest.
 11. Calka, P., Martinez, E., Lafond, D., Minoret, S., Tirano, S., Detlefs, B., Roy, J., Zegenhagen, J., and Guedj, C. (2011) *J. Appl. Phys.*, **109** (12), 124507.
 12. Strachan, J.P., Yang, J.J., Muenstermann, R., Scholl, A., Medeiros-Ribeiro, G., Stewart, D.R., and Williams, R.S. (2009) *Nanotechnology*, **20** (48), 485701.
 13. Dittmann, R., Muenstermann, R., Krug, I., Park, D., Menke, T., Mayer, J., Besmehn, A., Kronast, F., Schneider, C.M., and Waser, R. (2012) *Proc. IEEE*, **100** (6), 1979–1990.
 14. Wiemann, C., Patt, M., Cramm, S., Escher, M., Merkel, M., Gloskovskii, A., Thiess, S., Drube, W., and Schneider, C.M. (2012) *Appl. Phys. Lett.*, **100** (22), 223106/1–223106/3.
 15. Schneider, C.M., Wiemann, C., Patt, M., Feyer, V., Plucinski, L., Krug, I.P., Escher, M., Weber, N., Merkel, M., Renault, O., and Barrett, N. (2012) *J. Electron. Spectrosc. Relat. Phenom.*, **185** (10), 330–339.
 16. Fujii, T., Arita, M., Hamada, K., Kondo, H., Kaji, H., Takahashi, Y., Moniwa, M., Fujiwara, I., Yamaguchi, T., Aoki, M., Maeno, Y., Kobayashi, T., and Yoshimaru, M. (2011) *J. Appl. Phys.*, **109** (5), 53702/1–53702/5.
 17. Fujii, T., Arita, M., Hamada, K., Takahashi, Y., and Sakaguchi, N. (2013) *J. Appl. Phys.*, **113** (8), 83701/1–83701/7.
 18. Yang, Y., Gao, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012) *Nat. Commun.*, **3**, 732.
 19. Psiuk, B., Szade, J., Pilch, M., and Szot, K. (2009) XPS studies of perovskites surface instability caused by Ar+ ion and electron bombardment and metal deposition. 8th International Conference on Ion Implantation and Other Applications of Ions and Electrons, Kazimierz Dolny, Poland. *Vacuum*, **83**, S69–S72.
 20. Koehl, A., Kajewski, D., Kubacki, J., Lenser, C., Dittmann, R., Meuffels, P., Szot, K., Waser, R., and Szade, J. (2013) *Phys. Chem. Chem. Phys.*, **15**, 8311–8317.
 21. Stille, S., Lenser, C., Dittmann, R., Koehl, A., Krug, I., Muenstermann, R., Perlich, J., Schneider, C.M., Klemradt, U., and Waser, R. (2012) *Appl. Phys. Lett.*, **100** (22), 223503/1–223503/4.
 22. Lee, H.-S., Mizoguchi, T., Mistui, J., Yamamoto, T., Kang, S.-J.L., and Ikuhara, Y. (2011) *Phys. Rev. B: Condens. Matter*, **83** (10), 104110/10.
 23. Govoreanu, B., Kar, G.S., Chen, Y., Paraschiv, V., Kubicek, S., Fantini, A., Radu, I.P., Goux, L., Clima, S., Degraeve, R., Jossart, N., Richard, O., Vandeweyer, T., Seo, K., Hendrickx, P., Pourtois, G., Bender, H., Altimime, L., Wouters, D.J., Kittl, J.A., and Jurczak, M. (2011) 10x10nm² Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. 2011 IEEE International Electron Devices Meeting – IEDM '11, IEDM Technical Digest.
 24. Sawa, A. (2008) *Mater. Today*, **11** (6), 28–36.
 25. Hasan, M., Dong, R., Choe, H.J., Lee, D.S., Seong, D.J., Pyun, M.B., and Hwang, H. (2008) *Appl. Phys. Lett.*, **92**, 202102.
 26. Sawa, A., Fujii, T., Kawasaki, M., and Tokura, Y. (2004) *Appl. Phys. Lett.*, **85** (18), 4073–4075.
 27. Kawano, H., Shono, K., Yokota, T., and Gomi, M. (2008) *Appl. Phys. Express*, **1** (10), 101901/1.
 28. Yasuhara, R., Yamamoto, T., Ohkuba, I., Kumigashira, H., and Oshima, M. (2010) *Appl. Phys. Lett.*, **97**, 132111.
 29. Borgatti, F., Park, C., Herpers, A., Offi, F., Egoavil, R., Yamashita, Y., Yang, A., Kobata, M., Kobayashi, K., Verbeeck, J., Panaccione, G., and Dittmann, R. (2013) *Nanoscale*, **5** (9), 3954–3960.
 30. Meyer, R., Schloss, L., Brewer, J., Lambertson, R., Kinney, W., Sanchez,

- J., and Rinerson, D. (2008) Oxide Dual-Layer Memory Element for Scalable Non-Volatile Cross-Point Memory Technology Proceedings of the 9th Annual Non-Volatile Memory Technology Symposium, pp. 54–58.
31. Asanuma, S., Akoh, H., Yamada, H., and Sawa, A. (2009) *Phys. Rev. B*, **80** (23), 235113/1–235113/8.
32. Herpers, A., Lenser, C., Park, C., Offi, F., Borgatti, F., Panaccione, G., Menzel, S., Waser, R., and Dittmann, R. (2014) *Adv. Mater.*, **26**, 2730–2735.
33. Janousch, M., Meijer, G.I., Staub, U., Delley, B., Karg, S.E., and Andreasson, B.P. (2007) *Adv. Mater.*, **19**, 2232–2235.
34. Lenser, C., Kalinko, A., Kuzmin, A., Berzins, D., Purans, J., Szot, K., Waser, R., and Dittmann, R. (2011) *Phys. Chem. Chem. Phys.*, **13** (46), 20779–20786.
35. Andreasson, B.P., Janousch, M., Staub, U., Todorova, T., Delley, B., Meijer, G.I., and Pomjakushina, E. (2009) *Phys. Rev. B*, **80** (21), 212103/1–212103/4.
36. Meijer, G.I., Staub, U., Janousch, M., Johnson, S.L., Delley, B., and Neisius, T. (2005) *Phys. Rev. B*, **72**, 155102.
37. Lenser, C., Kuzmin, A., Purans, J., Kalinko, A., Waser, R., and Dittmann, R. (2012) *J. Appl. Phys.*, **111** (7), 76101.
38. Lenser, C., Connell, Z., Kovacs, A., Dunin-Borkowski, R., Koehl, A., Waser, R., and Dittmann, R. (2013) *Appl. Phys. Lett.*, **102**, 183504.
39. Jeong, D.S., Schroeder, H., and Waser, R. (2007) *Electrochem. Solid-State Lett.*, **10** (8), G51–G53.
40. Kwon, D.-H., Kim, K.M., Jang, J.H., Jeon, J.M., Lee, M.H., Kim, G.H., Li, X.-S., Park, G.-S., Lee, B., Han, S., Kim, M., and Hwang, C.S. (2010) *Nat. Nanotechnol.*, **5** (2), 148–153.
41. Strachan, J.P., Strukov, D.B., Borghetti, J., Yang, J.J., Medeiros-Ribeiro, G., and Williams, R.S. (2011) *Nanotechnology*, **22**, 254015.
42. Yang, J.J., Miao, F., Pickett, M.D., Ohlberg, D.A.A., Stewart, D.R., Lau, C.N., and Williams, R.S. (2009) *Nanotechnology*, **20** (21), 215201.
43. Yang, J.J., Zhang, M., Strachan, J.P., Miao, F., Pickett, M.D., Kelley, R.D., Medeiros-Ribeiro, G., and Williams, R.S. (2010) *Appl. Phys. Lett.*, **97** (23), 232102/1–232102/3.
44. Chen, Y., Goux, L., Clima, S., Govoreanu, B., Degraeve, R., Kar, G., Fantini, A., Groeseneken, G., Wouters, D., and Jurczak, M. (2013) *IEEE Trans. Electron Devices*, **60** (3), 1114–1121.
45. Chen, Y.S., Lee, H.Y., Chen, P.S., Liu, W.H., Wang, S.M., Gu, P.Y., Hsu, Y.Y., Tsai, C.H., Chen, W.S., Chen, F., Tsai, M.J., and Lien, C. (2011) *IEEE Electron Device Lett.*, **32** (11), 1585–1587.
46. Miao, F., Strachan, J.P., Yang, J.J., Zhang, M.-X., Goldfarb, I., Torrezan, A.C., Eschbach, P., Kelly, R.D., Medeiros-Ribeiro, G., and Williams, R.S. (2011) *Adv. Mater.*, **23**, 5633.
47. Park, G.-S., Kim, Y.-B., Park, S.Y., Li, X.S., Heo, S., Lee, M.J., Chang, M., Kwon, J.H., Kim, M., Chung, U.-I., Dittmann, R., Waser, R., and Kim, K. (2013) *Nat. Commun.*, **4**, 2382/1–2382/9.
48. Strachan, J.P., Medeiros-Ribeiro, G., Yang, J.J., Zhang, M.-X., Miao, F., Goldfarb, I., Holt, M., Rose, V., and Williams, R.S. (2011) *Appl. Phys. Lett.*, **98** (24), 242114.
49. Bersuker, G., Gilmer, D.C., Veksler, D., Kirsch, P., Vandelli, L., Padovani, A., Larcher, L., McKenna, K., Shluger, A., Iglesias, V., Porti, M., and Nafria, M. (2011) *J. Appl. Phys.*, **110** (12), 124518.
50. Iglesias, V., Porti, M., Nafria, M., Aymerich, X., Dudek, P., Schroeder, T., and Bersuker, G. (2010) *Appl. Phys. Lett.*, **97** (26), 262906/1–262906/3.
51. Calka, P., Martinez, E., Delaye, V., Lafond, D., Audoit, G., Mariolle, D., Chevalier, N., Grampeix, H., Cagli, C., Joussemaume, V., and Guedj, C. (2013) *Nanotechnology*, **24** (8), 85706/1–85706/9.

16

Interface-Type Switching

Akihito Sawa and Rene Meyer

16.1

Introduction

A homogeneous *interface-type* switching (also called: *area-scaling-type* switching) is often observed in resistive-switching memory (RS) cells consisting of oxides that are semiconducting or conducting in the pristine state. This type of switching takes place at the interface between the metal electrode and the oxide. By using multilead resistance measurements on perovskite oxide cells, Baikalov *et al.* [1] revealed that the contact resistance between the metal electrode and the perovskite oxide changes upon the application of an electric field. A number of models have been proposed for the mechanisms involved in the interface-type RS, such as electromigration (drift diffusion) of ionic point defects such as oxygen vacancies [1–7], trapping of charge carriers (holes or electrons) [8, 9], and interfacial Mott transition [10–13]. As explained in Chapter 1, only those mechanisms that involve a configuration change of atoms or ions will be able to support both reasonably low resistances (kilohms) and long retention times (years) at the same time. However, in the context of understanding the microscopic origin of the RS, both ionic and electronic mechanisms will be discussed.

In this chapter, we mainly focus on two model materials, which are most studied in the context of interface-type RS: mixed-valence manganites such as $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ or $(\text{Pr},\text{Ca})\text{MnO}_3$ (short: PCMO), which shows p-conductivity, and donor-doped SrTiO_3 , typically Nb-doped SrTiO_3 (short: Nb:STO), which is an n-conductor.

A significant characteristic of the interface-type switching is the scaling of the device resistance with the area [14, 15]. As an example, Figure 16.1 shows the resistance values in the low-resistance state (LRS) and high-resistance state (HRS) for cells based on Nb-doped SrTiO_3 and NiO, which are typical compounds of the interface-type and filament-type switching, respectively [14]. Both resistance states of the Nb-doped SrTiO_3 cells have resistance values that are inversely proportional to the cell area, whereas those of the NiO cells are much less dependent on the cell area. This result indicates that the RS in the Nb-doped SrTiO_3 cells

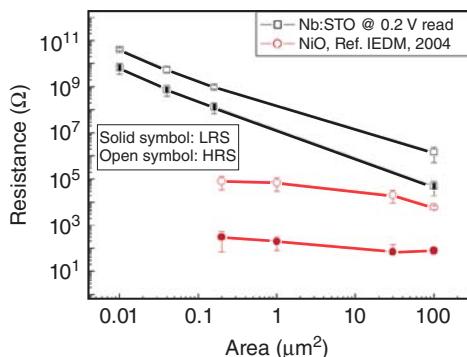


Figure 16.1 Area dependence of resistance values in high- and low-resistance states for Nb-doped SrTiO_3 (Nb:STO) and NiO memory cells. The resistance values of Nb:STO memory cells are linearly dependent on the area, suggesting that the resistive switching

takes place over the entire area of the interface, whereas those of NiO memory cells are almost independent of the area, suggesting that resistive switching is a local phenomenon. (Reprinted with permission from Ref. [14]. © 2005 IEEE.)

takes place over the whole area of the cell, that is, the entire interface, while the NiO cells show a filamentary RS. It should be noted that this view represents an idealized situation. For instance, multifilamentary scenarios in which the RS takes place in a regular manner with a periodicity on the nanometer scale are also considered here, because it shows a clear area scaling on the micrometer to millimeter scale. Furthermore, in many cases, no clear area scaling of one or both RS states is reported for the material systems discussed here. In fact, there are many indications that, often, more than one RS mechanism are active, which superimpose each other and, possibly, interact in an unforeseeable manner.

The first report on RS of metal/(Pr,Ca) MnO_3 cells dates back to 1997 [16], followed by many, for example, [8, 17–23]. A concept that deliberately introduces an additional tunnel barrier between the (noble) metal electrode and the conducting metal oxide turned out to be particularly interesting [24, 25].

It should be noted that there are reports on other material systems showing interface-type nanoionic RS based on tunneling barriers [26]. These systems are subject of ongoing research at the time of the publication of this book.

The chapter is organized as follows: in Section 16.2, we briefly provide some fundamentals of the oxides involved and of the current–voltage (I - V) characteristics of metal/oxide interfaces, and in Sections 16.3 and 16.4, we describe the RS of the two model systems including aspects of the mechanisms and modeling of interface-type switching. Section 16.5 focuses on systems in which a tunnel barrier is introduced between the metal electrode and the conducting oxide. It also includes aspects of the device implementation in cross-point memories, which can realize large-capacity nonvolatile memories. Section 16.6 reviews the special case of *ferroelectric* RS, an interface-type RS based on polarization reversal in ferroelectrics.

16.2

Metal/Conducting Oxide Interfaces: Characteristics and Fundamentals

16.2.1

Schottky-Like Metal/Conducting Oxide Interfaces

Before RS effects can be understood, it is necessary to have a basic understanding of the electronic properties of the electronically conducting oxides involved and of the *I-V* characteristics of metal electrode/conducting oxide interfaces.

Let us start with the *I-V* characteristics. There are two essential effects that control the contact resistance between the metal electrode and the oxide. One effect is a chemical reaction between the metal and the oxide. The other effect is the formation of space-charge layers due to differences in the Fermi level of the metal and the oxide. For an interface between a metal and a (semi)conductor, typically, a Schottky barrier gives rise to a high contact resistance due to the formation of a space-charge layer in which the majority carriers are strongly depleted (see Chapters 4 and 9). In fact, the contact resistance and the switching characteristics of conducting oxide memory cells generally depend on the electrode materials [8, 27–29]. Figure 16.2 shows the *I-V* characteristics for M/Pr_{0.7}Ca_{0.3}MnO₃/SrRuO₃ (M/PCMO/SRO) and M/SrTi_{0.99}Nb_{0.01}O₃/Ag (M/Nb:STO/Ag) cells, where M is the top electrode of Ti, Au, or SRO, with work functions of ~4.3, ~5.1, and ~5.3 eV, respectively [27]. In the cells, the SRO and Ag bottom electrodes form ohmic contacts with PCMO and Nb:STO, respectively. PCMO and Nb:STO are p- and n-type semiconducting oxides, respectively. For the p-type PCMO

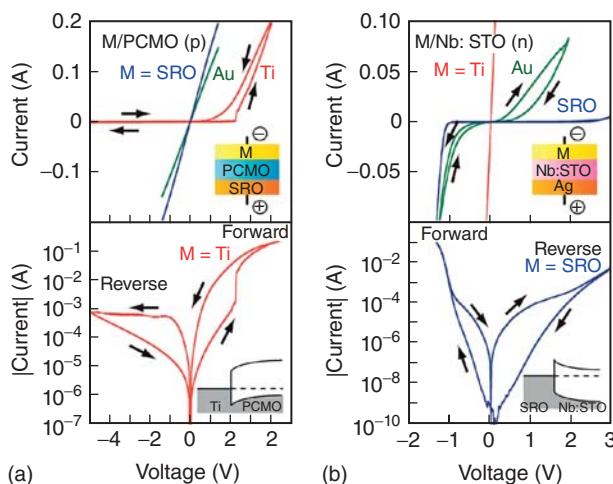


Figure 16.2 *I-V* curves for (a) p-type M/Pr_{0.7}Ca_{0.3}MnO₃/SrRuO₃ (M/PCMO/SRO) and (b) n-type M/SrTi_{0.99}Nb_{0.01}O₃/Ag (M/Nb:STO/Ag) cells (M = Ti, Au, and SRO). Current rectification with hysteresis is seen

in the *I-V* curves for the Ti/PCMO/SRO cell, and that with opposite polarity is seen in the Au/Nb:STO/Ag and SRO/Nb:STO/Ag cells. (Reprinted with permission from Ref. [27]. © 2008 Elsevier.)

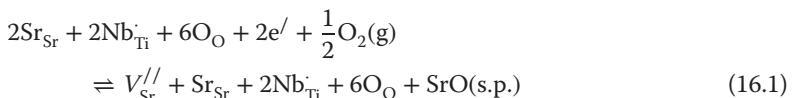
cells, the contact resistance of M/PCMO increases with decreasing work function of M, and the cell with a low work function metal of Ti shows rectifying *I*-*V* characteristics with hysteretic behavior indicative of RS, which is discussed later. On the other hand, for the n-type Nb:STO cells, as the work function of M increases, the contact resistance of M/Nb:STO increases, and the cell with a high work function metal of Au and SRO shows rectifying *I*-*V* characteristics (again with hysteretic behavior). The rectification of current is due to a Schottky-like barrier at the interface, as shown in the insets of the lower panels in Figure 16.2. Given that an RS is observed only in the cells showing rectifying *I*-*V* characteristics, the Schottky-like barrier plays a key role in the emergence of RS.

In the case of chemically very reactive Ti metal electrodes, the chemical reaction effect may also play a dominate role. Upon deposition (and annealing), Ti reacts with the oxide by extracting oxygen ions, for example, injecting oxygen vacancies from a region near the interface. In the case of n-type Nb:STO, these donor-type oxygen vacancies will increase the original donor concentration and, thus, only further increase the conductance of the ohmic contact. In the case of p-type PCMO, however, the oxygen vacancies will reduce the p-conductivity and may lead to a more insulating PCMO near the interface. This effect follows the same direction as the space-charge effect. Ti metal is turned into a low-conducting TiO_x contributing to the resistance of the interface. As we shall see in Section 16.4, the TiO_x may have a significant impact on the RS behavior.

16.2.2

Electronic Properties of Donor-Doped SrTiO_3

As described in detail in Chapter 4, SrTiO_3 is a band insulator with a band gap of approximately 3.2 eV, which can be turned into an n-type semiconductor, typically showing the characteristics of a degenerate semiconductor, by substituting host cations by higher valent cations. Typically donor dopants are pentavalent Nb on sites of the tetravalent Ti or trivalent La on sites of the bivalent Sr. Under normal synthesis conditions of single crystals, the positive extra charge of the donor is compensated by electrons, giving rise to the n-conductivity. If such a material is annealed in highly oxidizing atmospheres at high temperatures (>1400 K), cation vacancies are introduced from the surface (or inner surfaces such as grain boundaries). Cation vacancies (typically Sr vacancies, $V_{\text{Sr}}^{/\!/}$) are negatively charged and take over the compensation of the donors, leading to a highly insulating SrTiO_3 . The equilibrium between electron compensation (left) and Sr vacancy compensation (right) is given by



using the Kröger–Vink notation. The excorporated Sr reacts with oxygen from the gas phase (g) and forms a secondary phase (s.p.). A comprehensive overview of the defect chemistry of donor-doped SrTiO_3 has been provided by Moos and Härdtl

[30]. Cation vacancies in SrTiO_3 show extremely low diffusivities, even at high temperatures, so that it is impossible to fully oxidize thick layers of single crystalline, n-conducting donor-doped SrTiO_3 . However, close to the surface, such a reoxidation is possible even at moderately high temperature because the surface reaction is space-charge-controlled [31, 32]. Also, mechanical polishing and sputtering under oxidizing conditions may introduce surfaces states, which trap electrons and lead to thin insulating layers.

The oxygen-vacancy concentration in donor-doped SrTiO_3 is extremely low, unless a high-temperature anneal under extremely oxygen partial pressures or extremely reducing electrochemical conditions is applied. The negligible oxygen vacancy concentration in (at least the bulk of) donor-doped SrTiO_3 in both the electron compensated and the cation-vacancy compensated region has to be taken into account when models of RS are developed.

16.2.3

Electronic Properties of Mixed-Valent Manganites

Mixed-valent manganites represent a class of oxides that show a large diversity in their electronic (and magnetic) properties [33–37]. Their crystal structure is given by a distorted perovskite structure ABO_3 , where Mn is positioned on B-site, and the A-site is shared by trivalent rare earth cations (R) such as La, Pr, or Nd and divalent alkaline earth cations (A) such as Ca, Sr, or Ba. The composition $(\text{R}_{1-x}\text{A}_x)\text{MnO}_3$ can be regarded as a solid solution of RMnO_3 and AMnO_3 , in which the ratio of R to A determines the ratio of Mn^{3+} and Mn^{4+} . In the perovskite lattice, the oxygen ions are arranged in an octahedral manner about the Mn cations. As a result, a crystal field splitting occurs, that is, the five 3d orbitals of the Mn get split into three t_{2g} levels at lower energy and two e_g levels at higher energy. In the end-member AMnO_3 , Mn^{4+} ($[\text{Ar}]3s^23d^3$) has no occupied e_g -orbital and only three strongly bound t_{2g} electrons overlap with the oxygen orbitals. There is no Jahn–Teller distortion, and since the gap between the t_{2g} and the e_g levels is rather large, there are no unbound carriers; hence, the compound is insulating. In the other end-member RMnO_3 , Mn^{3+} ($[\text{Ar}]3s^23d^4$) has one loosely bound e_g electron per unit cell. As a consequence, a Jahn–Teller distortion occurs, and the two e_g levels are split into a lower and an upper e_g band. In addition, due to the electron correlation, namely the on-site Coulomb repulsion, each band is further split into two bands: the upper and lower Hubbard bands. As a result, the lowest e_g band is totally occupied, while other bands are empty, leading to an insulating system [33, 38, 39]. So, the end-member RMnO_3 and AMnO_3 are reasonably good insulators, while the solid solutions show p-conductivity. Figure 16.3 illustrates the density of states as a function of the energy for typically mixed-valent manganite, here: $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. In contrast to a band insulator with a dedicated band gap, there is no clear band gap, that is, there is a significant density of states at the Fermi level and its vicinity. Therefore, mixed-valent manganites can be regarded in first approximation as metals with low-mobility carriers. The carriers can be described as small polarons, because

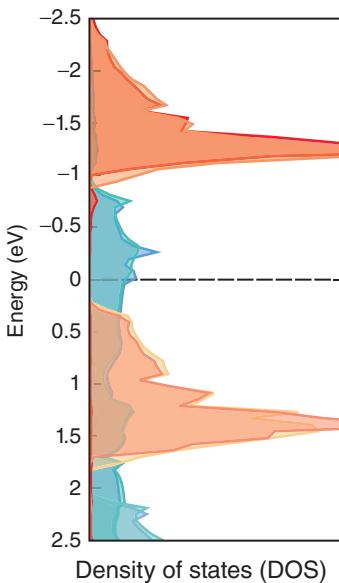


Figure 16.3 Sketch of the calculated density of states for a RAMnO_3 system (here: $\text{La}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$), which shows the finite density of (relatively localized) states near the Fermi level (at 0 eV). This reflects the fact that this compound rather is a metal than a band-type semiconductor. Reproduced with modifications from Ref. [41].

the holes are somewhat localized in the Mn ions, represented as Mn^{4+} , and move by a so-called double exchange mechanism. The polaronic nature of the conductivity in mixed-valent manganites has been the subject of many studies (see, e.g., [22, 40], and references cited therein). The details do not only depend on the R/A ratio but also on the specific nature of the R and A cations.

The defect chemistry describing the lattice disorder of $(\text{R}_{1-x}\text{A}_x)\text{MnO}_3$ is controlled by the perovskite crystal lattice, which means that interstitial oxygen ions have to be neglected and that the Schottky disorder is favored over the Frenkel disorder. If one starts to substitute R^{3+} by A^{2+} in RMnO_3 , the A ions act as acceptors. While in titanates, acceptors are compensated by oxygen vacancies, the dominant compensation in manganites is by holes, as this is energetically more favorable. As an example for a solid solution with 10 at% A, Figure 16.4 illustrates the calculated defect concentrations as a function of the oxygen partial pressure. In the oxidizing and moderately reducing regime, the majority defects are A acceptors compensated by holes. Toward further reducing conditions, oxygen vacancies take over in the compensation of the acceptors and the p-conductivity drops. This aspect will be of importance for the discussion of RS. Upon further reduction, the situation starts to get dominated by an increasing oxygen vacancy concentration, compensated by electrons. Under these extremely reducing conditions, the situation becomes similar as in titanates. However, when reading these diagrams (Figure 16.4), care must be taken with respect to the stability of the phases, in particular, at very low oxygen partial pressures. The stability range of $\text{R}_{1-x}\text{A}_x\text{MnO}_{3-\delta}$ is hardly reported in the literature. Furthermore, not only the concentration of the electronic carriers but also their mobility may be affected by the oxygen-vacancy concentration because of the role of oxygen ions in the double exchange mechanism [42].

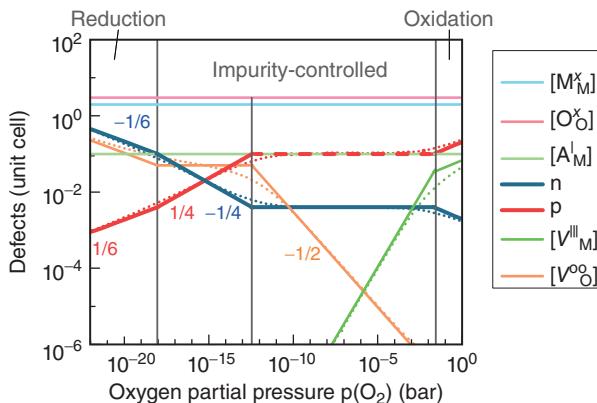


Figure 16.4 (Ausschnitt wählen, auf RAMnOx umschreiben, electron. Majoritätsträger herausstellen) Calculated defect concentrations as a function of the oxygen partial pressure $p(\text{O}_2)$ for $\text{R}_{0.9}\text{A}_{0.1}\text{MnO}_{3-\delta}$. (From Ref. [34] with modifications.)

16.3

Resistive Switching of Metal/Donor-Doped SrTiO_3 Cells

Besides the Schottky diode characteristics, Figure 16.2b shows RS effects for $\text{M}/\text{SrTi}_{0.99}\text{Nb}_{0.01}\text{O}_3/\text{Ag}$ ($\text{M}/\text{Nb:STO}/\text{Ag}$) with $\text{M}=\text{Au}$ and SrRuO_3 . The RS polarity of such cells typically displays a eightwise (8W) characteristics with respect to the electrode M. A variation of the donor concentration in the SrTiO_3 crystal leads to a systematic change in the reverse current and in the extent of the RS window as shown in Figure 16.5 for $\text{SRO}/\text{SrTi}_{1-x}\text{Nb}_x\text{O}_3/\text{Ag}$ [$\text{SRO}/\text{Nb:STO}(x)/\text{Ag}$] cells with various Nb content x [29]. Since Nb^{5+} acts as a donor, an increase in x leads to a reduction in the valence of Ti ions and a resultant increase in electron concentration of the n-type semiconductor $\text{Nb:STO}(x)$. All the cells show rectifying $I-V$ characteristics, and the rectification is enhanced with decreasing x , suggesting an increase in the depletion layer width. In addition to the enhanced rectification, the hysteretic behavior of the $I-V$ characteristics, that is, the RS, is enhanced with decreasing x . This result shows that the depletion layer width of the Schottky-like barrier is a crucial parameter governing the RS characteristics of Nb:STO cells. A similar result is obtained for a fixed Nb content, when the conditions of an oxidizing anneal prior to Pt electrode deposition are varied [43]. Increasing the O_2 anneal temperatures from 900 to 1200 °C leads to significant reduced currents and larger RS windows because the concentration of electrons is reduced in a near surface layer according to a shift in the defect toward the right in Eq. (16.1). This leads to an increase in the depletion layer width underneath the Pt electrode.

Apart from the common 8W polarity, which is typically observed for metal/n-doped SrTiO_3 cells, the literature shows a wide variety of seemingly contradicting results and explanations – from area-dependent to localized (filamentary) RS

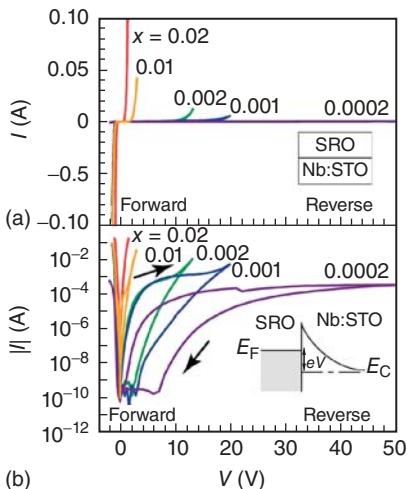


Figure 16.5 I - V characteristics of the SRO/ $\text{SrTi}_{1-x}\text{Nb}_x\text{O}_3/\text{Ag}$ [SRO/Nb:STO(x)/Ag] cells with various Nb content x [29]. (a) and (b) are plotted on linear and logarithmic current scales, respectively. Large hysteresis is seen in the reverse-bias region except for the $x = 0.02$ cell.

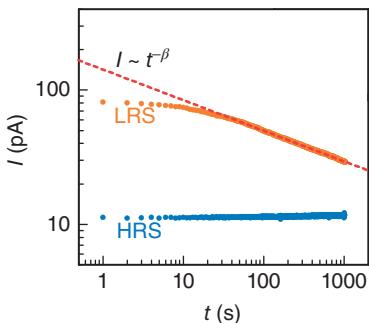


Figure 16.6 Resistance-state retention characteristics. Time-dependence of the small-signal ($+0.1$ V) current after switching to HRS and LRS, shown for a sample with a high-quality interface. The red dashed line is a fit to a Curie–von Schweidler power law for the decay of LRS. (From Ref. [46].)

and from nanoionic redox models to purely electronic models. The reason seems to be that at least two different RS mechanisms may be triggered in these cells [44]: an area-dependent RS before a forming process is invoked, and a filamentary RS after forming.

When an area-dependent RS prevails, the I - V and C - V characteristics show a hysteretic behavior and can be explained by a Schottky depletion model [45] based on an electronic trapping/detrapping mechanism. The Schottky barrier heights (SBHs) are not only determined by the work function of the metals and the electron affinity of the n-SrTiO₃ but also by interface states and, in particular, by additional thin low-permittivity interface layers, which are caused by inappropriate processing [46]. Often, the area-dependent RS displays relatively short retention times (Figure 16.6) and signatures of an electronic trapping/detrapping of band-gap states in the depletion region [44, 46, 47]. Because of the very small read-out currents, retention times in the 10^2 – 10^3 s regime do not violate the retention time–current relation of purely electronic RS described in Chapter 1. On the other hand, a nanoionic mechanism also cannot be excluded.

The filamentary RS encountered after forming leaves the $C-V$ characteristics unaffected by the resistive state [44]. This clearly indicates the local nature of the RS event while the majority of the interface remains invariant. In several cases, a lateral inhomogeneous SBH is discussed where reversible local SBH reductions lead to filamentary-type switching [48–50]. In some cases, the $I-V$ data of the filamentary RS do not fit to a Schottky depletion layer analysis and indicate a local change in the composition, which can be explained by nanoionic redox processes. This is consistent with large retention times that are observed [44].

In principle, scanning probe microscopy (SPM) techniques applied to surfaces of conducting oxides are not able to distinguish between filamentary and area-scaling RS because switching always occurs at the tip (which is a point contact at first approximation). A different story is the investigation of metal–insulator–metal (MIM) structures by SPM after peeling off the top electrode [7]. Still, SPM on surfaces may be useful to study the polarity of the RS and correlations with surface features. Typically, positive tip voltages are used to increase the conduction, consistent with an 8W RS polarity. A correlation of the SrO and TiO₂ termination with the $I-V$ characteristics is reported for local-conductivity atomic force microscopy (LC-AFM) induced switching [51].

An essential question on this class of RS systems is related to the 8W RS polarity. The counter-eightwise (C8W) RS often encountered in filamentary switching oxides, which become n-conductive if the O stoichiometry is lowered, is explained by the fact that mobile donors such as oxygen vacancies or cation interstitials are positively charged defects, which are attracted by a positively polarized electrode. As a consequence, the donor concentration is increased just in front of the electrode, which facilitates the electron conduction across this interface. Although in the bulk of n-doped SrTiO₃ crystals, no significant concentration of *mobile* donors (here: oxygen vacancies) is conceivable according to Section 16.2.2 and Chapter 4, these defects may be present in the space charge at the surface [32], and they may be electrochemically introduced at the electrode by anodic oxidation of oxygen ions. In fact, in Pt/Nb:SrTiO₃ cells without forming for which an area-dependent RS was found, a high concentration of oxygen vacancies has been detected at the electrode interface (in a depth of 3–10 nm) in the HRS using the STEM (Scanning transmission electron microscopy) EELS (electron energy loss spectroscopy) technique [52, 53]. The oxygen-vacancy concentration was reduced for the LRS. As a qualitative explanation of this apparent contradicting situation, a detachment of the region of high-oxygen-vacancy concentration from the n-conduction further afar from the interface has been suggested for a similar situation [7], supported by the drift–diffusion model [53, 54].

16.4

Resistive Switching of p-Type PCMO Cells

For the cells consisting of the strongly correlated electron system PCMO, the switching mechanism is usually different from that of the Nb:STO cells. In this

section, we describe the cells that are built by depositing an electrode metal directly onto PCMO (independent of whether or not a redox reaction occurs at the interfaces, which leads to a metal oxide layer formation), while cells with a deliberate tunnel oxide between the PCMO and the metal are described in Section 16.5.

When a positive voltage is applied to pristine $\text{Ti}/\text{Pr}_{0.5}\text{Ca}_{0.5}\text{MnO}_3/\text{SrRuO}_3$ (Ti/PCMO/SRO) cells, they undergo forming process into a HRS. Figure 16.7a shows cross-sectional transmission electron microscopy (TEM) images of such a cell after forming [6]. There is an amorphous TiO_y (a-TiO_y) layer with a quite uniform thickness of ~ 10 nm between the Ti electrode and the PCMO layer. This indicates that oxygen ions have drifted from the PCMO layer into the Ti electrode and part of the Ti electrode adjacent to the PCMO layer is oxidized. The drift diffusion of oxygen ions from the PCMO layer into the Ti electrode also forms oxygen vacancies in the PCMO layer. The formation of oxygen vacancies induces change in the valence of Mn ions in the PCMO layer. Figure 16.7b shows the Mn-L edge EELS spectra obtained at different positions in the PCMO layer, as indicated in Figure 16.7a. The peak intensity ratio of Mn-L₃ and Mn-L₂, $I(\text{L}_3)/I(\text{L}_2)$, decreased with increasing distance from the a-TiO_y/PCMO boundary, meaning that the Mn valence near the interface is smaller than that away from the interface.

When a negative voltage bias is applied to the Ti top electrode, a SET operation is conducted, that is, the resistance state is converted from HRS to LRS, as shown in Figure 16.2a, corresponding to a C8W polarity of these type of cells. In this case, positively charged oxygen vacancies are expected to drift from the PCMO layer into the a-TiO_y layer. During the RESET, oxygen vacancies are expected to drift back to the PCMO layer from the a-TiO_y layer.

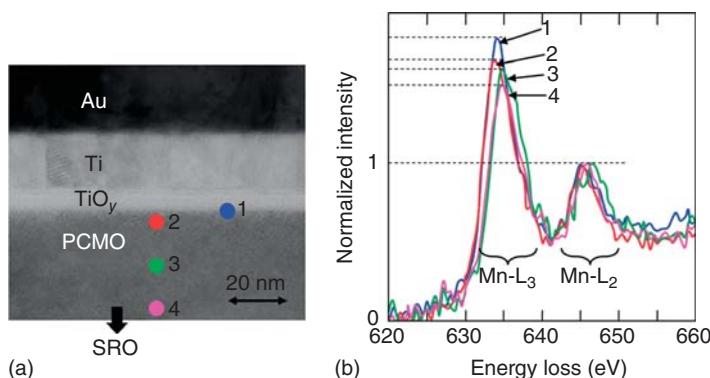


Figure 16.7 (a) Cross-sectional TEM images of a Ti/PCMO/SRO junction after forming (into a HRS). There is a thin amorphous TiO_y layer (~ 10 nm) between the Ti electrode and the PCMO layer [6]. (b) Mn-L edge electron

energy loss spectroscopy (EELS) spectra at several positions around the Ti/PCMO interface [6]. The positions are indicated in (a). Each spectrum is normalized by the intensity of the Mn-L₂ peak.

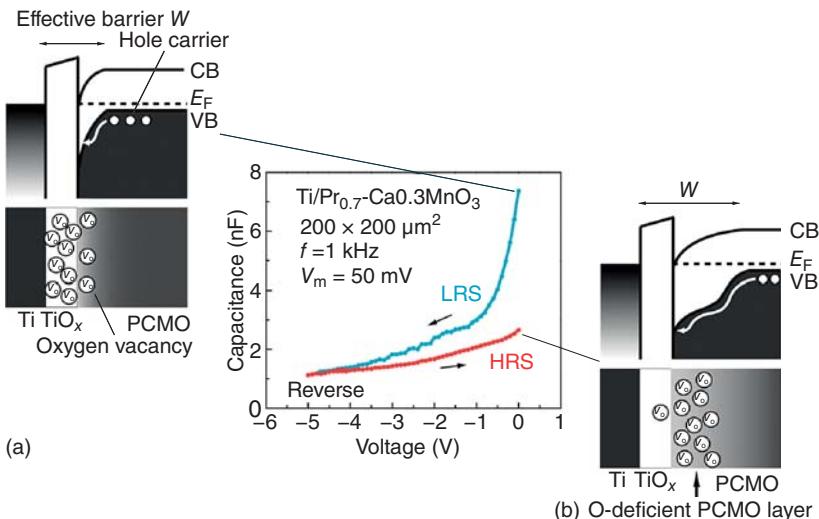


Figure 16.8 C - V curves under reverse bias for a $\text{Ti}/\text{PCMO}(x=0.3)/\text{SRO}$ device. The hysteretic characteristics indicate that the barrier width (W_d) at the Ti/PCMO interface is altered by applying an electric field. The insets show possible band diagrams of the Ti/PCMO interface (a) in a low-resistance state (LRS) and (b) in a high-resistance state (HRS). The work function of Ti is

~ 4.3 eV, that of *p*-type PCMO for $x < 0.5$ is $4.4\text{--}4.8$ eV, and the band gap is $0.2\text{--}0.6$ eV. Therefore, a Schottky-like barrier can be formed at the interface. In the HRS, the oxygen-deficient PCMO layer at the interface has a larger band gap, which interferes with the hole-carrier conduction. (Adapted with permission from Ref. [56]. © 2005 SPIE.)

There are several explanations for the nature of the resistive states in Ti/PCMO junction. For example, it has been reported that the oxygen-deficient PCMO film might have a larger band gap compared with that of the stoichiometric film [6, 55]. It is thus considered that in the HRS, the oxygen-deficient PCMO layer adjacent to the $a\text{-TiO}_y$ layer has a larger band gap compared with the bulk PCMO layer distant from the $a\text{-TiO}_y/\text{PCMO}$ boundary. This layer could act as an effective barrier for the hole-carrier conduction, as schematically drawn in Figure 16.8 [6]. In the LRS, when the oxygen vacancies in the PCMO are fewer, the hole carriers could flow through a thin depletion layer via a tunneling process as depicted in Figure 16.8. The change in the barrier at the interface may alter the hole-carrier conduction from tunneling to thermionic emission. This barrier width change is supported by capacitance measurements. The value of the capacitance per area is given by $C = \epsilon_{\text{TiO}}\epsilon_{\text{PCMO}}/(\epsilon_{\text{PCMO}}W_a + \epsilon_{\text{TiO}}W_d)$, where ϵ_{TiO} and ϵ_{PCMO} are the dielectric constants of $a\text{-TiO}_y$ and PCMO layers, respectively, W_a is the thickness of the $a\text{-TiO}_y$ layer, and W_d is the depletion layer width in the PCMO layer. The effective barrier width $W = W_a + W_d$. Assuming that ϵ_{TiO} , ϵ_{PCMO} , and W_a do not change, the change in C is attributable to the alteration of W_d . As seen in Figure 16.8, the C - V curve for the $\text{Ti}/\text{PCMO}/\text{SRO}$ cell shows hysteretic behavior, that is, capacitive switching, and the value of C in LRS is larger than that in HRS. This suggests

that W_d in LRS is narrower than that in HRS, consistent with the possible band diagram of the Ti/PCMO interface in HRS and LRS, as schematically shown in Figure 16.8.

Another possibility to explain the higher resistance of the HRS compared to the LRS is a reduced hole concentration as a result of a higher oxygen vacancy concentration as illustrated in Figure 16.4. The oxygen partial pressure (here: between approximately 10^{-10} and 10^{-15} bar) reflects the redox state of the PCMO. Toward increasing oxygen deficiency (as expected for the HRS), that is, with increasing oxygen-vacancy concentration in the PCMO layer adjacent to the $\text{a-TiO}_y/\text{PCMO}$ boundary, the hole concentration decreases, leading to a higher resistivity.

In a similar study of Ti/PCMO/SRO films epitaxially grown on STO substrates [23, 57], a more active role of the interface TiO_y layer was deduced, which led to a third possibility to explain the resistance change. Microscopy (TEM) and spectroscopy (EELS, hard X-ray photoelectron spectroscopy (HAXPES)) studies combined with electrical characterization have clearly proven that the RS in these cells is based on a redox process, which mainly occurs on the TiO_y side (Figure 16.9). The different resistance states are determined by the amount of fully oxidized Ti ions in the stack. The TiO_2 acts as a tunneling barrier, implying a reversible redox reaction at the interface that governs the increase or decrease in this insulating tunnel barrier.

In a number of cases, cells consisting of $(\text{R},\text{A})\text{MnO}_3/\text{n-doped SrTiO}_3$ have been investigated with respect to their RS properties. Here, the I - V feature of a typical

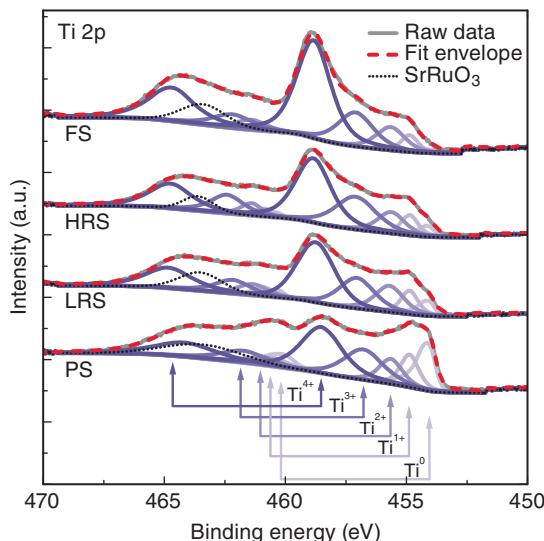


Figure 16.9 Ti 2p XPS spectra using HAXPES of Pt/Ti/PCMO/SRO collected for the pristine state (PS), the low-resistance state (LRS), the high-resistance state (HRS), and the formed state (FS), sorted from low resistance

(pristine) to high resistance (formed). The weights of the lines of the different Ti valence correspond to the stoichiometry of the TiO_x layer formed at the Ti/PCMO interface. (From Ref. [23].)

p–n junction is overlaying the electromigration of oxygen vacancies during the RS (see, e.g., [58, 59] and references cited therein).

For a Pt/(Pr,Ca)MnO₃/Pt system, the switching kinetics was studied comprehensively and two processes with very different time constants were suggested: a fast short-range exchange of oxygen ions at the electrode interface and a slow long-range redistribution of oxygen ions in the entire (Pr,Ca)MnO₃ film [60].

16.5

Resistive Switching in the Presence of a Tunnel Barrier

Another example of an oxide-based nonvolatile memory technology displaying area scaling is the tunnel ReRAM. The tunnel ReRAM was developed by Unity Semiconductor under the name of conductive metal oxide (brand name: CMOx). This work was later continued by Rambus Labs. Similarly to Magnetic RAM MRAM devices, the current is limited by quantum-mechanical tunneling through a thin insulating oxide layer. A possible advantage over other types of ReRAM devices involving oxygen ion motion seems to be that current levels for the SET, RESET, and Read operations can be controlled by the electronic tunnel properties of the barrier. The tunnel properties are supposed be changed by moving oxygen ions in and out of the tunnel barrier. The current levels of the device can be precisely controlled by varying the tunnel oxide thickness of the deposited tunnel oxide. In addition, highly nonlinear *I*-V characteristics in both the ON and the OFF state may allow for operating the memory device in a passive cross-point configuration (1R) without the need for an additional transistor or diode-like selector device. This is called a *self-select feature* (Chapter 22). Uniform conduction in conjunction with a well-defined geometry of the tunnel device results in excellent cell-to-cell and wafer-to-wafer uniformity and excellent device yield. Area scaling of the states allows predicting the device currents for future technology nodes. A further feature is the analog transition between resistance states, which enables multibit storage within a cell. Challenges to the device are insufficient data retention and some aspects of material integration and processing.

16.5.1

Device Structure and Materials

Figure 16.10 shows the device structure and a TEM cross section displaying the different layers of the tunnel ReRAM device. A thin tunnel oxide with a thickness of 2–3 nm is sandwiched between a noble metal top electrode and a conductive oxide electrode. The tunnel barrier is typically a high-k material such as ZrO₂ or HfO₂. The CMOx is a metallic-like perovskite such as (Pr,Ca)MnO₃, but can also be highly donor-doped SrTiO₃, for example, La- or Nb-doped SrTiO₃. Top and bottom electrodes are typically Pt or other noble metals. In the discussed device structure, the choice of a noble metal is mandatory to prevent a reaction with

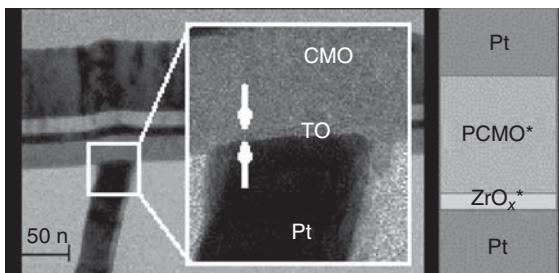


Figure 16.10 Structure and TEM cross section through a tunnel ReRAM device [61].

oxygen from the cell, which would create an additional undesired oxide layer in this type of devices. This is in contrast to the situation discussed in Section 16.3. Both CMOx and tunnel oxide are deposited at elevated temperature to ensure that the layers are crystalline. Crystalline film morphology structure of $(\text{Pr,Ca})\text{MnO}_3$ and most other conductive perovskites is required to ensure conductivity.

16.5.2

Electrical Characteristics

Figure 16.11 displays the quasistatic hysteretic I - V characteristics of the tunnel ReRAM based on PCMO. The I - V characteristic is highly nonlinear for both positive and negative polarities. When a positive voltage is applied to the top electrode, the device current decreases by $10\times$ to $100\times$. Applying negative voltages allows the reversal of the resistance change, that is, this device shows a typical C8W switching polarity with respect to the top electrode. In contrast to filamentary memory devices showing a threshold-like switching behavior, the transition between ON and OFF state is continuous. A continuous transition between these states is beneficial for multilevel storage. An external current limiter, commonly a transistor, is not required. Cycle-to-cycle variations of the I - V characteristic are small and are described by a continuous drift-like rather than stochastic changes as observed in filamentary-based memories devices.

Figure 16.12 shows the area scaling of the read currents (inset) and the read current densities measured at 0.5 V as a function of the device size. Electrode feature sizes range from $20 \mu\text{m}$ down to 30 nm. Data shown here are collected across a 300 mm wafer. The result clearly demonstrates the linear relationship between the current and the area, indicating that the conduction is uniformly distributed across the electrode and not localized. The observed current variations within each device size can be largely attributed to differences in the tunnel oxide thicknesses across the wafer.

The scaling of the device currents with the tunnel oxide thickness and the dependency on the CMO material is illustrated in Figure 16.13. An exponential relationship between the tunnel oxide thickness, here represented by the sputter deposition time, and the current densities is observed. The characteristic behavior as well as shape of the I - V curve and temperature dependence (not shown here) can be

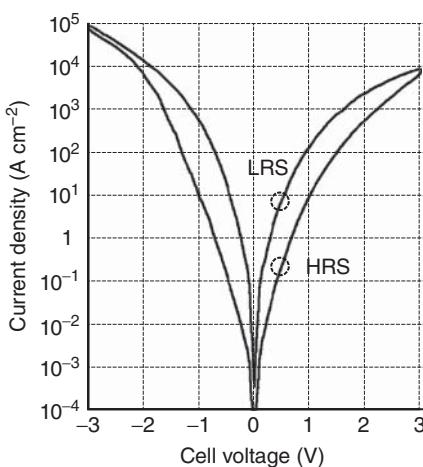


Figure 16.11 Quasistatic I - V characteristic of the tunnel ReRAM device [24].

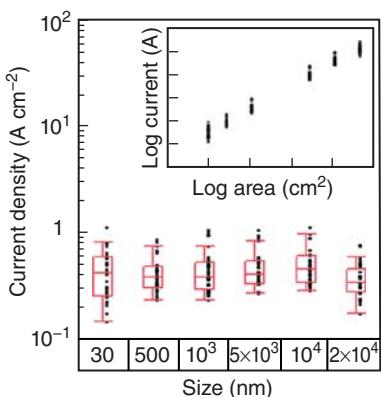


Figure 16.12 Read currents (inset) and read current densities at 0.5 V read voltage as a function of the device size. Data are collected from test structures across a 300 mm test wafer [61].

best described by a trap-assisted tunnel mechanism. Some CMO materials with higher conductivity as compared to $(\text{Pr},\text{Ca})\text{MnO}_3$ display current levels close to MIM devices. $(\text{Pr},\text{Ca})\text{MnO}_3$ devices have current values about 1 order of magnitude lower than MIM devices, suggesting the presence of a thin depletion layer at the $(\text{Pr},\text{Ca})\text{MnO}_3$ insulator interface. Shape of the I - V characteristics and functionality of the devices are largely independent of the CMO material.

The typical cycling endurance of the tunnel ReRAM devices is in the order of 10^3 – 10^6 cycles, depending on the material combination and the set/reset conditions [24]. For some material systems, a trade-off between memory effect and cycling endurance is observed. Data retention ranges from several days to a month at elevated temperatures and is a function of the material system and the set and reset conditions. A representative summary of the electrical characteristics of the tunnel ReRAM is given in Table 16.1.

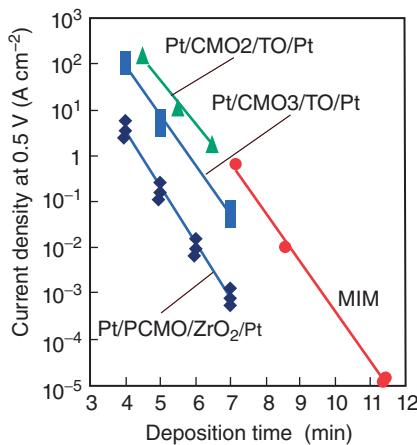


Figure 16.13 Dependence of the read current density at 0.5 V on tunnel oxide (TO) deposition time (tunnel oxide thickness) and material combinations. PCMO denotes $(\text{Pr},\text{Ca})\text{MnO}_3$ while CMO₂ and CMO₃ represent other conducting metal oxides [24].

Table 16.1 Device characteristics of tunnel ReRAM devices [61].

	TO = 23 Å	TO = 27 Å	
Program/erase voltage		<3 V	
Device size		30 nm	30 nm
Program current density	10^5 A cm^{-2}	1 μA	10^4 A cm^{-2}
Erase current density	10^5 A cm^{-2}	1 μA	10^4 A cm^{-2}
Forming current (A)		Not required	
Read current density at 1.0 V	10^3 A cm^{-2}	10 nA	10^2 A cm^{-2}
Cycling endurance		10^5	
Resistive memory effect		10×	
Program/erase time		<1 μs	
Data retention (type/best*)	24 h@110C/180 days@110C*		

16.5.3

Mechanism and Modeling

In the tunnel ReRAM presented here, the change in resistance can be explained by a field-driven exchange of oxygen between the $(\text{Pr},\text{Ca})\text{MnO}_3$ and the tunnel oxide and the resulting resistance change [24]. If a positive voltage is applied to the top electrode, oxygen vacancies are moved from the tunnel oxide into the $(\text{Pr},\text{Ca})\text{MnO}_3$, leaving a negative charge in the tunnel oxide. Due to the negative charge, the height of the tunnel barrier will increase and the electron tunneling current will decrease. The device is switched to the OFF state. After reversing the polarity, oxygen vacancies are attracted by the top electrode and are moved from the $(\text{Pr},\text{Ca})\text{MnO}_3$ into the tunnel barrier. The barrier height is reduced and the tunnel current increases. The device is switched into the ON state. The mechanism and the corresponding band diagrams including a charge-induced change in the tunnel barrier height are shown in Figure 16.14. Capacitance measurements

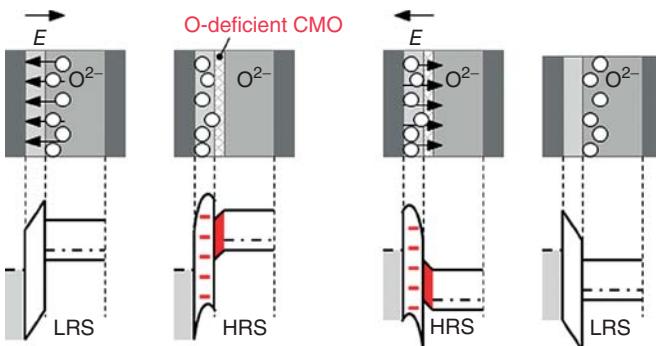


Figure 16.14 Schematic of the redistribution of oxygen between the CMO layer and the tunnel oxide for an external electric field. The accumulation of negative charge in the tunnel oxide causes an increase in

the potential barrier. A small change in the capacitance is explained by the formation of a thin insulating oxygen-deficient interface layer in the CMO. The effect can be reversed by changing the field polarity [61].

indicate that a thin insulating layer is formed at the PCMO/tunnel oxide interface in the OFF state, which is measured as a 3–5% decrease in the device capacitance. A model that explains the capacitance change includes the formation of a thin insulating layer of PCMO. The layer forms during the RESET operation, when oxygen is moved from the PCMO layer into the tunnel oxide and the PCMO is depleted in oxygen.

16.5.4

Passive Cross-Point Arrays

To be competitive in the ultrahigh-density memory market, a minimization of both memory array area and periphery area is essential. In order to achieve competitive chip sizes compared to the state-of-the-art planar and future 3D NAND Flash technologies, memory cells have to have a smaller footprint. The requirement of reducing the chip size becomes even more stringent if we take into consideration that the NAND Flash technology allows to store 2 bits/cell or 3 bits/cell (multilevel cells, MLCs).

Area reduction can be achieved by replacing the access transistor in the memory array (1T-1R configuration) by a two-terminal selector device (1S-1R configuration). Two-terminal selector devices do not require a third access line as needed by an access transistor. The selector device shifts the *I-V* characteristic by the threshold voltage of the selector device. Below a critical voltage, the resistance of the memory cell comprised of a selector device and a memory device is limited by the selector device. Above the critical voltage, the resistance of the current is limited by the resistive memory element. This allows to drastically reduce the leakage currents through partially selected memory cells. The ability to implement select devices in the layers on top of the CMOS Si wafer opens up the possibility of

building multiple memory layers on top of each other. Selector devices are extensively discussed in Chapter 22.

An alternative to reduce parasitic leakage currents is to utilize the non-linearity of the I - V characteristic of the memory device. Due to the highly nonlinear I - V characteristics both in the ON and OFF states, the tunnel ReRAM is the viable device for the selector-less (self-select) approach. Based on the intrinsic self-selecting behavior of the cell, arrays have been realized based on $\text{Al}/\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3/\text{Pt}$ cells [62, 63] and based on the CMOx concept with tunnel oxides [25]. In the latter case, arrays up to 8192×256 have been implemented. Asymmetric bit-line and word-line sides are due to differences in the limitations. While the length of the word line determines the maximum tolerable leakage current during SET and RESET operations, the bit-line length is determined by the maximum tolerable deterioration of the read signal due to parasitic currents from partially selected cells. Figure 16.15 illustrates perfect standard patterns of all zeros, all ones, diagonal, inverse diagonal, checker board, and inverse checker board for 64×64 bit arrays. A detailed description of SET, RESET, and read operations is given in Ref. [25]. Benefits of selector-less cross-point arrays are lower operation voltages, reduced number of process steps, and simpler design. Challenges to the approach are the limited array size, read and write disturb, increased SET and RESET currents due to the limited steepness of the I - V characteristic. Improvements can be expected from an increase in the steepness of the I - V characteristic. A successful increase in the steepness will directly

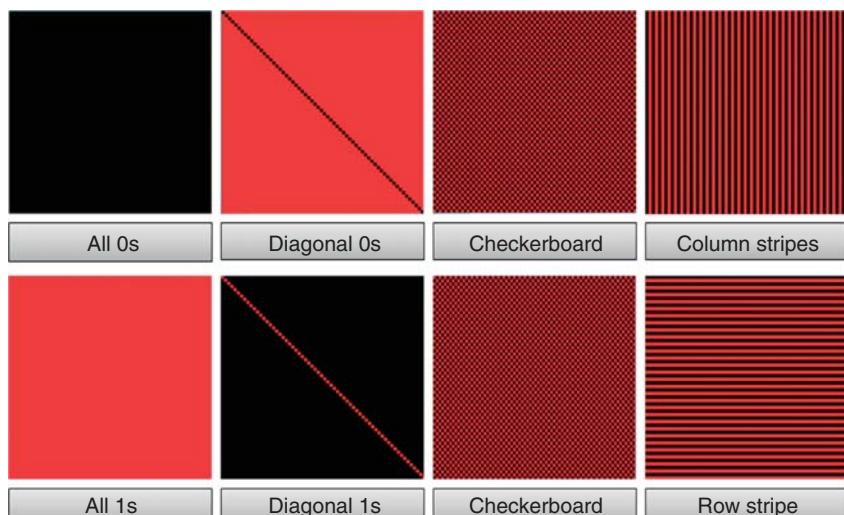


Figure 16.15 Demonstration of set, reset, and read operations in a true selector-less 64×64 bits cross-point array (1R) composed of tunnel ReRAM memory devices. The intrinsic self-selecting property of the devices is sufficient for small- to medium-size arrays [61].

improve the array efficiency by enabling larger arrays and will reduce the power consumption by reducing leakage currents.

16.6

Ferroelectric Resistive Switching

16.6.1

Classification of Ferroelectric Resistive Switching

The ferroelectric RS effect based on polarization reversal in ferroelectrics was first reported by Esaki in the 1970s [64]. Since then, ferroelectric RS has been conceptionally described [65–67] and reported for a number of ferroelectric perovskite titanates, such as PbTiO_3 [68–70], BaTiO_3 [71–73], and multiferroic BiFeO_3 [74–77].

In general, the ferroelectric RS memories are categorized into two major types depending on the conduction mechanism: a ferroelectric tunneling junction [64, 71–73] and a ferroelectric diode [68, 70, 74–77]. The ferroelectric tunneling junction consists of an ultrathin ferroelectric tunneling barrier, and reversal of its ferroelectric polarization induces a change in the tunneling barrier height. The barrier height modification due to polarization reversal requires an asymmetric potential distribution in the ferroelectric barrier (Figure 16.16). As schematically shown in Figure 16.16b, such an asymmetric potential distribution is caused by different top and bottom electrode metals having different screening lengths [78, 79]. Another way to tailor the asymmetric potential distribution is to insert an ultrathin dielectric (nonferroelectric) layer at the interface between an electrode and a ferroelectric layer, as schematically shown in Figure 16.16c [66, 80, 81]. The thin dielectric layer separates the screening charge in the electrode from the polarization charge in the ferroelectric layer, and thus, a finite potential distribution forms in the dielectric layer. For the ferroelectric diode, a Schottky-like barrier might form at an electrode/ferroelectric interface, and the potential distribution of this Schottky-like barrier changes depending on the ferroelectric polarization direction, as schematically shown in Figure 16.16d [68].

In this section, we focus on ferroelectric RS diodes and describe our current understanding of the mechanism and device characteristics.

16.6.2

Ferroelectric Resistive-Switching Diode

In general, ferroelectric oxides are insulating. However, when ferroelectric oxides have cation or oxygen deficiencies, they can be semiconducting. Capacitors consisting of such semiconducting ferroelectric oxides can show rectification (diode behavior) as well as RS [68, 70, 74–77, 82]. Figure 16.17 shows the I - V characteristics of a capacitor cell consisting of a Bi-deficient $\text{Bi}_{1-\delta}\text{FeO}_3$ film (100 nm), which acts as a p-type semiconductor sandwiched between a Pt top electrode

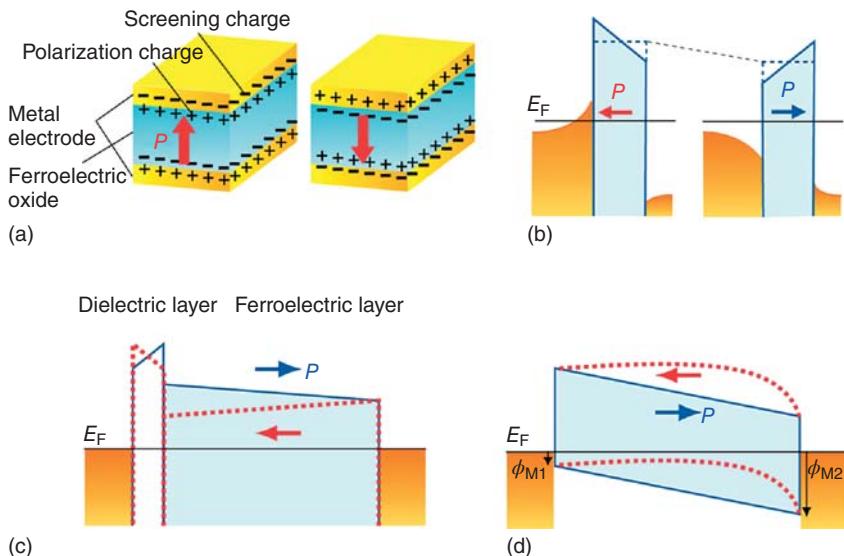


Figure 16.16 (a) Schematic of a ferroelectric capacitor. (b) Schematic of the potential distribution in a ferroelectric tunneling junction for different polarization directions, assuming that the screening length in the left electrode is greater than that in the right electrode. The dashed lines show the average barrier height, which changes depending on the polarization direction. (c) Schematic of the potential distribution

in a ferroelectric tunneling junction that has an interfacial dielectric layer. (d) Schematic of the potential distribution in a ferroelectric Schottky diode, assuming that the Schottky barrier (ϕ_{M2}) at the right interface is larger than that (ϕ_{M1}) at the left interface. The potential distribution in (c) and (d) changes depending on the polarization direction. (Sources: (b-d) were redrawn with modification from Refs. [78, 81, 68], respectively.)

and an SrRuO_3 bottom electrode (Pt/BFO/SRO) measured at a voltage-sweeping frequency (f) of 1 and 0.1 kHz [76]. The cell shows a rectifying and hysteretic I - V curve, indicating the RS. The rectification direction indicates that a p-type Schottky-like barrier forms at the Pt/BFO interface. In the I - V curve measured at $f = 1$ kHz, positive and negative current peaks can be observed around 6.5 and -2.5 V, respectively. The resistance state of the cell changes after the positive and negative current peaks, suggesting that the RS is mediated by a large current flow in the device. However, as f is decreased to 0.1 kHz, no current peak is observed in the I - V curve. This result suggests that the origin of the current peaks is a transient current, such as a ferroelectric displacement current, and that polarization reversal is involved in the RS in the cell.

It has been demonstrated that RS can be artificially induced in interface-engineered ferroelectric capacitors in which a thin LaFeO_3 (LFO) dielectric layer was inserted at one of the interfaces of SRO/BFO/SRO cells [77]. Figure 16.18a,b show the I - V characteristics of SRO/BFO/SRO and SRO/LFO(10 nm)/BFO/SRO cells measured at a voltage-sweeping frequency f of 1 kHz. The SRO/BFO/SRO

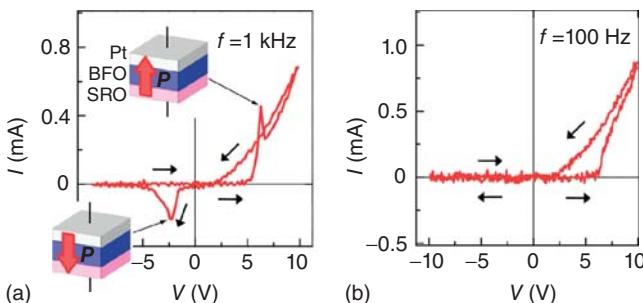


Figure 16.17 I - V characteristics of Pt/BFO/SRO cell measured at a voltage-sweeping frequency (f) of (a) 1 and (b) 0.1 kHz. Positive and negative current peaks originating from the ferroelectric

displacement current are observed in the I - V curve measured at $f = 1\text{ kHz}$. (Adapted with permission from Ref. [76]. © 2012 Wiley-VCH.)

cell shows practically ohmic I - V characteristics without hysteresis. On the other hand, the SRO/LFO/BFO/SRO cell shows hysteretic and rectifying I - V characteristics, indicating the emergence of RS as well as the formation of an asymmetric barrier at the SRO/LFO/BFO interface (Figure 16.18d). This result can be understood by an asymmetric potential distribution generated by the separation of the screening charge and the polarization charge due to the interfacial dielectric layer, as is the case with the ferroelectric tunneling junction, which has an ultrathin dielectric layer at the interface (Figure 16.16c).

Figure 16.19 shows the out-of-plane piezoresponse force microscopy (PFM) phase and amplitude hysteresis loops measured in an SRO/LFO/BFO/SRO cell [77]. These results suggest that the BFO layer is ferroelectric and has asymmetric coercive voltages of about +5.0 and -2.0 V. In addition, the PFM phase hysteresis loop is in good agreement with the current (resistance) switching hysteresis loop (Figure 16.19, upper panel), indicating that polarization reversal is involved in the RS in the cell.

A critical issue with the ferroelectric RS memories is the data retention property [77]. For the emergence of ferroelectric RS, an asymmetric potential distribution is required, meaning that a depolarization field does not completely vanish in the memory cell. It is well known that in conventional ferroelectric random access memory (FeRAM), the depolarization field reduces the ferroelectric polarization with time [83], resulting in poor data retention. For the Pt/BFO/SRO cell, however, relatively good retention values at room temperature have been reported (Figure 16.20a) [76]. This relatively good retention of the cell may be attributable to the high coercive field of BFO film [84]. However, the retention property of the cell degrades rapidly with increasing temperature [76, 77]. For practical use of ferroelectric resistive memory, the retention at higher temperature should be improved.

Figure 16.20b shows the endurance test of the Pt/BFO/SRO cell [76]. The cell showed a large R_H/R_L ratio of >10 up to 10^5 cycles, and the R_H/R_L ratio was as

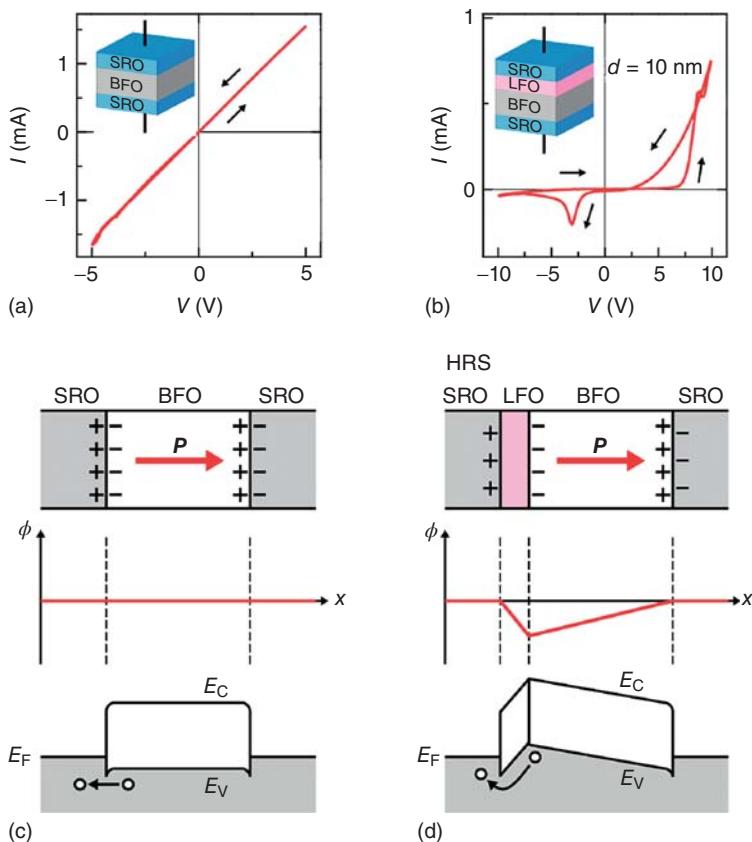


Figure 16.18 I - V characteristics of (a) SRO/BFO(100 nm)/SRO cell and (b) SRO/LFO(10 nm)/BFO(10 nm)/SRO cell measured at a voltage-sweeping frequency f of 1 kHz. The SRO/BFO/SRO cell shows practically ohmic I - V characteristics, whereas the SRO/LFO/BFO/SRO cell shows hysteretic and rectifying I - V characteristics,

indicating the emergence of resistive switching. Schematic illustrations of (top) device geometry with space charges, (middle) potential distributions, and (bottom) band diagrams of (c) SRO/BFO/SRO cell and (d) SRO/LFO/BFO/SRO cell (high-resistance state). (Adapted with permission from Ref. [77]. © 2013 AIP.)

large as 3 even at 10^6 cycles. Similar endurance values have been reported for the ferroelectric RS diode consisting of oxygen-deficient $\text{BiFeO}_{3-\delta}$,[75], implying that the limited endurance may result from these deficiencies (defects). For conventional FeRAM, it has been reported that defects have a significant impact on fatigue and endurance [85]. Therefore, the control of crystallinity and chemical composition in the ferroelectric oxide films is crucial for improving the endurance of ferroelectric resistive memory.

It should be noted that the topic of ferroelectric (and multiferroic) RS is in a research stage, and the mechanism is still under debate. The possibility that some

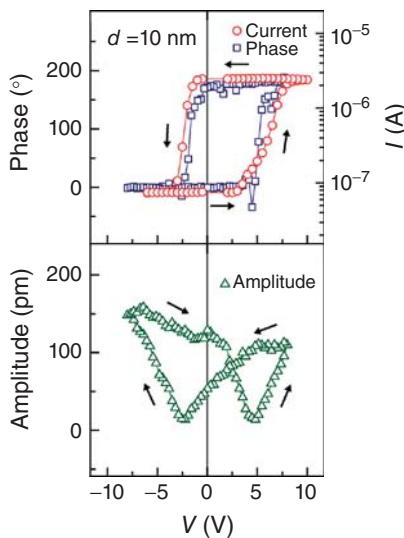


Figure 16.19 Out-of-plane PFM phase (blue squares), amplitude (green triangles) and pulsed-voltage-induced current (resistance) switching (red circles) hysteresis loops measured for an SRO/LFO(10 nm)/BFO/SRO device. For current-hysteresis measurement, a voltage pulse with τ_p of 10 ms was applied, and then a current value at $V = +2$ V was measured. (Reprinted with permission from Ref. [77]. © 2013 AIP.)

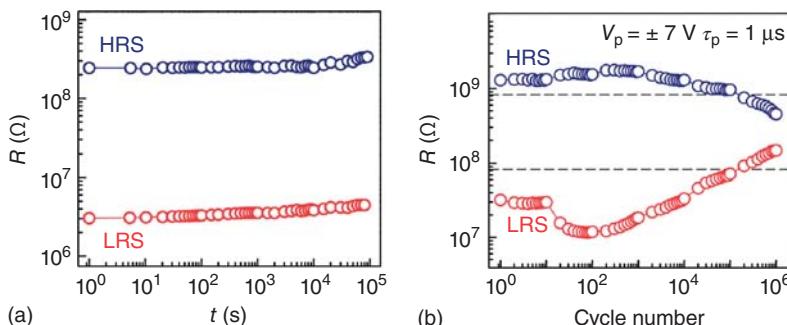


Figure 16.20 (a) Data retention and (b) endurance characteristics of a Pt/BFO/SRO cell. Broken lines in (b) represent a memory window of $R_H/R_L = 10$. (Adapted with permission from Ref. [76]. © 2012 Wiley-VCH.)

of the results are caused by redox reactions in some of the reports cannot be excluded. Further detailed studies are thus required for a definitive understanding of the mechanism.

16.7 Summary

Homogeneous *interface-type* switching, which is also called *area-scaling-type* switching, is often observed in RS memory cells consisting of oxides that are semiconducting or conducting in the pristine state. This type of switching takes place at the interface between the metal electrode and the oxide. In this chapter,

we have focused on two model materials, which are most studied in the context of interface-type RS: mixed-valence manganites such as $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ (short: PCMO), which shows p-conductivity, typically with a thin binary oxide layer (formed by redox reaction or deliberately by deposition) between the manganite and the active electrode, and donor-doped SrTiO_3 , typically Nb-doped SrTiO_3 (short: Nb:STO), which is an n-conductor. The possible electronic and ionic mechanisms as well as the competition between area-dependent switching and filamentary switching are mentioned. In addition, a brief and preliminary review of ferroelectric RS is provided.

Acknowledgment

The authors thank one of the editors (RW) for the discussions and for contributing to the text in Sections 16.2–16.4.

References

1. Baikalov, A. *et al.* (2003) *Appl. Phys. Lett.*, **83** (5), 957–959.
2. Chen, X., Wu, N.J., Strozier, J., and Ignatiev, A. (2005) *Appl. Phys. Lett.*, **87** (23), 233506-1–233506-3.
3. Seong, D.J., Jo, M., Lee, D., and Hwang, H. (2007) *Electrochem. Solid-State Lett.*, **10** (6), H168–H170.
4. Nian, Y.B., Strozier, J., Wu, N.J., Chen, X., and Ignatiev, A. (2007) *Phys. Rev. Lett.*, **98** (14), 146403/1–146403/4.
5. Jeon, S.H., Park, B.H., Lee, J., Lee, B., and Han, S. (2006) *Appl. Phys. Lett.*, **89** (4), 42904-1–42904-3.
6. Asanuma, S., Akoh, H., Yamada, H., and Sawa, A. (2009) *Phys. Rev. B*, **80** (23), 235113/1–235113/8.
7. Muenstermann, R., Menke, T., Dittmann, R., and Waser, R. (2010) *Adv. Mater.*, **22** (43), 4819–4822.
8. Sawa, A., Fujii, T., Kawasaki, M., and Tokura, Y. (2004) *Appl. Phys. Lett.*, **85** (18), 4073–4075.
9. Fujii, T., Kawasaki, M., Sawa, A., Akoh, H., Kawazoe, Y., and Tokura, Y. (2005) *Appl. Phys. Lett.*, **86** (1), 12107-1–12107-3.
10. Fors, R., Khartsev, S.I., and Grishin, A.M. (2005) *Phys. Rev. B: Condens. Matter Mater. Phys.*, **71** (4), 45305-1–45305-10.
11. Rozenberg, M.J., Inoue, I.H., and Sanchez, M.J. (2004) *Phys. Rev. Lett.*, **92** (17), 178302/1–178302/4.
12. Oka, T. and Nagaosa, N. (2005) *Phys. Rev. Lett.*, **95** (26), 266403/1–266403/4.
13. Xue, K.H., de Araujo, C.A.P., Celinska, J., and McWilliams, C. (2011) *J. Appl. Phys.* (21st International Symposium on Integrated Ferroelectrics and Functionalities, Colorado Springs, CO) 109, 091602/1–091602/6.
14. Sim, H. *et al.* (2005) Technical Digest of IEEE International Electron Devices Meeting 2005, Washington, DC, pp. 758–761.
15. Hasan, M. *et al* (2008) *Appl. Phys. Lett.*, **92**, 202102.
16. Asamitsu, A., Tomioka, Y., Kuwahara, H., and Tokura, Y. (1997) *Nature*, **388** (6637), 50–52.
17. Xing, Z.W., Wu, N.J., and Ignatiev, A. (2007) *Appl. Phys. Lett.*, **91** (5), 52106/1.
18. Fujimoto, M., Koyama, H., Nishi, Y., and Suzuki, T. (2007) *Appl. Phys. Lett.*, **91** (22), 223504/1–223504/3.
19. Seong, D.J. *et al.* (2009) *IEEE Electron Device Lett.*, **30** (9), 919–921.
20. Asanuma, S., Akoh, H., Yamada, H., and Sawa, A. (2009) *Phys. Rev. B: Condens. Matter*, **80** (23), 235113/1.
21. Liao, Z. *et al* (2011) *Appl. Phys. Lett.*, **99** (11), 113506/1.

22. Scherff, M., Hoffmann, J., Meyer, B., Danz, T., and Jooss, C. (2013) *New J. Phys.*, **15**, 103008/1.
23. Herpers, A. *et al.* (2014) *Adv. Mater.*, **26**, 2730–2735.
24. Meyer, R. *et al.* (2008) 2008 9th Annual Non-Volatile Memory Technology Symposium, Proceedings, pp. 54–58.
25. Chevallier, C.J. *et al.* (2010) Technical Digest 2010 IEEE international Solid-State Circuits Conference (ISSCC), p. 260.
26. Hansen, M. *et al.*
27. Sawa, A. (2008) *Mater. Today*, **11** (6), 28–36.
28. Tokunaga, Y. *et al.* (2006) *Appl. Phys. Lett.*, **88** (22), 223507-1-3.
29. Fujii, T., Kawasaki, M., Sawa, A., Kawazoe, Y., Akoh, H., and Tokura, Y. (2007) *Phys. Rev. B*, **75** (16), 165101.
30. Moos, R. and Härdtl, K.H. (1997) *J. Am. Ceram. Soc.*, **80** (10), 2549–2562.
31. Meyer, R. and Waser, R. (2001) *J. Eur. Ceram. Soc.*, **21** (10-11), 1743–1747.
32. Meyer, R., Waser, R., Helmbold, J., and Borchardt, G. (2003) *Phys. Rev. Lett.*, **90** (10), 105901/1–105901/4.
33. Imada, M., Fujimori, A., and Tokura, Y. (1998) *Rev. Mod. Phys.*, **70** (4), 1039–1263.
34. Coey, J.M.D., Viret, M., and von Molnar, S. (1999) *Adv. Phys.*, **48** (2), 167–293.
35. Herpers, A. (2014).
36. Coey, J., Viret, M., and von Molnár, S. (2009) *Adv. Phys.*, **58**, 571–697.
37. Malavasi, L. (2008) *J. Mater. Chem.*, **18** (28), 3295–3308.
38. Tokura, Y. and Tomioka, Y. (1999) *J. Magn. Magn. Mater.*, **200**, 1.
39. Raabe, S. *et al* (2012) *Adv. Funct. Mater.*, **22** (16), 3378–3388.
40. Dionne, G.F. (2009) *Magnetic Oxides*, Springer-Verlag.
41. Medvedeva, J.E., Anisimov, V.I., Mryasov, O.N., and Freeman, A.J. (2002) *J. Phys. Condens. Matter*, **14**, 4533–4542.
42. Quintero, M., Levy, P., Leyva, A.G., and Rozenberg, M.J. (2007) *Phys. Rev. Lett.*, **98** (11), 116601/1–116601/4.
43. Rodenbücher, C., Speier, W., Bihlmayer, G., Breuer, U., Waser, R., and Szot, K. (2013) *New J. Phys.*, **15**, 103017.
44. Yang, M. *et al* (2014) *J. Appl. Phys.*, **115** (13), 134505/1.
45. Park, C., Seo, Y., Jung, J., and Kim, D. (2008) *J. Appl. Phys.*, **103** (5), 54106/1.
46. Mikheev, E., Hoskins, B.D., Strukov, D.B., and Stemmer, S. (2014) *Nat. Commun.*, **5**, 3990.
47. Bourim, E.M. and Kim, D.W. (2013) *Curr. Appl. Phys.*, **13** (3), 505–509.
48. Seong, D.J., Lee, D., Pyun, M., Yoon, J., and Hwang, H. (2008) *Jpn. J. Appl. Phys.*, **47** (12), 8749–8751.
49. Li, J., Ohashi, N., Okushi, H., and Haneda, H. (2011) *Phys. Rev. B*, **83** (12), 125317/1.
50. Lee, E., Gwon, M., Kim, D.W., and Kim, H. (2011) *Appl. Phys. Lett.*, **98** (13), 132905/1–132905/3.
51. Lu, H.X., Liu, Y.B., Chen, Y.S., Wang, J., Shen, B.G., and Sun, J.R. (2014) *J. Appl. Phys.*, **116** (17), 173710.
52. Park, J., Kwon, D.-H., Park, H., Jung, C.U., and Kim, M. (2014) *Appl. Phys. Lett.*, **105**, 183103.
53. Lee, S., Lee, J.S., Park, J.-B., Kyoung, Y.K., Lee, M.-J., and Noh, T.W. (2014) *APL Mater.*, **2**, 066103.
54. Lee, J.S., Lee, S.B., Kahng, B., and Noh, T.W. (2013) *Appl. Phys. Lett.*, **102** (25), 253503/1–253503/4.
55. Lee, H.S., Choi, S.G., Park, H.H., and Rozenberg, M.J. (2013) *Sci. Rep.*, **3**, 1704/1.
56. Sawa, A., Fujii, T., Kawasaki, M., and Tokura, Y. (2005) *Proc. SPIE*, **5932**, 59322C.
57. Borgatti, F. *et al.* (2013) *Nanoscale*, **5**, 3954.
58. Sawa, A. *et al* (2007) *Appl. Phys. Lett.*, **90** (25), 252102/1–252102/3.
59. Norpoth, J., Mildner, S., Scherff, M., Hoffmann, J., and Jooss, C. (2014) *Nanoscale*, **6** (16), 9852–9862.
60. Scherff, M., Meyer, B., Hoffmann, J., Jooss, C., Feuchter, M., and Kamlah, M. (2015) *New J. Phys.*, **17**, 033011.
61. Meyer, R. (2014) European Materials Research Society (EMRS) 2014 Spring Meeting, Lille, May 26-30, 2014.
62. Minseok, J. *et al.* (2010) Technical Digest of 2010 Symposium on VLSI Technology, pp. 53–54.
63. Lee, J., Jo, M., Seong, D.J., Shin, J., and Hwang, H. (2011) *Microelectron. Eng.*, **88** (7), 1113–1118.

64. Esaki, L., Laibowitz, R.B., and Stiles, P.J. (1971) *IBM Tech. Discl. Bull.*, **13** (8), 2161.
65. Tsymbal, E.Y. and Kohlstedt, H. (2006) *Science*, **313**, 181–183.
66. Meyer, R. and Waser, R. (2006) *J. Appl. Phys.*, **100** (5), 1611-1-8–1611-8.
67. Tsymbal, E., Gruverman, A., Garcia, V., Bibes, M., and Barthélémy, A. (2012) *MRS Bull.*, **37**, 138–143.
68. Blom, P.W.M., Wolf, R.M., Cillessen, J.F.M., and Krijn, M.P.C.M. (1994) *Phys. Rev. Lett.*, **73** (15), 2107–2110.
69. Contreras, J.R., Kohlstedt, H., Poppe, U., Waser, R., Buchal, C., and Pertsev, N.A. (2003) *Appl. Phys. Lett.*, **83** (22), 4595–4597.
70. MakSYMovich, P., Yu, S., Jesse, P., Ramesh, R., Kalinin, A.P., and Baddorf, S.V. (2009) *Science*, **324**, 1421–1425.
71. Garcia, V. *et al* (2009) *Nature*, **460** (7251), 81–84.
72. Chanthbouala, A. *et al* (2012) *Nat. Nanotechnol.*, **7**, 101.
73. Kim, D.J. *et al*. (2012) *Nano Lett.*, **12**, 5697.
74. Choi, T., Lee, S., Choi, Y.J., Kiryukhin, V., and Cheong, S.W. (2009) *Science*, **324**, 63.
75. Jiang, A.Q. *et al* (2011) *Adv. Mater.*, **23** (10), 1277–1281.
76. Tsurumaki, A., Yamada, H., and Sawa, A. (2012) *Adv. Funct. Mater.*, **22**, 1040–1048.
77. Tsurumaki-Fukuchi, A., Yamada, H., and Sawa, A. (2013) *Appl. Phys. Lett.*, **103** (15), 263502.
78. Zhuravlev, M., Sabirianov, R., Jaswal, S., and Tsymbal, E. (2005) *Phys. Rev. Lett.*, **94**, 246802.
79. Pantel, D., Goetze, S., Hesse, D., and Alexe, M. (2011) *ACS Nano*, **5** (7), 6032–6038.
80. Meyer, R., Rodriguez-Contreras, J., Petrarau, A., and Kohlstedt, H. (2004) *Integr. Ferroelectr.*, **68**, 77–88.
81. Zhuravlev, M.Y., Wang, Y., Maekawa, S., and Tsymbal, E.Y. (2009) *Appl. Phys. Lett.*, **95**, 052902.
82. Kim, T.H. *et al*. (2012) *Adv. Funct. Mater.*, **22**, 4962.
83. Mehta, R.R., Silverman, B.D., and Jacobs, J.T. (1973) *J. Appl. Phys.*, **44** (8), 3379–3385.
84. Wang, J. *et al*. (2003) *Science*, **299**, 1719–1722.
85. Scott, J.F. and Paz-de-Araujo, C.A. (1989) *Science*, **246** (4936), 1400–1405.

17

Electrochemical Metallization Memories

Michael N. Kozicki, Maria Mitkova, and Ilia Valov

17.1

Introduction

In this chapter, we describe a resistance change technology that is based entirely on metal cation transport and redox reactions in simple two-terminal metal–electrolyte–metal/metal–insulator–metal (MEM/MIM) device structures 005B [1]. The device uses an oxidizable (soluble) active electrode, typically (but not limited to) Ag or Cu. The counter electrode is an inert (not soluble in the solid electrolyte) electrode, for example, W, TiN, and Pt. An oxide, chalcogenide, or halide material between the electrodes serves to transport the metal cations [2]. A positive voltage applied to the oxidizable electrode leads to the dissolution of the metal and deposition of a metallic (Ag- or Cu-containing) filament at the opposite electrode. This metallic filament ultimately bridges the relatively insulating ion conductor and defines the low-resistance on state (or states). The filament is dissolved by applying a voltage of the opposite polarity, which returns the cell to a high-resistance off state, defined by the area of the cell and the resistivity of the ion conductor.

The basic elements of the device may be configured laterally or vertically. Whereas lateral devices, which typically have the electrodes placed at different points on the surface of a thin ion conducting film, may be utilized in a variety of applications, for example, radio frequency (RF) switches [3], it is the vertical configuration that is most applicable to memory devices. Vertical cells occupy the smallest possible area, which is critical for high-density memory arrays. These devices are nonvolatile, have the requisite low power consumption, and possess excellent prospects for scalability to atomic dimensions [4]. In addition, the distance between the electrodes is defined by the thickness of the ion conducting layer and can therefore be very small, for example, a few nanometers to a few tens of nanometers. This nanoscale distance allows rapid bridging by the electrodeposit, which results in fast switching. An example of the operation of a vertical cell is shown in Figure 17.1.

As with any novel technology that is being examined by several research groups simultaneously, various names and acronyms have been applied to this

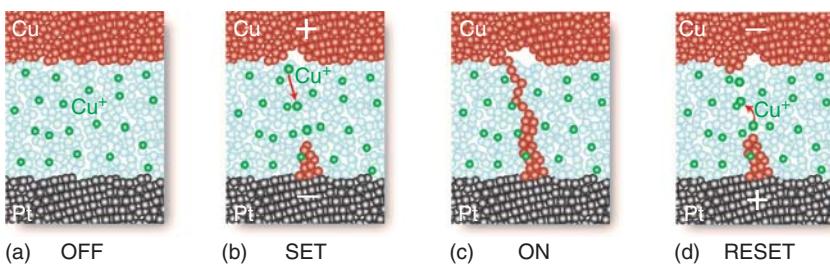


Figure 17.1 Operation principles of cation-based memory cells with a Cu oxidizable electrode and Pt inert electrode. (a) The initial/OFF state. (b) The oxidation of metallic

Cu to Cu⁺ ions and the growth of a Cu filament in the “program” or “set” operation. (c) The ON state. (d) The reverse bias removal of the filament in the “erase” or “reset” [1].

type of redox-based resistive-switching devices. The original term programmable metallization cell (PMC) now refers to the technology platform that encompasses a diverse range of applications beyond memory [3, 5]. The term conductive bridging random access memory (CBRAM) became popular during product development, and this is now typically used by the semiconductor industry exclusively for the memory arrays [6]. The term *ECM* (electro-chemical metallization) is typically used in the scientific literature when describing the underlying electrochemical processes [7]. “Atomic switch” or “gap-type atomic switch” was introduced to differentiate cells in which switching occurs in a vacuum gap from MIM structures [8]. In addition, the term *electrochemical memory* and the generic expressions *ionic memory* and *nanoionic memory* have also been applied.

17.2 Metal Ion Conductors

17.2.1 Materials

An assortment of chalcogenides, oxides, and halides has been employed in cation-based resistive-switching devices [2]. Chalcogenide materials contain group VI or “chalcogen” atoms, S, Se, Te, and O, although oxides are usually considered separately from the others in the literature [9]. Interestingly, even though the transport properties of these materials can be quite different, their resistive-switching behavior tends to be quite similar; the dissimilar macroscopic properties of these materials tend to become less obvious when the material thickness (between the electrodes) is small. This means that materials ranging from well-known superionic electrolytes such as AgI to insulators such as SiO₂ can behave as solid electrolytes at room temperature [10, 11]. Table 17.1 gives examples of the ion conducting materials used in ECM (CBRAM).

Table 17.1 Solid ion conductors used in cation-based devices.

Insulators	Mixed conductors	Ion electrolytes
SiO_2 [12–15]	$\text{Ag}_{2+\delta}\text{S}$ [8, 16–18]	RbAg_4I_5 [19, 20]
Ta_2O_5 [21, 22]	$\text{Cu}_{2-\delta}\text{S}$ [23–25]	AgI [26, 27]
ZrO_2 [28]	$\text{Cu}_{2-\delta}\text{O}$ [29–32]	$\text{Ag}-\text{PEO}$ [33]
HfO_2 [34]	$\text{Ag}-\text{AsS}_x$ [35–37]	$(\text{AgI})_{0.5}(\text{AgPO}_3)_{0.5}$ [38]
TiO_2 [39]	$\text{Ag}-\text{GeSe}_x$ [40–44]	
WO_3 [45–47]	$\text{Ag}-\text{GeS}_x$ [37, 41, 46] $\text{Cu}-\text{GeSe}_x$ [44]	
GeO_x [48]	Cu-doped SiO_2 [49]	
GeS_x [44, 50]	Cu-doped WO_3 [51]	
GeSe_x [43, 52]	Cu-doped ZrO_2 [53] $\text{Cu}-\text{GeTe}$ [54] $\text{Cu}-\text{TCNQ}$ [55, 56]	
	a-Si [57–59]	

Adapted from Ref. [1].

Unlike electrons, ions in a solid cannot be delocalized, that is, they are constrained to lattice sites and move by random hopping in the crystal lattice or through a network of channels (see Chapters 4 and 5). These channels may be a consequence of order in the material, as in the case of the interstitial channels in certain directions in crystalline materials, or may be due to long-range disorder, as in amorphous (glassy) and/or nanoscopically porous materials. In general, the smallest ions tend to possess high mobilities as they can readily move through the physical pathways in the supporting material. So, small ionic radius elements such as Li and Na are particularly mobile in a variety of materials but their high chemical reactivity makes them unsuitable for integrated memory applications. Ag is much more appropriate in this context due to its nobility, high conductivity, and low overpotential for both reduction and oxidation processes. Ag halides such as AgI and silver chalcogenides, Ag_2S , Ag_2Se , and Ag_2Te , have been widely studied as solid electrolytes. The low-temperature crystalline phases of these materials are less conductive with mixed ionic and electronic contributions, but their high temperature polymorphs (e.g., the cubic phase of Ag_2Se , which is stable above 133 °C) are excellent ion conductors [60, 61]. In these materials, the number of sites that can accommodate Ag is typically much higher than the number of mobile ions, and this ensures that there are always sites for ions to move into. This, in combination with the low potential barriers between these sites (typically a few tenths of an electronvolts), results in the superionic nature of these materials. Deviations from stoichiometry (δ in Table 17.1) as low as 100 ppm can lead to the existence of both mobile ions and electrons and so, nonstoichiometric materials are regarded as mixed conductors [62]. Cu-containing analogs (e.g., Cu_2S) have comparable properties, although the participation of the electrons from the Cu d orbital will result in a more complex bonding structure [63].

Thermal stability is important for practical purposes as it allows the materials to withstand the temperatures used in integrated circuit fabrication. There are two major factors contributing to the glass transition temperature of the materials (T_g) – the coordination of the amorphous network and the strength of the chemical bonding, with coordination exerting the most influence. Some chalcogenide glasses have relatively low coordination and, therefore, low glass transition temperature, but Ge-containing glasses have a relatively high T_g since Ge is usually fourfold coordinated. These glasses are, therefore, preferred for the fabrication of devices. Note that the inclusion of elements with high melting points does not necessarily improve T_g as these may form weak chemical bonds within the system or have a coordination number lower than 4, resulting in lower glass transition temperatures. The sensitivity of the properties of binary chalcogenide superionic conductors to composition and temperature/phase tends to limit their use in memory applications, and so, as Table 17.1 illustrates, the emphasis has been on other materials including oxides and higher order compounds.

In chalcogenide ternaries, several tens of atomic percent of the mobile metal, Ag or Cu, is usually added to binary base glasses involving a group IV or group V atom, such as As–S or Ge–Se, thus forming the active layer (electrolyte) of the memory device. The metal can be incorporated during the formation of the source material, during fabrication, or even during the initial stages of device operation in a process known as *forming* or *conditioning*. In general, these ternary glasses are more rigid than organic polymers but more flexible than a typical oxide glass and their other physical properties follow the same tendency. This structural flexibility allows the formation of voids through which ions can easily move from one equilibrium position to another and also allows the formation of electrodeposits within the electrolyte (discussed later). Adding Ag to a chalcogenide base glass can be achieved by diffusing the metal from a thin surface film via the process of photodissolution. This technique uses light energy greater than the optical gap of the glass to create charged defects near the interface between the metal or metal-rich and unreacted chalcogenide layers [64]. The holes created are trapped by the metal to form ions while the electrons move into the chalcogenide film. The electric field so formed is sufficient to allow the ions to overcome the energy barrier at the interface and move into the chalcogenide [65]. For convenience, the UV light used for photodissolution typically is the same as that used for photolithography, with a wavelength of 365 nm. This corresponds to an energy of 3.39 eV, which is more than sufficient for chalcogenide glasses, which typically have a gap less than 3 eV. However, oxides tend to have much larger band gaps (SiO_2 is around 9 eV), and so, thermal diffusion (discussed later) is favored over photodissolution.

In metal-doped chalcogenide glasses, the metal ions are associated with the nonbridging chalcogen atoms but the bonds formed are relatively long, for example, 0.25 nm in Ag–Ge–S ternaries [66]. Since the coulombic energy is proportional to the inverse of the cation–anion separation, the long bonds lead to relatively small attractive forces; the Ge-chalcogenide glasses are therefore among the electrolytes with the lowest coulombic energies [66]. This gives an opportunity to apply the Anderson–Stuart model for the diffusion of Ag

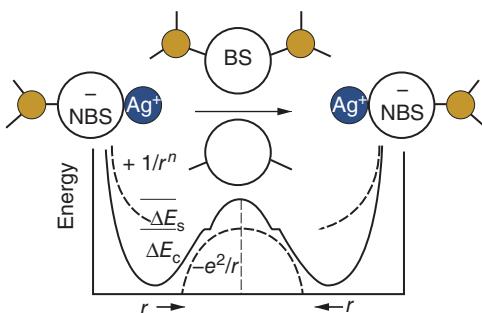


Figure 17.2 Schematic illustration of the Anderson–Stuart model of ionic conduction. The activation energy of the charged Ag^+ is shown to be a function of E_s and E_c , where E_s is caused by the physical impediment of

the bonding Se (BS) atoms. E_c is determined by the difference in coulombic energy between at the saddle point and initial lattice position. Adapted from Ref. [68].

in these materials. In this model, ions are assumed to hop between available equivalent sites separated by an activation energy composed of two terms – a *coulombic energy* and a *strain energy* [67, 68]. The coulombic energy arises from an interaction of the backbone with the charged ion. The strain energy arises from the physical impediment that neighboring atoms from the glass backbone pose to the migrating ion. Considering both, it becomes clear that the charge and the size of the migrating atom will determine its mobility through the glass. Figure 17.2 illustrates an example of the Anderson–Stuart model proposed for Ag–Ge–Se system.

The figure shows the Ag^+ ion migrating between two equivalent sites of non-bonding Se (NBS) atoms. The bonding Se (BS) atoms create a “doorway” and inherent strain energy (E_s) the Ag^+ must overcome before migrating through the material. The coulombic energy (E_c) is difference between the highest energy barrier at the doorway, that is, the saddle-point energy, and the energy of the Ag^+ at the initial lattice position.

For the simple Anderson–Stuart model, the coulombic energy is given by

$$\Delta E_c = \frac{M}{\gamma} \left[\frac{-Z_{ce} Z_{ae} e^2}{\lambda/2} - \frac{-Z_{ce} Z_{ae} e^2}{(r_c + r_a)} \right], \quad (17.1)$$

where M = Madelung constant (used to approximate the electrostatic potential of an ion in a crystal), γ = correction term (typically set equal to the material dielectric constant), Z_{ce} = cation charge, $-Z_{ae}$ = anion charge, r_c = cation radius, r_a = anion radius, and λ = jump distance from one NBS to the other. In Eq. (17.1), the first term in the parentheses represents the coulombic energy at the transition site while the second term represents the bonding energy between the NBS anion and the migrating Ag^+ cation. Rearrangement of Eq. (17.1) results in Eq. (17.2).

$$\Delta E_c = \frac{M Z_{ce} Z_{ae} e^2}{\gamma} \left[\frac{1}{(r_c + r_a)} - \frac{2}{\lambda} \right], \quad (17.2)$$

where M is dependent on the structure of the material and, thus, will change as the stoichiometry changes. The strain energy, E_s , is approximated as the energy required to dilate a cylinder for the Ag^+ cation to jump between NBS sites and is given by

$$\Delta E_s = \frac{\pi G(r_c - r_d)^2 \lambda}{2}, \quad (17.3)$$

where G = shear modulus and r_d = radius of the doorway (i.e., the distance between the BS atoms). Considering that the doorway radius will change with the structure of the backbone material, this strain energy represents a critical structure term that will dictate the Ag^+ migration through the chalcogenide glass. It should be noted here that the structure of the chalcogenide films is strongly dependent on the formation method. For example, thermally evaporated and chemical vapor deposition (CVD)-deposited films usually result in a relaxed structure, which resembles the equilibrium structure of the bulk material with the same composition, while sputtering results in stressed films due to the specific kinetics of their formation. These formation-related differences can lead to different results from different research groups.

Thermal processes allow partial dissociation, which results in defect formation followed by ion migration, and the activation energy for this strongly depends on the distance between the hopping cation and the anion located at the next-nearest neighbor position and the height of the barrier between these locations. The sites are plentiful and, therefore, are closely spaced, and the barriers are typically low, in the order of a few tenths of an electron volt. These factors, in conjunction with the structural voids through which ions can easily move, mean that Ge-chalcogenide glasses have relatively low activation energies for ion transport [66, 69] and, therefore, possess high ion mobility even at room temperature. However, due to the disordered structure of the films, there are many traps in their structure. By performing *ab initio* thermal molecular dynamics (MD) simulations at various temperatures, it has been found that Ag motion may be tracked on timescales that are accessible (tens of picoseconds). In Figure 17.3a, the trajectories of three different Ag atoms are shown for a simulation at 1000 K [70].

The figure provides direct evidence of the existence of trapping centers (TCs) of Scher–Lax–Phillips type. Estimated Ag self-diffusion constants were within a factor of 10 of the experimental value extrapolated to room temperature. To detail the nature of the TCs and check the local-orbital basis set, calculations with VASP have also been reported [71]. Here, silver dynamics was studied by constant-temperature dynamics at 300 and 700 K. Extended trajectories of 20 ps were obtained. At 300 K, silver is largely trapped: only six hopping events were observed. The silver traps fall into two categories. Type 1 (32%) are strongly bound: 4 Ag atoms sit at single TC with no neighboring TC within a radius of 2.0 Å, and 7 Ag atoms occupy two overlapping TCs with the host pairs making an angle of about 90° with each other. Figure 17.3b indicates typical dynamics of the two types of Ag. The motion is oscillatory within a trap and ballistic between traps.

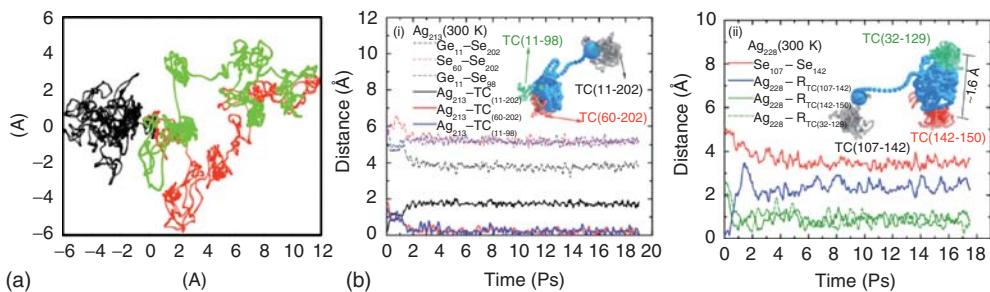


Figure 17.3 (a) 2D projection of Ag motion in $(\text{GeSe}_3)_{0.9}\text{Ag}_{0.1}$. Red and green curves represent paths of most mobile two Ag atoms, black, least mobile [70]. (b) Characteristic examples of silver dynamics: (i) “Type 1” (isolated) trap and (ii) “Type 2” trap (rapid Ag hopping among local group of trapping centers). Inset: trajectories of Ag sites (blue) and neighboring TCs (gray, green, and red) [71].

1” (isolated) trap and (ii) “Type 2” trap (rapid Ag hopping among local group of trapping centers). Inset: trajectories of Ag sites (blue) and neighboring TCs (gray, green, and red) [71].

Because of the strong affinity between the chalcogenide glass and Ag, a strong chemical interaction occurs in the Ag/GeS_x system when GeS_x is not saturated with Ag, leading to a partial or complete dissolution of the active electrode [72]. As stated, the chemical dissolution of Ag or Cu in chalcogenide systems is activated by UV irradiation in GeS_x or by VIS for GeSe_x [73] and also leads to formation of a interfacial monolayers of, for example, an Ag^+ -rich phase (compared to the equilibrated bulk concentration [74]). Also, for other systems, for example, Zn/GeS_x , the formation of metal– GeS_x compounds at the interface has been reported [74, 75]. The mechanisms of this chemical dissolution have been studied in detail by Cho *et al.* [72], reporting that during oxidation, the electron transfers to the Ge 4s states instead of to S 3p. It is not necessary to irradiate the chalcogenide glasses/Ag samples by means of an artificial UV source to dissolve sufficiently thin (up to 10 nm) silver layers, as the UV fraction of the natural daylight apparently provides a sufficient dose to drive the Ag migration into the GeS_x . Thus, to prevent Ag diffusion, one needs to saturate the Ge–Ge bonds. This can be achieved either by surface oxidation (formation of Ge–O bonds) or by the introduction of a barrier film [76]. Moreover, the sequence of film deposition is essential for the determination of a specific electrode/electrolyte interfacial chemical composition and structure. Thus, Ag deposited on GeS_x leads to a preferable formation of an Ag–Ge–S surface layer, whereas deposition of GeS_x on Ag results in an Ag–S interface [74]. The deposition conditions also influence the stoichiometry and the morphology of the films [50]. The reactivity of the interfaces and dissolution of Ag can influence the OFF and even ON state stability for Ge-chalcogenide-based devices, causing a nanobattery effect or shift in the switching voltages [77, 78]. These instabilities can be minimized by using Ag predoped GeS_x but can also be useful as selector elements in cells with complementary resistive-switches (CRSSs) [79].

It should be noted that the addition of several tens of atomic percent of metal to a chalcogenide (or oxide)-based glass typically does not result in a homogeneous material. Considerable changes in the nanostructure occur, which have a profound effect on the macroscopic characteristics. In the case of chalcogenide ternaries,

these changes are due to phase separation. For example, in Ag–Ge–Se, the Ag reacts with the chalcogen to form distributed nanocrystalline Ag_2Se superionic phases in a Ge-rich insulating supporting material [80]. This inhomogeneity is ideal for resistance change devices as the off state is dominated by the electrolyte's high resistivity phase but the availability of mobile ions in the dispersed Ag-rich phases means that the effective ion mobility is high and a metallic filament may still be rapidly formed to switch the structure to its low-resistance state.

The case of Cu in SiO_2 has a different chemical story, but the end result is not entirely dissimilar. When introduced into a deposited oxide film, the Cu does not tend to react with the constituents of the SiO_2 network but instead remains in elemental form [14, 81]. However, at the Cu/ SiO_2 interface, formation of both Cu_2O and CuO has been reported [82]. For annealing temperatures above 600 °C, Cu nanoparticle precipitation is observed [14, 83]. Our own work has shown that carrying out a Cu diffusion at 660 °C leads to optimal device performance, with a high-resistance off state defined by the SiO_2 matrix and stable switching enabled by the abundance of Cu in the distributed nanoscale precipitates [14]. It has been shown that Cu^{2+} is bonded weakly to the oxygen in SiO_2 compared to Cu^+ and, thus, appears to be faster. Significantly higher processing temperatures lead to Cu appearing at the lower interface, which destabilizes the device operation. In the case that SiO_2 (or any other oxide material) is not predoped, the condition is more complicated. To electrochemically dissolve copper into the glassy matrix, a counter charge/reaction is required [84]. It has been shown that moisture plays a significant role in this situation and provides the required counter charge [84–86].

17.2.2

Ion Transport

As described in detail in Chapter 5, mobile ions in a solid electrolyte reside in potential wells separated by low potential barriers, but they possess kinetic energy governed by Boltzmann statistics, which results in thermal diffusion and, in the presence of an electric field, in ionic conduction.

The ion current density created by this drift component simplifies to

$$\mathcal{J} = \mathcal{J}'(T) \exp\left(\frac{ue}{kT}\right) \quad (17.4)$$

This nonlinear effective mobility is common in ionic devices but, as we will see later in this chapter, it is not the only exponential process related to switching. Indeed, in most memory device configurations, ion transport is not likely to be the rate-limiting step in filament formation, due to the high diffusivity/mobility (the latter being in the range of 10^{-4} to 10^{-2} $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$ in the best solid electrolytes at fields in the order of 10^5 to 10^6 V cm^{-1}) and the small distances involved (typically 10–50 nm separation between the electrodes) [1].

The ion conductivity is given as

$$\sigma = z_e c \mu \quad (17.5)$$

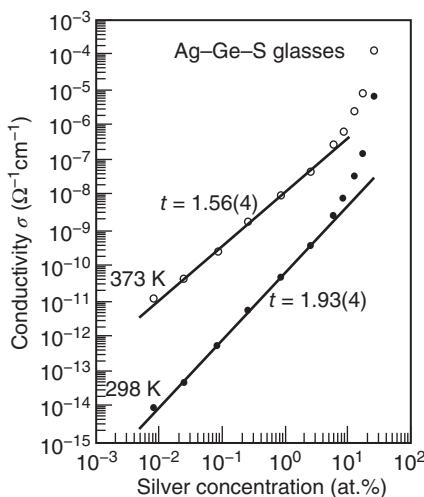


Figure 17.4 Conductivity as a function of Ag concentration in an Ag–Ge–S ternary. (From Ref. [87].)

For a singly charged ion such as Ag^+ ($z=1$), $c=10^{19} \text{ cm}^{-2}$, and $\mu=10^{-4}$ to $10^{-2} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, the ion conductivity is in the order of 10^{-4} to $10^{-2} \Omega^{-1} \text{ cm}^{-1}$, which is typical of Ag-doped chalcogenide-based glasses. It should be noted, however, that the conductivity and, interestingly, the activation energy for ion conduction of the ternary glasses are a strong function of the mobile ion concentration. Using Ag-doped Ge–S glass as an example, the room temperature conductivity in the Ag concentration range between 0.01 and 3 at% changes from 10^{-14} to around $5 \times 10^{-10} \Omega^{-1} \text{ cm}^{-1}$ and the activation energy drops from 0.9 to 0.65 eV. However, when the Ag concentration reaches tens of atomic percent, both conductivity and activation energy change more rapidly with concentration, as shown in Figure 17.4, so that the conductivity and activation energy values for “good” solid electrolytes quoted earlier are attained [1, 87]. This change in the slopes of the conductivity and activation energy curves with Ag content is a direct result of the transformation of the ternary material itself caused by the presence of such large quantities of Ag. It is also interesting to note from Figure 17.4 that a modest temperature increase from 298 to 373 K increases the conductivity (and, therefore, the mobility) by more than an order of magnitude.

As mentioned previously, ion mobility may not be the most critical factor for device performance in vertical devices, but the electrolyte conductivity plays an important role in defining the off-state resistance of the device. The off-state resistance, R_{off} , will be determined by

$$R_{\text{off}} = \frac{L}{\sigma A}, \quad (17.6)$$

where L is the device length (equal to the electrolyte thickness in a vertical device) and A is the device area. In a $0.001 \mu\text{m}^2$ device (corresponding to the 32 nm technology node) with $L = 30 \text{ nm}$ and $\sigma = 10^{-4} \Omega^{-1} \text{ cm}^{-1}$, R_{off} will be $3 \text{ G}\Omega$. In practice, it is possible to obtain a similarly high R_{off} with a much larger area device by “grading” the Ag concentration so that the conductivity is orders of magnitude lower toward the bottom of the film.

17.3

Electrochemistry of CBRAM (ECM) Cells

17.3.1

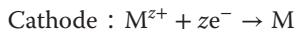
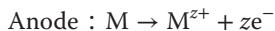
Fundamental Processes

Electrochemistry deals with the relationship between electricity and chemical change. Batteries are perhaps the most obvious example of the application of electrochemical principles; the movement of ions and the change in their oxidation state within the cell are used to release electrical energy over time. Ions not only carry charge but also have significant mass, so ion transport is a means to redistribute material in a highly controlled manner. For example, a metal atom that becomes oxidized at one location can be moved as a cation through an electrolyte by an electric field. On receiving an electron at another location, the ion is reduced and becomes a neutral metal atom. In this case, the net change in the system is the redistribution of metal using energy from an external power source. The electronics industry has benefited from “deposition electrochemistry” for many decades. Electroplating, in which metal ions in a liquid solution are reduced to create a uniform metal film on a substrate, is used in printed circuit boards and chip packages and in the processes used to make copper interconnect within integrated circuits. In these examples, physical dimensions, such as electrode spacing, are typically large and the electric fields are relatively small. The term *nanoionics* is applied when electrochemical effects occur in materials and devices with interfaces, for example, electrodes or electrochemically different material phases, that are closely spaced (a few tens of nanometers or less). In this regime, the functionality of ionic systems is different from macroscale versions. Particularly, as discussed previously, a large range of integrated-circuit-compatible thin-film materials are good ion conductors, and ion mobilities can be extremely high due to the high electric fields in nanoscale structures. In addition, electrodeposition of minuscule quantities of a highly conductive metal such as Ag or Cu can produce localized, persistent but reversible changes to electrical properties such as resistance and capacitance.

Regardless of ion mobility, a sustainable ion current will only flow if there is a source of ions at one location and a sink of ions at another; otherwise, the movement of cations away from the fixed anions would create an internal field, which will prevent current flow. In the process of electrodeposition of a conducting filament, cations are reduced by electrons from the negative

electrode (cathode); the filamentary electrodeposit is essentially a sink of ions from the electrolyte. An oxidizable positive electrode (anode) is the ion source to maintain ion concentration and overall charge neutrality. In the case of an Ag (or Cu)-ion-containing electrolyte, this oxidizable anode is merely Ag (or Cu) or a compound or alloy containing the metal. So, our most basic device consists of a solid electrolyte between an electron-supplying cathode and an ion-supplying anode, as shown in Figure 17.1. Note that in the absence of an ion source, reduction in the ions will occur only for materials allowing deviation from stoichiometry at the expense of the ion concentration in the electrolyte and electrodeposition will result in a change of ion concentration and possibly the electrolyte properties. In addition, a depleted electrolyte could allow the subsequent chemical dissolution of the filament into the matrix, which would not occur if the electrolyte was maintained near the chemical saturation point. This has important consequences for the stability of the filament and the retention of a memory device using this principle.

In the device described, the anode will oxidize when a bias is applied if the oxidation potential of the metal, M, is greater than that of the solution. As current flows in the cell, the metal ions will be reduced at the cathode. The electrode half-reactions are:



with the electrons (e^-) being supplied by the external power source. At the boundary between an electrode and an electrolyte, there is a potential difference due to charge transfer and the change of state associated with the electrode reactions. This potential difference causes a polarization effect known as the *double layer* [88], consisting of positive ions absorbed on the electrode and a net negatively charged diffuse layer, which extends over a few tens of nanometers or less in our typical solid electrolyte system. Its magnitude depends on system properties such as ion concentration, mobility, and electrode materials and must be determined for specific cell arrangements. The charge-transfer process at the electrode–electrolyte interface is described by the Butler–Volmer relationship,

$$\mathcal{J} = \mathcal{J}_0 \left[\exp \left(\frac{\alpha z_e V}{kT} \right) - \exp \left(\frac{(1-\alpha) z_e V}{kT} \right) \right], \quad (17.7)$$

where \mathcal{J}_0 is the exchange current density, α is the transfer coefficient (ranging from 0 to 1 but typically close to 0.5, depending on the nature of the reaction), and V is the potential deviation from the equilibrium voltage, that is, from an externally applied bias. This equation is representative of the reduction in the barrier height (by $\alpha z_e V$) for an ion at the interface, which leads to the increase in ion current density across this interface. Details on electrochemical fundamentals are presented in Chapter 9.

Direct experimental evidence for the involvement of electrochemical redox processes (reactions) are rarely shown; however, redox peaks in the I - V sweeps due to

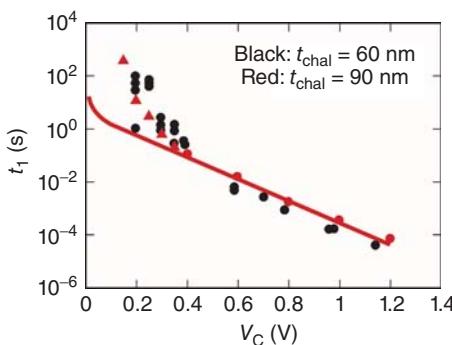


Figure 17.5 Example of initial switching time (t_1) versus applied voltage (V_c) for two electrolyte thicknesses, t_{chalc} (black = 60 nm, red = 90 nm) in a Ag/Ag-doped GeS_2/W cell. The relationship between t_1 and V_c is

exponential for $V_c > 0.4 \text{ V}$, but the switching time becomes extremely long below about 0.2 V, leading to an effective threshold voltage. (From Ref. [89].)

Faraday reactions [12, 13] and the nanobattery effect based on the differences in the redox and/or diffusion potentials [77] have recently been demonstrated. These results provide an unequivocal proof for the electrochemical nature of resistive switches.

By using the relation between the current density j and the charge Q , that is, $tjA_f = Q$, one can modify the Butler–Volmer equation, deriving it with respect to the time t required for filament to form:

$$t = \frac{Q}{jA_f} = \frac{Q}{A_f j_0 \exp(\alpha z_e V/kT)}, \quad (17.8)$$

assuming that the filament has a cross-sectional area of A_f . This means that the switching time declines exponentially with applied voltage, a phenomenon that is evident in most if not all ionic resistive memory devices, regardless of the material system employed, for example, [10, 89, 90]. However, a significant deviation from this exponential relationship is typically evident at low applied bias as illustrated in Figure 17.5.

This deviation from the exponential dependence is not unexpected as the filament will not evolve (or will grow extremely slowly) until the applied bias overcomes the overpotential associated with the double layer and other overpotential contributions such as nucleation [91], that is, the work required to create a stable nucleus on the cathode for subsequent filament growth [20, 44, 91, 92], and any series resistance components. The different rate-limiting steps can appear depending on the material properties as well as thermodynamic and the operational conditions.

A transition between nucleation mode and charge-transfer mode has been experimentally demonstrated for the $\text{Cu}/\text{Ta}_2\text{O}_5$ system where the increasing voltage is shown to minimize the influence of the nucleation on the switching time [22]. A model covering the regimes involving different rate-limiting steps is given

in the example of Ag/AgI/Pt cells [93] and discussed in more detail in Chapter 14. The work function differences in the MEM/MIM structure also contribute to the effective growth/switching threshold voltage in such devices [94]. In addition, the threshold in Cu–SiO₂ devices has been explained as being due to the difficulty at low bias for anode atoms to be oxidized and injected/diffused into SiO₂ and then be transported/reduced without being reoxidized on the cathode, thus effectively stopping filament growth [95]. Experimental results show that the minimum (near DC) threshold voltage of a Ni/Ag–Ge–Se/Ag structure is in the order of 0.18 V, whereas the threshold of W/Ag–Ge–Se/Ag is around 0.25 V. Considering higher band-gap materials, the threshold of a W/Ag–Ge–S/Ag device is closer to 0.45 V and Cu-doped SiO₂ devices typically exhibit thresholds in excess of 0.5 V. The threshold has an Arrhenius nature with an activation energy that is in the order of a few tenths of an electron volt; for the W/Ag–Ge–S/Ag structure, the threshold becomes 0.25 V even at an operating temperature of 135 °C. Once an electrodeposit has formed on the cathode, the new metal influences the overpotential/work function difference and the minimum threshold for further deposition is reduced. This effect is particularly evident in slow voltage ramp plots, where there is ample time for substantial “plating” to occur on the cathode. As the cathode material exhibits particular electrocatalytic activity toward the reduction process, the cathode metal can also significantly influence the reaction rate [96].

Before leaving the discussion on switching time dependence, two other factors should be noted. Figure 17.5 shows the results from two electrolyte thicknesses, 60 and 90 nm. In the exponential regime, the initial switching time for the thinner electrolyte is consistently less than that for the thicker material as there is less material to “bridge” with the filament. This is not evident in the nonexponential regime, where electrode effects likely dominate. In addition, the choice of electrolyte material determines the slope of the curve; if we substitute an Ag-doped Ge–Se material for the Ag-doped Ge–S, the switching time in the exponential region decreases by an order of magnitude, which is commensurate with the difference in the activation energies for filament formation in the two materials.

17.3.2

Filament Growth and Dissolution

In structures with thin electrolytes in which the nanoscale roughness of the electrode materials is significant and the fields are relatively high, nonuniformities in the ion concentration and in the electrode topography will tend to promote localized nucleation and deposition. The ions nearest the electron-supplying cathode will be reduced first, and even if multiple nuclei are initially formed, the one with the highest field and best ion supply will be favored for subsequent growth, and this will consume the local ion supply, effectively starving the other nuclei. Thus, it is likely that a “winner takes all” scenario will prevail with a single filament extending out from the cathode. Electrodeposition on the

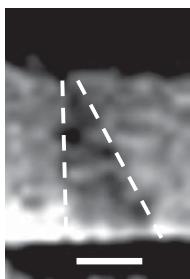


Figure 17.6 Transmission electron micrograph of an Ag filament grown in a 50 nm thick Ag–Ge–S film between W (bottom) and Ag (top) electrodes. The dashed lines show the approximate outline of the filament and the size bar is 20 nm. (Micrograph courtesy of Adesto Technologies.)

cathode does not imply that ions entering from the oxidizable anode have to travel the entire length of the electrolyte to replace those that are reduced. The ions move through the electrolyte via a coordinated motion [94]; the ion closest to the site vacated by the reduced ion will move into this position and those upstream will do similarly, each filling the vacated site of the one downstream until the last vacated space closest to the anode is taken by the incoming ion (and thereby maintaining overall charge neutrality). The tip of the growing filament moves toward the source of the ions, so the ions have less distance to travel before becoming reduced and the field increases as the growth proceeds. The conductive filament, which is physically connected to the cathode, can supply electrons for subsequent ion reduction, so the filament will harvest ions from the electrolyte not only to extend itself outward from the cathode but also to grow radially. This results in a generally conical shape to the electrodeposit (wider at the base and narrowing along its length) as the parts of the filament deposited first will have the most radial growth, as illustrated in Figure 17.6. It should be said that this is not always the case as many factors, such as nanoscale variations in the ion concentration, can disrupt this shape formation, but in any case, it is reasonable to assume that the anode end will be narrower than the cathode end.

There are however, reports on a reverse growth mode where the filament apparently grows from the anode toward the cathode [59, 97–99]. This phenomenon has been explained on both theoretical and experimental levels by a bipolar electrode effect [99]. It has been demonstrated that the particular growth mode is controlled by the combination of factors such as reduction/oxidation rate, ion mobility, local charge/mass distribution, morphological factors (e.g., nanovoids, and channels), temperature, and presence of moisture [99].

As in any deposition process, the growth rate of the electrodeposit, v , will depend on the ion flux feeding the growth, $f = j/z_e$, and the atomic density of the material being deposited, N , by

$$v = \frac{f}{N} = \frac{j}{z_e N} \quad (17.9)$$

If we simplify the filament shape so that it has an average or effective cross-sectional area A_f , so that $j = i_i/A_f$, where i_i is the ion current or, more specifically, the Faradaic current that supplies the filament growth, then the growth rate may

be given by

$$\nu = \frac{i_f M_A}{(A_f z_e N_A \rho_f)}, \quad (17.10)$$

where M_A is the atomic mass, ρ_f is the density of the filament material, and N_A is Avogadro's number. For Ag as the filament material ($M_A = 107.9 \text{ g mol}^{-1}$, $\rho_f = 10.5 \text{ g cm}^{-3}$), this equation approximates to

$$\nu = \frac{10^{-4} i_i}{A_f}. \quad (17.11)$$

For an ion current of $1 \mu\text{A}$ and an effective filament area of 100 nm^2 , the filament growth velocity is 10^2 cm s^{-1} , which is a reasonable estimate given the measured switching times of typical devices. As seen previously, the ion current density feeding the filament growth j and therefore i_i increase exponentially with the applied voltage and so the growth rate will also have this dependence. This means that the switching time, that is, the time it takes to bridge the electrolyte, will have an inverse exponential relationship with respect to the applied voltage, as we noted earlier.

An obvious question arises when faced with any exponential function – what are the practical limits? In this case, the limiting factor must be how fast the ions can replenish those that are removed from the electrolyte. The simplest way to estimate this is to consider the ion mobility (μ is in the order of $10^{-2} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for the best chalcogenide-based solid electrolytes and $10^{-6} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for oxide-based materials) and the electric field ($E = 10^5$ to 10^6 V cm^{-1}), which when multiplied give us a range for ion velocity – 10^4 cm s^{-1} at the high end and $10^{-1} \text{ cm s}^{-1}$ for low field/low mobility. So, if ion supply velocity represents a physical limit for practical devices, switching times in a 50-nm-thick electrolyte should range from 0.5 ns to 50 μs , depending on the material and the bias conditions used, which is indeed observed in practice.

As long as the potential drop along the filament is above the minimum required for the electrodeposition, radial growth continues after the initial switching/bridging event [89]. This growth follows the same rate–bias dependency as the initial growth (exponential) but is typically slower due to the collapse in the voltage following the initial bridging event [89, 100]. Note that as this growth proceeds, the resistance will continue to fall, although not quite as dramatically as during the initial switching event. All growth will cease if the (programming) current is limited so that the voltage drop decreases below the minimum necessary for electrodeposition as the on-state resistance falls. The effect of this is discussed later in this section.

One other aspect of filament growth is the amount of charge required to form the filament. For an Ag filament volume $V_f = 5 \times 10^3 \text{ nm}^3$ (which is close to the volume of electrodeposited Ag in the filament shown in Figure 17.6), Faraday's constant ($9.65 \times 10^4 \text{ C mol}^{-1}$) gives us a total charge to form this $Q = 4.70 \times 10^{-14} \text{ C}$. For an ion current of $1 \mu\text{A}$, this will take 47 ns, but if we increase the bias by a factor of 2.3 (considering the exponential relationship between voltage and current)

to take the current to $10\text{ }\mu\text{A}$, this charge will be supplied in 4.7 ns . The amount of charge supplied will determine the volume and therefore the resistance of the electrodeposit, which in turn determines the on-state resistance of the device. By ignoring the electrode and contact series resistance effects and assuming that the filament has a uniform area A_f along its length L , R_{on} is given by

$$R_{on} = \frac{L}{\sigma_f A_f}, \quad (17.12)$$

where σ_f is the conductivity of the metal. If the electrodeposited material is Ag, σ_f will be lower than the bulk conductivity due to surface and grain boundary scattering in the nanoscale feature; a reasonable value is $1.25 \times 10^4\text{ }\Omega^{-1}\text{ cm}^{-1}$ as this has been reported for thin Ag films [101]. In the example where $A_f = 100\text{ nm}^2$ and $L = 50\text{ nm}$ with $\sigma_f = 10^4\text{ }\Omega^{-1}\text{ cm}^{-1}$, R_{on} should therefore be around $400\text{ }\Omega$. In reality, the value of R_{on} will be higher than this simple analysis suggests, partly due to the nonuniform cross-sectional area involved (i.e., the resistance will be dominated by the area of the tip of the cone) and partly due to the nano-inhomogeneity of the filament. We will discuss filament morphology later in this chapter.

The erase mechanism is less well understood. A few studies regarding the kinetics of filament dissolution have been carried out [97, 102–104], and the general mechanism involves the oxidation of the filament metal. The inert electrode in device structures is typically electrochemically inert (not readily oxidizable) so that the electrodeposition process is reversible by switching the polarity of the applied bias. This is called *bipolar* programming as different polarity is used for “write” (or set) and “erase” (or reset). When the electrodeposit is made positive with respect to the original oxidizable electrode, it is dissolved via oxidation. The charge balance is maintained by reduction of these metal ions at the place where the excess metal for the electrodeposition resulted from during the growth process (the Ag or Cu electrode). The amount of charge required to oxidize the filament and plate the metal onto the electrode is essentially equivalent to that required to grow the link. Once the filament has been completely dissolved, the process self-terminates. It is important to note that it is the asymmetry of the electrodes that allows the deposition/dissolution process to be cycled repeatedly.

It should be said that there is more to the erase process than just the oxidation of the electrodeposited metal. The link must first be broken so that the reverse bias will appear between the filament and the electrode on which the material is deposited. This initiation process is most likely to occur at the narrowest part of the filament (the point of highest resistance), the tip of the cone in the most ideal case, where there is concentration of the potential [102]. Ion transport away from the dissolving filament also must be considered [102, 103]. Unlike the write process, drift and diffusion will work in tandem during erase with a reverse bias. In the case of Ag–GeS₂ electrolyte devices under a reverse bias of less than a few hundred millivolts, the relationship between the time to produce a particular change in resistance and bias appears to be similar to the exponential relationship seen in programming, but when a larger range of erase bias is considered,

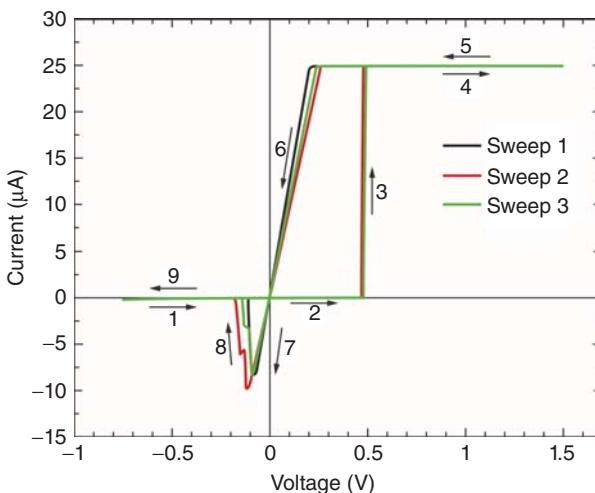


Figure 17.7 Current–voltage plot for a Cu-doped SiO_2 device with a current limit of $25\ \mu\text{A}$. The numbered arrows show the sequence and direction of the voltage sweep for each of the three sweeps shown.(From Ref. [14].)

Poole–Frenkel transport appears to be a factor [102]. The effective diameter of the conductive filament also has a significant effect on the erase process in that higher programming currents (which result in a thicker filaments) lead to longer erase times for the same bias [102]. Joule heating in the filament will also play a role in the erase, and so, depending on the experimental conditions, departures from the exponential or Poole–Frenkel relationship between erase time and voltage are also likely to be seen. Note that for comparable write and erase voltages, the maximum current flow during the erase is no more than that used for programming as it is set by the initial filament resistance (prior to breaking) [105]. This low erase current/power is one of the unique and highly desirable aspects of these resistive memory devices.

An example of switching is given in Figure 17.7, which shows a typical slow voltage ramp (taking several seconds to complete each up-down cycle) current–voltage plot for a Cu/Cu-doped SiO_2/W device [14]. The device is initially in the high-resistance off state (arrow 1), and as the applied bias is increased from $-0.75\ \text{V}$ to approximately $0.5\ \text{V}$ (arrows 1, 2), it switches to its on state by the formation of a bridging filament (arrows 3, 4). This switching threshold depends on the sweep rate due to the exponential nature of the relationship between switching time and voltage as discussed; a fast voltage ramp will result in a higher threshold [90]. The current in this example is limited to $25\ \mu\text{A}$, and this sets the on-state resistance to approximately $10\ \text{k}\Omega$, a relationship that is discussed later in this chapter. The device remains in its low-resistance state (arrows 4, 5, 6, 7) until the reverse sweep reaches $-0.2\ \text{V}$ where the conducting link is broken and the device returns to the off state (arrows 8, 9). Note that the section of the plot shown by arrows 4 and 5 is the compliance-limited region (controlled

by the source-measurement unit) and the slope of the curve shown by arrow 6 represents the on-state resistance of the device. The transition between these two regions is at 0.2 V, which represents the minimum bias for electrodeposition in this slow ramp rate example.

It should also be mentioned that, as some other resistive memory technologies, the filament can be dissolved using a bias that is the same as that used for filament formation, that is, “unipolar” programming. For our cation-based devices, this involves a current pulse that can be orders of magnitude higher than that used to program the cell, as the filament must be dissolved using Joule heating/thermal diffusion rather than a redox reaction (forward bias means that the redox reaction is actually contributing to the reinforcement of the link!). Our work has shown that for DC programming, the unipolar erase current is in the order of several milliamperes, regardless of programming current; a cell programmed at 50 μ A takes 1.6 mA in a unipolar erase process. Unipolar operation allows some functional flexibility, but it clearly is not the lowest power option.

17.3.3

Filament Morphology

The reduction in ions results in the formation of an electrodeposit in (or on) the solid electrolyte, but its precise morphology depends on a number of factors involving not only the basic principles of electrochemistry but also the structure of the hosting material and the stochastic aspects of ion transport. In the most general case, the process of electrodeposition starts with the nucleation of the new metal atom phase on the cathode. The growing feature captures ions that move close enough to be attracted to it to be reduced and attach and form the electrodeposit. As discussed previously, in small-scale structures, electrodeposition will typically be limited by charge transfer at the interface of the electrode/electrodeposit and electrolyte, but drift–diffusion processes will largely define electrodeposit morphology as these processes direct the ions to the growing feature. If the electrodeposit morphology is not constrained by the host (e.g., it forms within a liquid electrolyte or on the surface of a homogeneous solid ion conductor), the addition of atoms to the nucleus results in outward growth similar to a diffusion-limited aggregation (DLA) form [106, 107] but directed to some extent by the electric field [108]. For the high-field/high-ion-density conditions common in solid electrolyte devices, dendrite formation predominates. Dendrites have a branched nature but grow in a preferred axis defined by the electric field. Electrodeposits grown on the surface of solid electrolytes typically exhibit this dendritic form, but it should be noted that the (nano-)inhomogeneity of the material also plays a significant role in their final form [109]. Unconstrained three-dimensional structures growing out from an electrode surface can be dendritic but also have the general conical shape described in the previous section, as illustrated in Figure 17.8.

In the case of growth that is constrained by the structure of the host, the picture of filament morphology is not so clear as direct analysis of growth within the glass

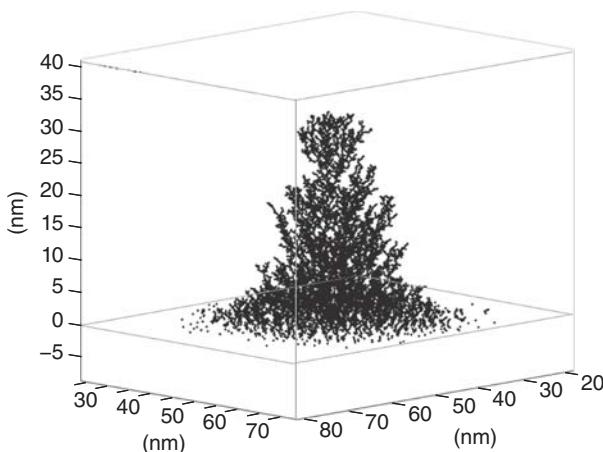


Figure 17.8 Unconstrained dendritic electrodeposit grown outward from an electrode surface. (Courtesy of S. Menzel, FZ Jülich.)

is currently unavailable, but we can make deductions from glass structure theory and also from electrical measurements. As mentioned previously, good solid electrolytes tend to possess channels and voids through which ions may readily pass. Flexibility of the glass may also be beneficial to both ion transport and filament formation with the body of the material.

Considering the structure at the atomic level of a typical solid electrolyte glass, an interesting picture emerges of where the filament may form and therefore what it might look like. Figure 17.9 illustrates the structure of $\text{Ag}_{0.12}\text{As}_{0.35}\text{S}_{0.53}$, created using density functional theory (DFT) techniques and confirmed using X-ray diffraction and neutron scattering. The 500 atom model shown in Figure 17.9a looks quite dense until the open “cavities” in the structure are highlighted – the

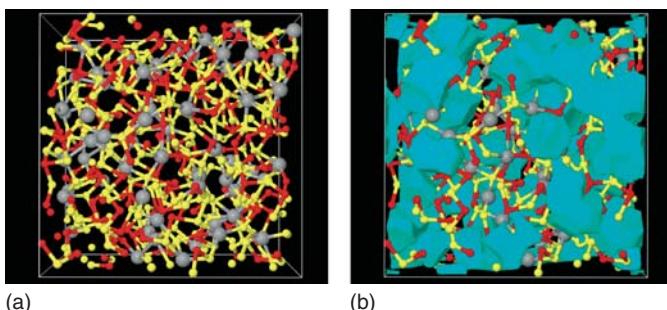


Figure 17.9 (a) 500 atom model of $\text{Ag}_{0.12}\text{As}_{0.35}\text{S}_{0.53}$ glass created using full DFT. (b) The same material showing the cavities in the glass. Here, the gray spheres denote the Ag atoms, the red spheres the As atoms, and the yellow spheres the S atoms. (Cour-

tesy J. Akola, University of Jyväskylä and Tampere Technological University, Finland; R. Jones, Jülich Research Center, Germany; T. Wagner, University of Pardubice, Czech Republic.)

light blue regions in Figure 17.9b. These large (on the atomic scale) cavities account for 24% of the volume of this sponge-like material. It is, therefore, reasonable to assume that the electrodeposits that make up the conducting filament in an on-state device form within these cavities and are thereby largely shaped by them. As mentioned previously, it is extremely difficult to confirm this using material analysis as the techniques that have sufficient resolution (e.g., high-resolution transmission electron microscopy or scanning tunneling microscopy) tend to disrupt the fine structures involved, either during sample preparation or with the probe energy or charge. We, therefore, have to infer the structure of the filament via indirect means, specifically, using electrical measurements.

If our macroscopically conical but nanoscopically dendritic filament is indeed shaped by the nanocavities in the glass, it would comprise multiple “nanowire” subcomponents, and these should have a unique electrical character. Specifically, these nanowires should be narrow enough at points for conduction to be dominated by single or small numbers of metal atoms. This would lead to quantized conduction in units of $2e^2/h = 7.75 \times 10^{-5} \Omega^{-1}$ (12.9 k Ω), where h is Planck’s constant [37]. Interestingly, strong evidence for this conductance quantization has indeed been found in a variety of systems such as Ag–Ge–S, AgI, and SiO₂ [26, 38, 110], although it should be said that this could also be due to a single atomic point contact at or near the tip of the filament [8]. This single point connection (rather than an ensemble of such elements) would probably lead to significant on-state instability [111]. In any case, it is evident that the simple calculation of R_{on} presented earlier is insufficient and that quantization will have to be taken into account.

Before leaving the topic of filament morphology, one other issue should be discussed – the modeling of on-state resistance, R_{on} . Several models have been proposed for this [26, 46, 54, 89, 105], including geometric (filament diameter), quantum (single atom conduction), or tunneling (gap) effects. However, considering the complexities of the material systems as discussed earlier in this chapter, it is likely that an amalgam of these models will be necessary to accurately describe R_{on} in both low (<12.9 k Ω) and high (>12.9 k Ω) resistance regimes. This is especially true when considering phase-separated materials or the nanowire ensemble thesis as the inhomogeneity inherent in both will lead to complexities. In the low-resistance regime, it is reasonable to assume that the quantity of metal involved will be such that continuous metal pathways dominate, even if they are composed of ensembles of nanowires. This leads to the ohmic on-state characteristics often seen in this regime. In the transition zone (around 12.9 k Ω), quantized resistance effects become most obvious. The high resistance range is more problematic. Relatively stable R_{on} values in this regime have been reported, but this implies that the diameter of the filament is less than that of the metal atoms involved, which is clearly impossible [1]. So, the conduction pathway must be discontinuous and the overall resistance dominated by the material within the breaks in the metallic portions of the filament.

No matter what the precise form of the filament is, there is no doubt that it is extremely small and perhaps could even be scaled to atomic dimensions [112]. The

diminutive size of the conducting pathway in comparison to the device area helps to explain why the on-state resistance is observed to be independent of device diameter, whereas the off-state resistance increases with decreasing area [113]. An electrodeposit this small suggests that the entire device can be shrunk to true nanoscale dimensions without compromising its operating characteristics.

17.4 Devices

17.4.1 Device Operation

The previous sections detailed the material and electrochemical foundations of cationic resistance-change devices, but we now focus on operational characteristics. Programming is relatively straightforward with a forward bias (oxidizable electrode positive with respect to the inert electrode) in excess of the electrodeposition threshold used to write the device and a reverse bias used to erase. Reading the state (off or on) of the device involves a bias that will not “disturb” or flip the current state, for example, a forward bias below the minimum required to write under normal operating conditions. For example, the device whose I - V plot is shown in Figure 17.7 would be written using a bias above 0.5 V, erased below -0.2 V, and read using 0.2 V. These represent minimum values in this near-DC biasing situation and would increase with decreasing time (e.g., for diminishing write, erase, read pulse widths) due to the exponential relationships involved.

As discussed earlier in this chapter, the off-state resistance, R_{off} , is a function of device geometry and metal concentration in the ion conductor. In the simplest terms, the on-state resistance, R_{on} , depends on how much charge is applied (the product of Faradaic current and programming time) as this will determine the volume of the electrodeposit, but as we saw, voltage is also an important factor in the programming operation. The existence of a threshold voltage leads to an interesting relationship between R_{on} and the programming current limit, I_{pr} , as illustrated in Figure 17.10 [2]. As is evident from the figure, R_{on} decreases as I_{pr}^{-1} , regardless of the electrolyte material.

This may be explained by the fact that a minimum potential difference is required to allow electrodeposition to proceed – the threshold discussed previously. Since continued electrodeposition reduces the resistance of the filament, for any current limit (I_{pr}), there exists an R_{on} at which the potential drop falls below the threshold and electrodeposition is effectively halted. R_{on} is simply related to I_{pr} by

$$R_{\text{on}} = \frac{V_t}{I_{\text{pr}}}, \quad (17.13)$$

where V_t is the threshold voltage for electrodeposition in the material system used. This is a simplification as V_t depends not only on the materials in the device

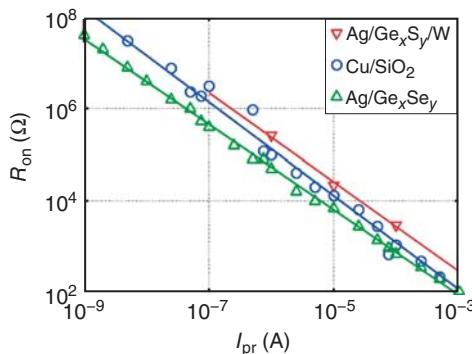


Figure 17.10 ON-state resistance R_{on} versus programming current, I_{pr} , for a variety of solid electrolytes in cationic resistive memory cells. (From Ref. [2].)

but also on how much of the anode material is electrodeposited on the cathode. So, in practice, V_t will depend on a number of experimental factors, including sweep rate/programming time, device programming history, and, considering the underlying physics of the electrodeposition process, temperature.

A more complete model for R_{on} as a function of programming current has been developed [89]. The model is based on a one transistor–one resistor (1T-1R) configuration, which is common in resistive memory arrays (discussed later), and incorporates both the initial switching event and the radial growth of the filament. The model uses a simple approach to include the temperature rise in the filament due to Joule heating. This is important because, as we saw earlier, charge transfer is temperature-activated (as is transport) and will, therefore, be enhanced if the temperature within the cell rises. By assuming that the heat flow through the metal filament with high thermal conductivity dominates, the temperature rise in the filament, ΔT_f , may be estimated by

$$\Delta T_f = \frac{V_c^2 R_{\text{th}}}{R_c}, \quad (17.14)$$

where V_c is the voltage drop across the cell, R_{th} is the thermal resistance of the filament, and R_c is the electrical resistance of the cell. We can approximate the thermal resistance by assuming that the length and area for heat flow in the filament are 10 nm and 100 nm^2 , respectively, and the thermal conductivity for the Ag metal is around $400 \text{ W m}^{-1} \text{ K}^{-1}$, so that R_{th} is roughly $2 \times 10^5 \text{ K W}^{-1}$. For a cell that has an $R_c = 1 \text{ k}\Omega$ with a voltage drop of 0.45 V (the write threshold for a typical Ag/Ag–Ge–S/W cell), the maximum temperature rise will be 40°C . Of course, the temperature rise prior to the initial switching event will depend on the instantaneous resistance of the cell, which will be very high until the conducting bridge is formed, making the temperature rise minuscule. The model also accounts for the select transistor, which as a dynamic resistive load acts to supply a constant current when the voltage drop across it pushes it into saturation. The outcome of the model is that the cell resistance (R_c) is seen to be controlled by the transistor drain

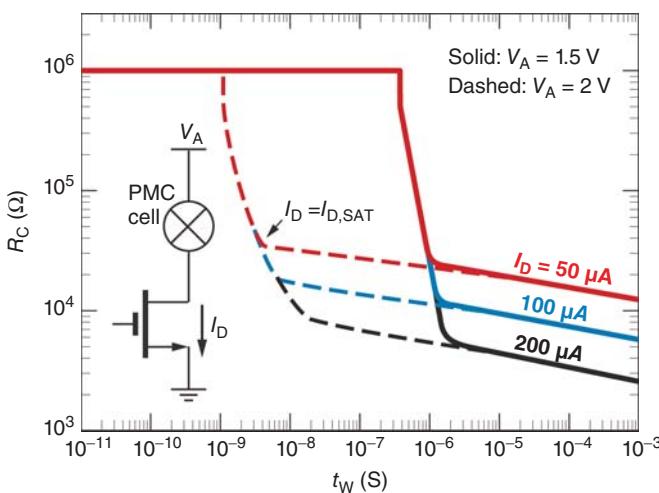


Figure 17.11 On-state resistance of an Ag/Ag-doped GeS₂/W PMC cell, showing dependence on programming current, programming pulse width, and cell voltage in a

1T-1R cell. The arrow shows where the transistor is in saturation and the drain current is constant. (From Ref. [89].)

current (I_D), programming pulse width (t_w), and power supply voltage (V_A). An example of an output from this model for an Ag/Ag–Ge–S/W device is shown in Figure 17.11. An arbitrary R_{off} of 1 MΩ was chosen, and as expected, R_c decreases with increasing drain (programming) current and the pulse width necessary to switch the device to the on state (t_w) decreases with increasing voltage. In addition, R_c falls with longer programming pulses as the filament growth is maintained (as discussed previously). For a t_w of 40 ns, a programming current of 50 μA, and a supply voltage of 2 V, corresponding to 4 pJ write energy, an R_{on} value of 35 kΩ is obtained for this Ag–Ge–S device. Note that devices based on solid electrolytes with lower activation energy for filament formation switch faster than this, for example, Ag–Ge–Se devices will switch in 20 ns at 0.6 V across the device [100], whereas the Ag–Ge–S devices require around 1 μs at this voltage.

This control of on-state resistance via programming parameters means that it is possible to create multiple discrete resistance levels to represent multiple binary digits in each physical cell. With 2^n resistance levels, we can store n logical bits, so that the off-state plus three on-state resistances (2^2) could be used to represent 00, 01, 10, and 11. Such multilevel cell (MLC) programming has already been demonstrated in cation memory arrays [102, 114, 115].

Memory performance is generally judged by parameters such as program and erase energy (power–speed product), data retention, which in this case is the stability of the on- and off-state resistances with time, and the number of program-erase cycles the device can withstand without loss of state differentiation (endurance). Program and erase energy capability in the picojoules range has been demonstrated with commercially produced CBRAM cells [116], which is below

most existing and proposed memory types. Programming currents in the tens of microampere range or higher lead to good data retention for nonvolatile memory applications [103], with considerable off-to-on resistance ratio being maintained after 10 years at 100 °C [117]. As far as endurance is concerned, published results confirm that cation-based cells can be considered well beyond the practical minimum for commercial nonvolatile technologies of 10^4 program-erase cycles, and our own work has shown that 75 nm diameter devices can be considered beyond 10^{10} cycles with no apparent degradation [118].

Ultrahigh-density memories require cells that not only are physically small but also can be operated at low voltages and currents. The physical scalability of cation-based cells has been demonstrated to the sub-20 nm regime [119]. In these nanoscale devices, the off-state resistance was above $100\text{ G}\Omega$ and a $1\text{ }\mu\text{A}$ programming current resulted in an on-state resistance of $50\text{ k}\Omega$. This high off to on resistance ratio is useful as it facilitates a simple reading scheme and permits some drift in the resistance states over time without loss of state differentiation. In addition, sub-20 nm devices based on a Cu–Te/GdO_x-layered electrolyte/dielectric structure have also been shown to demonstrate excellent switching characteristics [120]. Voltage scaling is a necessity for future devices as they must operate at voltages and currents that are commensurate with the operational goals of the ITRS at the deep nanoscale nodes. Low voltage programming of cation memory devices has previously been achieved using slow voltage sweeps [118] and small numbers of fast pulses [116]. Cycling beyond 10^6 operations has been demonstrated for Ag/Ag–Ge–S/W devices at voltages below 0.6 V with unambiguous differentiation of on and off states achieved for currents as low as $1.6\text{ }\mu\text{A}$ [121]. Extreme current scaling has also been verified with operation in the 10 pA range of Ir/Cu–SiO₂/Cu structures [122].

17.4.2

Memory Arrays

To create a working memory circuit, individual cells must be connected together to form an array. Resistive memory arrays can be active, which involves a transistor to access each memory cell during write, erase, or read operation (1T–1R), or passive, in which each location has a diode (or other nonlinear element) in series with the storage device (one-diode one-resistor (1D–1R)) to prevent sneak current paths between the cells (discussed elsewhere in this book) and thereby allow them to be isolated from one another during operation. These array schemes are introduced in Chapter 1 and described in detail in Chapter 22.

The 1T-1R format of active arrays has been demonstrated extensively [6, 114, 115, 123]. The drain of the transistor is typically connected to the inert electrode of the storage element at each location, and the cell is selected by turning the transistor on via appropriate voltages on the specific row (word line) and column (bit line). The voltage on the oxidizable electrode can be set above the voltage on the column to provide a positive bias across the cell for a write operation or set below the column voltage to reverse bias the cell to erase it. The 1T-1R format is

appropriate for high-performance applications as the speed and drive characteristics of the transistors can be used to get the most out of the dynamic range and switching capabilities of the storage elements. Since the area of a typical cation memory storage element is defined by how small we can make the metal contacts, that is, $4F^2$ where F is the minimum feature size, the area of the entire memory location in the 1T–1R scheme is really governed by the select transistor area. This is a disadvantage as far as density is concerned as transistor areas are larger than those of the storage elements, ranging from around $15F^2$ to $20F^2$ in a logic process from $8F^2$ to $6F^2$ if specialized dynamic random access memory (DRAM) transistors are used. Of course, the effective cell size will be smaller if an MLC scheme is employed; if two logical bits are stored in one physical location, the virtual cell size is half the physical area. Figure 17.12 shows an example of an integrated 1T-1R active array, fabricated using a 130 nm complementary metal-oxide-semiconductor (CMOS) process with copper multilevel interconnect [6]. The optical image shown in Figure 17.12a shows the entire chip, which contains four 256 kb subarrays and control circuitry (address decoders, sense amplifiers, etc.). Figure 17.12b is a transmission electron micrograph cross section of part of the array with the standard logic select transistors on the bottom plane and three levels of interconnect above. Each transistor is connected to an individual storage cell via Cu plugs, which link the drain to the inert electrode through the various levels of interconnect. Figure 17.12c shows a transmission electron micrograph cross section of an individual storage cell, with W, a via plug connected to the underlying Cu metallization acting as the cathode during programming. In this particular format, the ion conductor, oxidizable metal, and top level of metallization are common to several storage cells to simplify both design and processing (Figure 17.12b), but the area of the storage device is defined by the area of tungsten plug (Figure 17.12c). Note that the storage elements are built near the end of the fabrication of the circuit in a “back-end-of-line” (BEOL) process, which means that the CMOS fabrication scheme need not be changed. A further advantage is

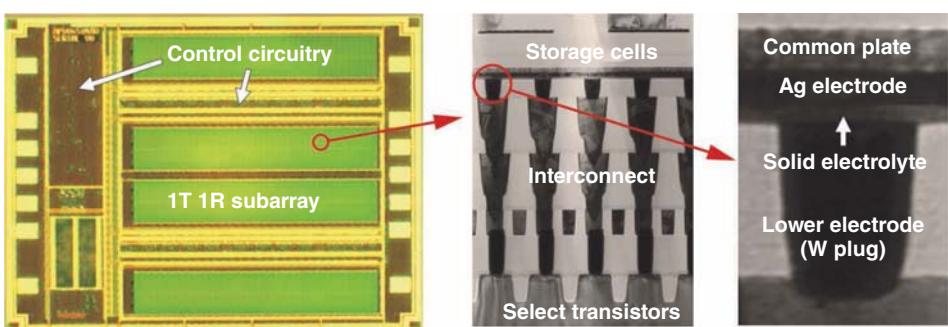


Figure 17.12 Integrated active 1T-1R array of cation memory devices. (a) Optical image of the chip. (b) Transmission electron micrograph of a cross section of the array. (c) Transmission electron micrograph of an individual storage cell. (From Ref. [9].)

that only two extra (noncritical) masks are required to define which tungsten plugs are used in storage elements and which are through connections and to define the memory stack regions. This minimizes the cost of integration and facilitates the embedding of the memory arrays with standard logic.

In the case of a 1D-1R cell, an individual cell is programmed or read by applying some fraction (e.g., half) of the programming/read voltage to its column and half to its row. All nonselected rows and columns have 0 V applied so that only the selected device has the full programming or read voltage across it. Since cation-based devices are bipolar in operation, that is, the erase requires the opposite bias to the write, the diode must have a low-voltage reverse breakdown or Zener characteristic to permit the erase but still be able to block half-program/read voltages on the array rows and columns to prevent false programming and reading. Such a Zener-diode-isolated scheme has been described [124]. Diode-isolated passive arrays can be a compact alternative to 1T-1R arrays, as long as the diode area can be minimized so that the overall physical cell area is no more than $4F^2$. This is a challenge for high-current resistive-memory technologies but not for cation/electrochemical cells as both write and erase currents are small enough to allow reasonable current densities in the diode element (a topic that is discussed elsewhere in this book). In addition to their compact nature, diode-isolated arrays can be layered to create an ultradense 3-D memory, suitable for mass storage applications. Such multilayer passive arrays have been created with deposited silicon p–n diodes in one-time programmable (OTP) antifuse memory systems [125].

One of the unique aspects of the cation cells is that the metal filament of the on state will form a Schottky diode with a doped silicon inert electrode (instead of metal). This is a simple way to achieve cell isolation as the diode is inherent in the memory structure and no additional device layers are required [126]. Interestingly, the reverse bias leakage current in these devices is proportional to the on-state resistance as both are determined by the diameter of the metal filament. An alternative approach to passive cell isolation lies in the concept of CRSs [127], which consist of two antiserial (complementary) interconnected storage structures. Only one of the two elements is programmed, and the other effectively acts as a threshold switch, blocking the sneak currents during array operation.

17.5

Technological Challenges and Future Directions

Considering the characteristics of cation-based devices described in this chapter, they appear to have the necessary attributes to fit the requirements of highly scaled, low-energy memories. Indeed, it is these advantageous attributes that have led to the commercialization of CBRAM technology, which entered the NOR Flash marketplace in 2012 and continues to make significant technological strides, particularly in the area of retention at temperatures exceeding 200 °C [128]. To attain such high-temperature retention, the switching layer

should be an oxide with a separate ion supply layer in the anode. Similar dual-layer devices have been investigated in the past [120] and are likely to grow in importance as the technology develops. The devices have also been used in novel low-voltage sensor systems utilizing CMOS that operates in the subthreshold regime, in which the memory programming and read core voltages were 0.6 and 0.35 V, respectively [129]. In addition, the technology has been shown to be radiation-tolerant [130], which opens up applications in the space, defense, and medical device sectors. The on-state resistance programmability not only allows high-density storage via MLC techniques but also allows advanced applications in the area of neuromorphic systems [119, 131]. The work on the bridging of nanometer-sized gaps between an electrode and a solid electrolyte seems to suggest that atomic-scale devices are possible and could represent the ultimate scaling of the technology [8, 112].

Even though there are many research and development efforts underway on these device structures and the materials they use, many questions regarding the physics of the technology still exist. For example, whereas the concept of the filament as an ensemble of nanowires within the “Swiss cheese” structure of the host material is reasonable, given the available configurational models and results of electrical characterization (which shows the expected quantized conduction effects), we still have no direct way of confirming this. Imaging filaments at the atomic scale within solid ion conductors has proven to be difficult as the energetic beams or the sample preparation techniques involved tend to distort the physical situation; for example, electron-beam analysis will actively grow electrodeposits by providing incidental bias and electron current for ion reduction, obscuring the original filament structure (the image in Figure 17.6 was taken at cryogenic temperatures in an attempt to reduce this effect). Tomography at the atomic level will need to be refined in order to see the true nature of the filament but until then, we can still get tantalizing glimpses of its structure and associated reliability aspects through sophisticated electrical analysis [128].

As a final comment, for any memory technology, laboratory-made structures inevitably tell a highly optimistic story as we tend to have a selection bias toward the best devices. In reality, the array is only as good as its worst cell. Endurance that is less than expected, retention failures, and stuck bits plague even the most advanced memory technologies. It may still take many years before we fully understand these issues in CBRAM-like devices, but the existence of commercial products has provided a huge volume of memory cells (trillions of cells have been produced to date), and this has generated the statistical analyses necessary to uncover the fine details of the physics of the underlying mechanisms.

Acknowledgment

The authors would like to acknowledge the support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

1. Valov, I. and Kozicki, M.N. (2013) *J. Phys. D Appl. Phys.*, **46**, 074005.
2. Valov, I., Waser, R., Jameson, J.R., and Kozicki, M.N. (2011) *Nanotechnology*, **22** (25), 254003/1–254003/22.
3. Nessel, J.A., Lee, R.Q., Mueller, C.H., Kozicki, M.N., Ren, M., and Morse, J. (2008) 2008 IEEE MTT-S International Microwave Symposium Digest, Atlanta, GA, Vols. 1–4 1312–1315.
4. Hasegawa, T., Terabe, K., Tsuruoka, T., and Aono, M. (2012) *Adv. Mater.*, **24** (2), 252–267.
5. Kozicki, M., Maroufkhani, P., and Mitkova, M. (2003) *Superlattices Microstruct.*, **34**, 467–473.
6. Gopalan, C. et al. (2011) *Solid-State Electron.*, 2nd International Memory Workshop (IMW 2010), Seoul, South Korea, **58** (1), 54–61.
7. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) *Adv. Mater.*, **21** (25–26), 2632–2663.
8. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005) *Nature*, **433** (6), 47–50.
9. Tanaka, K. (2001) Elsevier *Encyclopedia of Materials: Science and Technology*, p. 1123.
10. Sakamoto, T., Lister, K., Banno, N., Hasegawa, T., Terabe, K., and Aono, M. (2007) *Appl. Phys. Lett.*, **91**, 092110/1–092110/3.
11. Yang, B., Liang, X.F., Guo, H.X., Yin, K.B., Yin, J., and Liu, Z.G. (2008) *J. Phys. D: Appl. Phys.*, **41** (11), 115304/1–115304/5.
12. Tappertzhofen, S., Menzel, S., Valov, I., and Waser, R. (2011) *Appl. Phys. Lett.*, **99** (20), 203103/1–203103/3.
13. Tappertzhofen, S., Mündelein, H., Valov, I., and Waser, R. (2012) *Nanoscale*, **4** (10), 3040–3043.
14. Thermadam, S.P., Bhagat, S.K., Alford, T.L., Sakaguchi, Y., Kozicki, M.N., and Mitkova, M. (2010) *Thin Solid Films*, **518** (12), 3293–3298.
15. Buckwell, M., Montesi, L., Mehonic, A., Reza, O., Garnett, L., Munde, M., Hudziak, S., and Kenyon, A. J. (2015) *Phys. Status Solidi C* **12** (1–2), 211–217.
16. Morales-Masis, M., van der Molen, S.J., Fu, W.T., Hesselberth, M.B., and van Ruitenbeek, J.M. (2009) *Nanotechnology*, **20**, 095710 (6pp).
17. Nayak, A. et al. (2010) *J. Phys. Chem. Lett.*, **1** (3), 604–608.
18. Xu, Z., Bando, Y., Wang, W., Bai, X., and Golberg, D. (2010) *ACS Nano*, **4**, 2515–2522.
19. Liang, X.F., Chen, Y., Chen, L., Yin, J., and Liu, Z.G. (2007) *Appl. Phys. Lett.*, **90**, 022508 (3pp).
20. Valov, I. et al. (2012) *Nat. Mater.*, **11** (6), 530–535.
21. Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) *Nanotechnology*, **22**, 254013 (9pp).
22. Tsuruoka, T., Hasegawa, T., Valov, I., Waser, R., and Aono, M. (2013) *AIP Adv.*, **3** (3), 32114/1–32114/7.
23. Nayak, A., Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) *Nanotechnology*, **22** (23), 235201/1–235201/7.
24. Sakamoto, T., Iguchi, N., and Aono, M. (2010) *Appl. Phys. Lett.*, **96** (25), 252104/1–252104/3.
25. Kim, S.-W. and Nishi, Y. (2007), IEEE Proc. Non-volatile memory symposium, 2007, albuquerque NM, USA, 10–13 Nov. 2007, p. 76–78, ISBN 978-1-4244-1361-4.
26. Tappertzhofen, S., Valov, I., and Waser, R. (2012) *Nanotechnology*, **23** (14), 145703 (6pp).
27. Liang, X.F., Chen, Y., Shi, L., Lin, J., Yin, J., and Liu, Z.G. (2007) *J. Phys. D Appl. Phys.*, **40** (16), 4767–4770.
28. Liu, Q. et al. (2009) *IEEE Electron Device Lett.*, **30** (12), 1335–1337.
29. Lv, H.B. et al. (2008) *IEEE Electron Device Lett.*, **29** (4), 309–311.
30. Choi, S.-J. et al. (2013) *Appl. Phys. Mater. Sci. Process.*, **112** (4), 807–815.
31. Ebrahim, R., Zomorodian, A., Wu, N., and Ignatiev, A. (2013) *Thin Solid Films*, **539**, 337–341.
32. Wei, L.L., Shang, D.S., Sun, J.R., Lee, S.B., Sun, Z.G., and Shen, B.G. (2013) *Nanotechnology*, **24** (32), 325202/1–325202/7.

33. Wu, S. *et al.* (2011) *Adv. Funct. Mater.*, **21**, 93–99.
34. Haemori, M., Nagata, T. & Chikyow, T. (2009) *Appl. Phys. Express*, **2** (6), 61401/1–61401/3.
35. Hirose, Y. and Hirose, H. (1976) *J. Appl. Phys. (USA)*, **47** (6), 2767–2772.
36. Kozicki, M.N., Yun, M., Hilt, L., and Singh, A. (1999) Proceedings of Solid-State Ionic Devices, Seattle, WA, USA, 02/05/1999–07/05/1999) *J. Electrochem. Soc.*, 298–309.
37. Jameson, J.R. *et al.* (2012) *IEEE Electron Device Lett.*, **33** (2), 257–259.
38. Guo, H.X. *et al.* (2007) *Appl. Phys. Lett.*, **91** (24), 243513/1–243513/3.
39. Tsunoda, K., Fukuzumi, Y., Jameson, J.R., Wang, Z., Griffin, P.B., and Nishi, Y. (2007) *Appl. Phys. Lett. (USA)*, **90** (11), 113501/1–113501/3.
40. Kozicki, M.N. and Mitkova, M. (2006) *J. Non-Cryst. Solids*, **352**, 567–577.
41. Bruchhaus, R., Honal, M., Symanczyk, R., and Kund, M. (2009) *J. Electrochem. Soc.*, **156** (9), H729–H733.
42. Mitkova, M. and Kozicki, M.N. (2002) *J. Non-Cryst. Solids*, **209**–302, 1023–1027.
43. Rahaman, S.Z. *et al.* (2010) *Elec-trochim. Solid-State Lett.*, **13** (5), H159–H162.
44. Soni, R. *et al.* (2011) *J. Appl. Phys.*, **110** (5), 54509/1–54509/10.
45. Li, Y. *et al.* (2010) *Phys. Status Solidi RRL*, **4** (5–6), 124–126.
46. Gopalan, C., Kozicki, M.N., Bhagat, S., Thermadam, S.C.P., Alford, T.L., and Mitkova, M. (2007) *J. Non Cryst. Solids*, **353** (18–21), 1844–1848.
47. Dong, C.Y. *et al.* (2011) *Appl. Phys. Lett.*, **98** (7), 72107/1–72107/3.
48. Rahaman, S.Z. *et al.* (2012) *Jpn. J. Appl. Phys.*, **51**, 04DD11.
49. Sacchetto, D., Zervas, M., Temiz, Y., De Micheli, G., and Leblebici, Y. (2012) *IEEE Trans. Nanotechnol.*, **11** (1), 8–11.
50. van den Hurk, J., Valov, I., and Waser, R. (2012) *Thin Solid Films*, **527**, 299–302.
51. Kozicki, M.N., Gopalan, C., Balakrishnan, M., and Mitkova, M. (2006) *IEEE Trans. Nanotechnol.*, **5** (5), 535–544.
52. Schindler, C., Valov, I., and Waser, R. (2009) *Phys. Chem. Chem. Phys.*, **11** (28), 5974–5979.
53. Liu, Q. *et al.* (2009) *Appl. Phys. Lett.*, **95** (2), 23501/1–23501/3.
54. Choi, S. *et al.* (2011) *Adv. Mater.*, **23** (29), 3272–3277.
55. Kever, T., Boettger, U., Schindler, C., and Waser, R. (2007) *Appl. Phys. Lett.*, **91** (8), 083506/1–083506/3.
56. Oyamada, T., Tanaka, H., Matsushige, K., Sasabe, H., and Adachi, C. (2003) *Appl. Phys. Lett. (USA)*, **83** (6), 1252–1254.
57. Hajto, J., Owen, A., Snell, A., Le Comber, P., and Rose, M. (1991) *Philos. Mag. B*, **63**, 349–369.
58. Kim, K.-H. *et al.* (2011) *Nano Lett.*, **12**, 389–395.
59. Yang, Y., Gao, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012) *Nat. Commun.*, **3**, 732/1–732/8.
60. Rickert, H. (1982) *Electrochemistry of Solids*, Springer.
61. Kartini, E., Collins, M., Indayaningsih, N., and Svensson, E. (2000) *Solid State Ion.*, **138**, 115–121.
62. Ogawa, H. and Kobayashi, M. (2002) *Solid State Ion.*, **148**, 211–217.
63. Shahi, K. (1977) *Phys. Status Solidi A*, **41**, 11–44.
64. Rennie, J.H.S. and Elliott, S.R. (1987) *J. Non Cryst. Solids*, **97**–98, 1239–1242.
65. Kolobov, A.V., Elliott, S.R., and Taguirdzhanov, M.A. (1990) *Philos. Mag. B*, **61**, 859–865.
66. Tanaka, K., Miyamoto, Y., Itoh, M., and Bychkov, E. (1999) *Phys. Status Solidi A*, **173**, 317–322.
67. Anderson, O. and Stuart, D. (1954) *J. Am. Ceram. Soc.*, **37**, 573–580.
68. Rao, K.J. (2002) *Structural Chemistry of Glasses*, Elsevier.
69. Elliott, S.R. (1991) *Chalcogenide Glasses*, VCH Publishers, Weinheim.
70. Tafen, D., Drabold, D.A., and Mitkova, M. (2005) *Phys. Rev. B*, **72** (5), 54206/1–54206/9.
71. Drabold, D.A. (2009) *Eur. Phys. J. B*, **68** (1), 1–21.
72. Cho, D.-Y., Valov, I., van den Hurk, J., Tappertzhofen, S., and Waser, R. (2012) *Adv. Mater.*, **24** (33), 4552–4556.

73. Oldale, J., Rennie, J., and Elliott, S. (1988) *Thin Solid Films*, **164**, 467–473.
74. Horton, J.H., Peat, K.L., and Lambert, R.M. (1993) *J. Phys. Condens. Matter*, **5**, 9037–9048.
75. Horton, J.H., Hardacre, C., Baddeley, C.J., Moggridge, G.D., Ormerod, R.M., and Lambert, R.M. (1995) *Phys. Rev. B*, **52**, 2054–2062.
76. van den Hurk, J. *et al.* (2014) *Phys. Chem. Chem. Phys.*, **16** (34), 18217–18225.
77. Valov, I. *et al.* (2013) *Nat. Commun.*, **4**, 1771/1–1771/9.
78. Tappertzhofen, S., Linn, E., Böttger, U., Waser, R., and Valov, I. (2014) *IEEE Electron Device Lett.*, **35** (2), 208–210.
79. van den Hurk, J., Linn, E., Zhang, H., Waser, R., and Valov, I. (2014) *Nanotechnology*, **25** (42), 425202/1–425202/6.
80. Kozicki, M.N., Mitkova, M., Zhu, J., and Park, M. (2002) *Microelectron. Eng. (Netherlands)*, **63** (1–3), 155–159.
81. Dallaporta, H., Liehr, M., and Lewis, J. (1990) *Phys. Rev. B: Condens. Matter*, **41**, 5075.
82. Cho, D.Y., Tappertzhofen, S., Waser, R., and Valov, I. (2013) *Nanoscale*, **5** (5), 1781–1784.
83. Pan, J. *et al.* (2008) *Vacuum*, **83**, 641–644.
84. Tappertzhofen, S., Valov, I., Tsuruoka, T., Hasegawa, T., Waser, R., and Aono, M. (2013) *ACS Nano*, **7** (7), 6396–6402.
85. Tsuruoka, T., Terabe, K., Hasegawa, T., Valov, I., Waser, R., and Aono, M. (2012) *Adv. Funct. Mater.*, **22** (1), 70–77.
86. Tappertzhofen, S., Hempel, M., Valov, I., and Waser, R. (2011) Materials Research Society Symposium Proceedings, p. 1330–1336.
87. Bychkov, E. (2000) *Solid State Ion.*, **136**, 1111–1118.
88. Delahay, P. (1965) *Double Layer and Electrode Kinetics*, John Wiley & Sons Inc..
89. Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A.L., and Kozicki, M.N. (2009) *IEEE Trans. Electron Devices*, **56** (5), 1040–1047.
90. Schindler, C., Staikov, G., and Waser, R. (2009) *Appl. Phys. Lett.*, **94** (7), 072109/1–072109/3.
91. Valov, I. and Staikov, G. (2013) *J. Solid State Electrochem.*, **17** (2), 365–371.
92. Milchev, A., Stoyanov, S., and Kaischev, R. (1974) *Thin Solid Films*, **22**, 255–265.
93. Menzel, S., Tappertzhofen, S., Waser, R., and Valov, I. (2013) *Phys. Chem. Chem. Phys.*, **15** (18), 6945–6952.
94. Jameson, J.R., Gilbert, N., Koushan, F., Saenz, J., Wang, J., Hollmer, S., Kozicki, M. *et al.* (2012) *Appl. Phys. Lett.*, **100** (2), 23505/1–23505/4.
95. Pan, F., Yin, S., and Subramanian, V. (2011) *IEEE Electron Device Lett.*, **32** (7), 949–951.
96. Tappertzhofen, S., Waser, R., and Valov, I. (2014) *ChemElectroChem*, **1**, 1287–1292.
97. Liu, Q. *et al.* (2012) *Adv. Mater.*, **24** (14), 1844–1849.
98. Gao, S., Song, C., Chen, C., Zeng, F., and Pan, F. (2012) *J. Phys. Chem. C*, **116** (33), 17955–17959.
99. Yang, Y. *et al.* (2014) *Nat. Commun.*, **5**, 4232/1–4232/9.
100. Gilbert, N.E., Gopalan, C., and Kozicki, M.N. (2005) *Solid-State Electron. (UK)*, **49** (11), 1813–1819.
101. Logeeswaran, V., Kobayashi, N. P., Islam, M. S., Wu, W., Chaturvedi, P., Fang, N. X., Wang, S. Y. Williams, R. S. *et al.* (2008) *Nano Lett.*, **9**, 178–182.
102. Kamalanathan, D., Russo, U., Ielmini, D., and Kozicki, M.N. (2009) *IEEE Electron Device Lett.*, **30** (5), 553–555.
103. Symanczyk, R., Bruchhaus, R., Dittrich, R., and Kund, M. (2009) *IEEE Electron Device Lett.*, **30** (8), 876–878.
104. Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2010) *Nanotechnology*, **21** (42), 425205/1–425205/8.
105. Choi, S., Ambrogio, S., Balatti, S., Nardi, F., and Ielmini, D. (2012) 4th IEEE International Memory Workshop (IMW), 2012.
106. Meakin, P. (1983) *Phys. Rev. A: At. Mol. Opt. Phys.*, **27**, 604.
107. Witten, T.A. and Sander, L.M. (1981) *Phys. Rev. Lett.*, **47** (19), 1400–1403.
108. Sawada, Y., Dougherty, A., and Gollub, J.P. (1986) *Phys. Rev. Lett.*, **56**, 1260.

109. Mitkova, M.I., Kozicki, M.N., and Aberouette, J.P. (2003) *J. Non-Cryst. Solids*, **326&327**, 425–429.
110. Tappertzhofen, S., Linn, E., Menzel, S., Waser, R., and Valov, I. (2014) IEEE Silicon Nanoelectronic Workshop (CFP14SNW-PRT), pp. 65–66.
111. Valov, I. (2014) *ChemElectroChem*, **1**, 26–36.
112. Zhirnov, V.V., Meade, R., Cavin, R.K., and Sandhu, G. (2011) *Nanotechnology*, **22** (25), 254027/1–254027/21.
113. Symanczyk, R., Balakrishnan, M., Gopalan, C., Happ, T., Kozicki, M. N., Kund, M., Mikolajick, T., Mitkova, M., Park, M., Pinnow, C., Robertson, J., Ufert, K. *et al.* (2003) Proceedings of the Non-Volatile Memory Technology Symposium 17.
114. Gilbert, N.E. and Kozicki, M.N. (2007) *IEEE J. Solid-State Circuits*, **42** (6), 1383–1391.
115. Schrögmeier, P. (2007) VLSI Circuits Symposium, pp. 18–23.
116. Derhacobian, N., Hollmer, S.C., Gilbert, N., and Kozicki, M.N. (2010) *Proc. IEEE*, **98** (2), 283–298.
117. Symanczyk, R., Dittrich, R., Keller, J., Kund, M., Mueller, G., Ruf, B., Albareda, P.-H., Bournat, S., Bouteille, L., Duch, A. *et al.* (2007) Proceedings of NVMTS pp. 71–75.
118. Kozicki, M.N., Park, M., and Mitkova, M. (2005) *IEEE Trans. Nanotechnol. (USA)*, **4** (3), 331–338.
119. Yu, S. and Wong, H.-S. (2011) *IEEE Trans. Electron Devices*, **58** (5), 1352–1360.
120. Aratani, K. *et al.* (2007) IEDM Technical Digest, 2007, p. 783–786.
121. Kamalanathan, D., Akhavan, A., and Kozicki, M.N. (2011) *Nanotechnology*, **22**, 254017/1–254017/6.
122. Schindler, C., Weides, M., Kozicki, M.N., and Waser, R. (2008) *Appl. Phys. Lett.*, **92** (12), 122910–122913.
123. Schroegmeier, P., Angerbatier, M., Dietrich, S., Ivanov, M., Hoenigschmid, H., Liaw, C., Markert, M., Symanczyk, R., Altimirrie, L., Bournat, S., Mueller, G. *et al.* (2007) 2007 Symposium On Vlsi Circuits, Digest of Technical Papers, pp. 186–187.
124. Toda, H. (2009) US Patent 7606059. Three-dimensional programmable resistance memory device with a read/write circuit stacked under a memory cell array.
125. Johnson, M. *et al.* (2003) *IEEE J. Solid-State Circuits*, **38** (11), 1920–1928.
126. Puthenthaleradam, S.C., Schroder, D.K., and Kozicki, M.N. (2011) *Appl. Phys. A: Mater. Sci. Process.*, **102** (4), 817–826.
127. Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010) *Nat. Mater.*, **9** (5), 403–406.
128. Jameson, J. c (2013) IEEE International, Electron Devices Meeting (IEDM), 2013.
129. Gilbert, N., Zhang, Y., Dinh, J., Calhoun, B., and Hollmer, S. (2013) 2013 Symposium on VLSI Circuits, pp. C204–C205.
130. Dandamudi, P., Barnaby, H., Kozicki, M., Gonzalez-Velo, Y., and Holbert, K. (2013) 2013 14th European Conference on Radiation and Its Effects on Components and Systems (RADECS), 4 pp.
131. Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J.K., and Aono, M. (2011) *Nat. Mater.*, **10**, 591–595.

18

Atomic Switches

Kazuya Terabe, Tohru Tsuruoka, Tsuyoshi Hasegawa, Alpana Nayak, Takeo Ohno, Tomonobu Nakayama, and Masakazu Aono

18.1

Introduction

The last century was in a sense a century of solid electronic devices. In 1947, the first transistor was invented by Bardeen, Brattain, and Shockley, and thanks to Kilby's idea of integrating such transistors in 1958, integrated circuit (IC) technology experienced rapid development with the aid of microfabrication technology, leading to today's revolutionary information and communication technology. In IC technology, as typified by metal-oxide-semiconductor (MOS) transistor technology, the movement of electrons (or holes) is controlled by electric field, and the atoms (or ions) constructing the stage in which the electrons (or holes) move are stationary except for thermal vibration.

At the dawn of this century (2001), some of the authors developed a novel electronic device operated by a totally different mechanism in which the movement of atoms (or ions) is controlled by electric potential [1]. The device was later named the "atomic switch" [2]. Compared with conventional MOS transistor switches, the atomic switch is a much-awaited nonvolatile switch, smaller in size, simpler in structure, and lower in power consumption due to the lower resistance of the ON state, while remaining comparable in switching speed (see review articles [3–6]). Interestingly, the conductance of the atomic switch can be controlled not only at the ON and OFF binary values but also at various quantized conductance values n ($2e^2/h$) ($n = 1, 2, \dots$). Moreover, it has been revealed that the atomic switch closely resembles the synapses of the human brain in functionality [7]. In 2013, random networks of tens of millions of atomic switches have been constructed and have been found to exhibit unexpected novel properties of great interest [8].

This chapter on "atomic switches" is organized as follows. In Section 18.1, a brief history of the development of the atomic switch and its basic working principles are described. The various types of atomic switches are discussed specifically in detail in Sections 18.2–18.4. A summary is provided in Section 18.5.

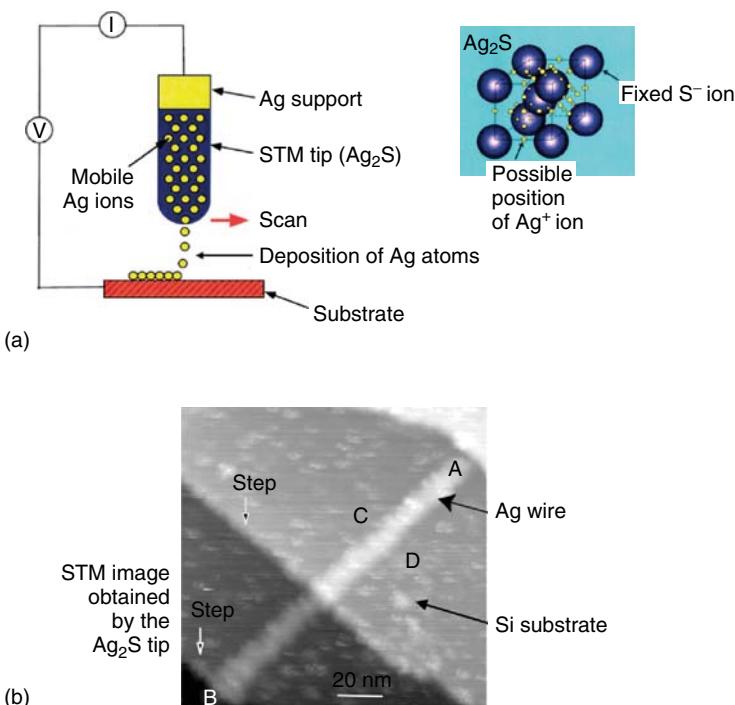


Figure 18.1 (a) Schema of nanofountain pen with Ag ink realized using metal atoms transferred from the STM tip of ionic/electronic mixed conductor Ag_2S , with crystal structure illustrated. (b) An Ag wire

formed by the Ag atom deposition from the scanned Ag_2S -STM tip. (Reproduced with permission from Ref. [9]. © (2002), AIP Publishing LLC.)

18.1.1

Brief History of the Development of the Atomic Switch

The development of the atomic switch was serendipitous in a sense. In earlier studies by some of the authors related to the development of novel nanofabrication methods, the new method illustrated in Figure 18.1a was tested; namely, a needle-like single crystal of Ag_2S was grown and used as the probe tip of a scanning tunneling microscope (STM). Since Ag_2S is a solid electrolyte in which Ag^+ ions are moving around as if in a liquid in a rigid lattice made of S^- ions (see the inset of Figure 18.1a), if a positive bias voltage is applied to the STM tip, Ag^+ ions in the Ag_2S tip will be extracted from the tip and deposited onto a counter electrode (substrate) one by one along the scan of the STM tip parallel to the substrate. Such fabrication of a nanowire made of Ag atoms by this “Nanofountain pen with Ag ink” was actually possible, as shown in Figure 18.1b [9–11].

However, it was found that when a relatively low positive bias voltage was applied to the Ag_2S tip, Ag^+ ions in the Ag_2S tip were not deposited onto the substrate but rather formed a cluster of Ag atoms at the apex of the tip

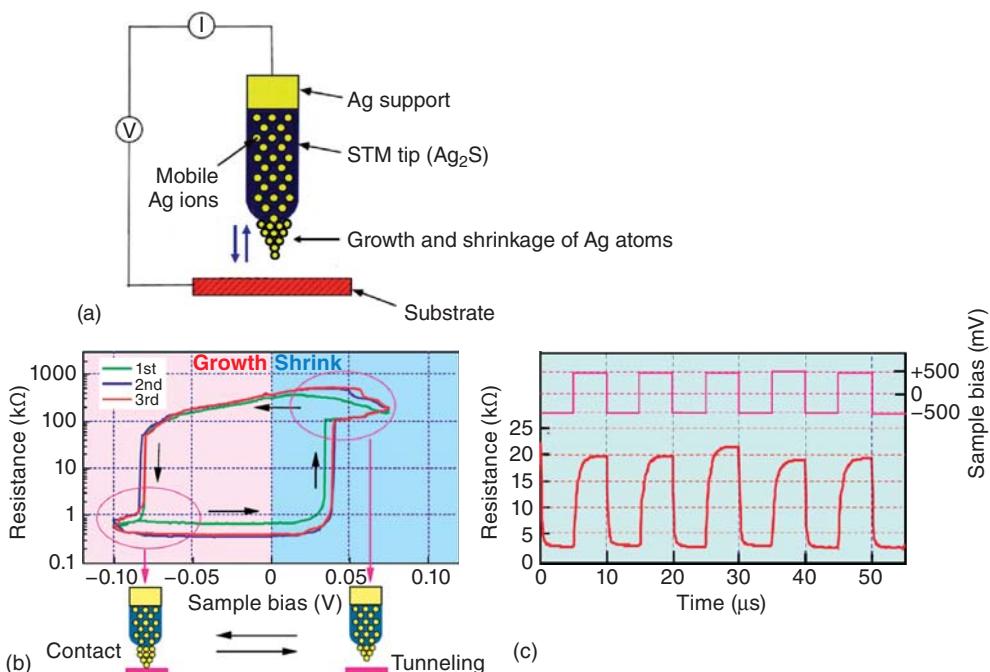


Figure 18.2 (a) Schema of growth and shrinkage of Ag atoms at the apex of an Ag₂S tip operated by applying positive and negative bias to the tip, respectively. (b) Electrical switching hysteresis achieved by

cyclic formation and destruction of the Ag cluster bridge between the Ag₂S tip and a Pt substrate. (c) Switching operation at 1 MHz [1].

(Figure 18.2a). Interestingly, when the polarity of the bias voltage was reversed, the Ag atoms of the cluster dissolved into the Ag₂S tip reversibly (Figure 18.2a). We immediately came up with the idea of using these reversible processes for atomic-scale electrical switching. In order to examine this idea, we cut off the feedback circuit of the STM so as to keep the height of the STM tip relative to the substrate constant, and the bias voltage applied to the STM tip was varied between positive and negative values repeatedly. As shown in Figure 18.2b, a beautiful switching hysteresis was observed. This switching was stable even at 1 MHz, as depicted in Figure 18.2c [1] (as we discuss in Section 18.3, ~1 GHz switching is possible in recent atomic switches).

18.1.2

Basic Working Principle of the Atomic Switch

The mechanism of the ON/OFF switching observed in Figure 18.2a,b is illustrated in Figure 18.3. The switching occurs between ON and OFF states (Figure 18.3a,b, respectively) reversibly. By applying a positive bias voltage to the Ag₂S STM tip, the electron tunneling current flows from the tip to the counter electrode through

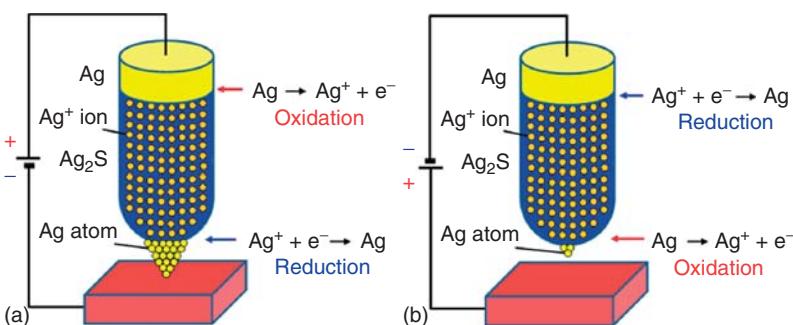


Figure 18.3 Principle of the switching caused by a solid electrochemical reaction using an $\text{Ag}/\text{Ag}_2\text{S}$ STM tip and a Pt substrate. (a) Switching-ON process and (b) switching-OFF process.

a certain point closest to the counter electrode. Therefore, Ag^+ ions near that point are partly neutralized into Ag atoms: $\text{Ag}^+ + \text{e}^- \rightarrow \text{Ag}$. Such neutral Ag atoms are unstable in Ag_2S and are, therefore, pushed out of the Ag_2S tip, forming a cluster of Ag atoms at the tip's apex (Figure 18.3a); when the Ag cluster touches the counter electrode, switching-ON occurs (Figure 18.3a). Under this condition, if the polarity of the bias voltage applied to the Ag_2S STM tip is reversed, the electron tunneling current flows from the counter electrode to the tip through the Ag cluster. Therefore, Ag atoms in the Ag cluster are partly ionized into Ag^+ ions: $\text{Ag} \rightarrow \text{Ag}^+ + \text{e}^-$. Such ionized Ag^+ ions are unstable in the Ag cluster and are, therefore, dissolved into the Ag_2S tip (Figure 18.3b), resulting in the OFF state.

The atomic switch shown in Figures 18.2 and 18.3, which utilizes an STM of macroscopic size, has accomplished its historical mission. We can now fabricate various atomic switches of much smaller sizes ranging from the micrometer to the nanometer.

Such atomic switches can be classified into “gap-type” and “gapless-type” (discussed in Sections 18.2 and 18.3, respectively). The gap-type atomic switches have a vacuum gap between two electrodes (as in Figures 18.2 and 18.3) or a gap filled with a soft and inert polymer, whereas the gapless-type atomic switches have neither the vacuum gap nor the soft and inert polymer gap. In the latter case, the gap is rather filled with an electronically insulating material such as a solid electrolyte at a stoichiometric composition (Ag_2S , Cu_2S , etc.), a polymer electrolyte (polyethylene oxide with mobile ions, etc.), or a metal oxide (Ta_2O_5 , etc.) that acts as an ion-transport medium. It should be noted that in both the gap-type and gapless-type atomic switches, the basic working principle is the creation and annihilation of a conducting path made of *metal atoms*. How many metal atoms are incorporated in the conducting path? Depending on the structure and size of the atomic switch, the number of metal atoms incorporated in the conducting path ranges widely from one to several thousand or even more. Even when a large number of metal atoms are incorporated in the conducting path, actual switching often occurs by the movement of several atoms at the apex of the conducting path. In

such a case, the conductance of the atomic switch can be controlled at one of the quantized values $n (2e^2/h)$ ($n = 1, 2, \dots$).

At the end of this introductory Section 18.1, the authors recommend the readers go through the last two paragraphs of Section 18.5, where the uniqueness of the atomic switch among various related devices is clarified.

18.2

Gap-Type Atomic Switches

The gap-type atomic switches consist of metal electrode A/ionic conductor material/vacuum gap/metal electrode B structures in which metal electrodes A and B function as an electrochemically active electrode (e.g., Ag or Cu) and an inert metal electrode (e.g., Pt or Au), respectively. Mixed conductors such as Ag_2S [12, 13] or Cu_2S , in which not only metal ions (e.g., Ag^+ ions or Cu^+ ions) but also electrons migrate, have been used as ionic conductor materials. The first switch to be developed was the gap-type switch that had a nanometer-sized gap between the inert metal electrode and ionic conductor. The basic characteristics and device structures are described.

18.2.1

Switching Time

The switching event in this sort of atomic switch takes place in a nanometer-sized gap between the inert metal electrode and an ionic conductor material. Apart from the electrochemical reaction, the switching kinetics involves a complex interplay of the chemistry of ionic conductor materials, the process of nucleation, and the electric field in the nanogap. The literature has investigated the switching times of gap-type atomic switches made from different ionic conductor materials such as Cu_2S [14–16], Ag_2S [17, 18], and RbAg_4I_5 [19] in order to identify their rate limiting steps.

A gap-type configuration can be realized with STM by placing its Pt tip above the surface of an ionic conductor material under a constant tunneling current, as shown schematically in Figure 18.4a for a Cu_2S system [17]. For the switching operation, a bias is applied between the tip and surface after fixing the position of the tip by disabling the feedback loop. Most of the potential drop occurs in the nanogap because its resistance is much higher than that of the Cu_2S , and the potential drop in the Cu_2S ($V_{\text{Cu}_2\text{S}}$) mainly occurs at the surface and interface. Electrochemical formation of a Cu cluster on the surface of the Cu_2S directly below the tip bridges the nanogap. The switching time (t_{sw}) is defined as the time interval in which the resistance between the tip and Cu_2S decreases from the initial OFF-resistance to $12.9 \text{ k}\Omega$, after applying a switching bias (V_{sw}). The initial OFF-resistance is that of the tunnel gap before applying a V_{sw} . The value of $12.9 \text{ k}\Omega$ is considered to be the ON-resistance corresponding to a single atomic contact.

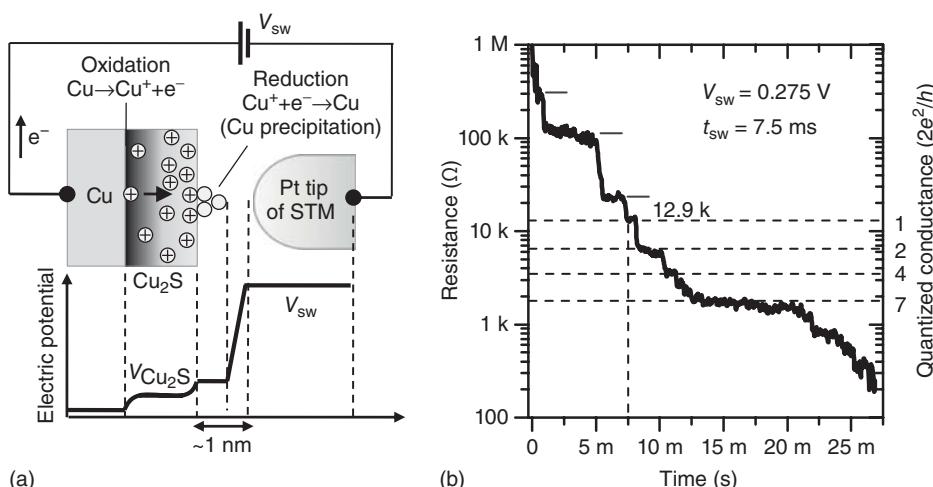


Figure 18.4 (a) Schematic representation of a Cu_2S gap-type atomic switch realized using a scanning tunneling microscope and the corresponding electric potential distribution induced by an applied bias V_{sw} for switching. The bias voltage effectively applied to the Cu_2S material is indicated

by $V_{\text{Cu}_2\text{S}}$. (b) Typical example of the change in resistance during the process of switching from an initial OFF-resistance of $1 \text{ M}\Omega$ and a V_{sw} of 0.275 V [14]. (Reprinted (adapted) with permission from Ref. [17]. © (2010) American Chemical Society.)

Figure 18.4b shows a typical switching operation of a Cu_2S gap-type atomic switch for a V_{sw} of 0.275 V and an initial OFF-resistance of $1 \text{ M}\Omega$ [14]. A stepwise decrease in resistance can be observed, and the steps after the point contact correspond to the quantized units of $1/NG_0$ ($G_0 = 2e^2/h = 1/12.9 \text{ k}\Omega$), where N indicates the number of quantized conductance steps. The time taken to reach the point contact after the bias application gave a t_{sw} of 7.5 ms in this case. The blue bars in the figure indicate that there are at least three steps before reaching the point contact. The number of steps suggests that at least three atomic layers were needed to bridge the nanogap in this case. Since the rate of electrochemical reaction increases exponentially with increasing bias, the values of t_{sw} should depend on V_{sw} . Figure 18.5a shows the bias dependence of t_{sw} of a Cu_2S gap-type atomic switch at room temperature for initial OFF-resistances of 1 , 10 , and $100 \text{ M}\Omega$, respectively [14]. As can be observed, t_{sw} decreased exponentially with increasing V_{sw} ($t_{\text{sw}} \propto \exp^{-\rho V_{\text{sw}}}$) and exhibited two distinct exponential decay components, β_1 and β_2 , indicating that more than one process participates in controlling the rate of switching. For a given OFF-resistance condition, the exponent β_1 of the lower bias range is larger than the exponent β_2 of the higher bias range. Furthermore, the switching became comparatively faster by decreasing the initial OFF-resistance. As reported in the literature, different ionic conductor systems have different t_{sw} values, which can be attributed to material-based differences in the activation energies for the chemical reaction [16].

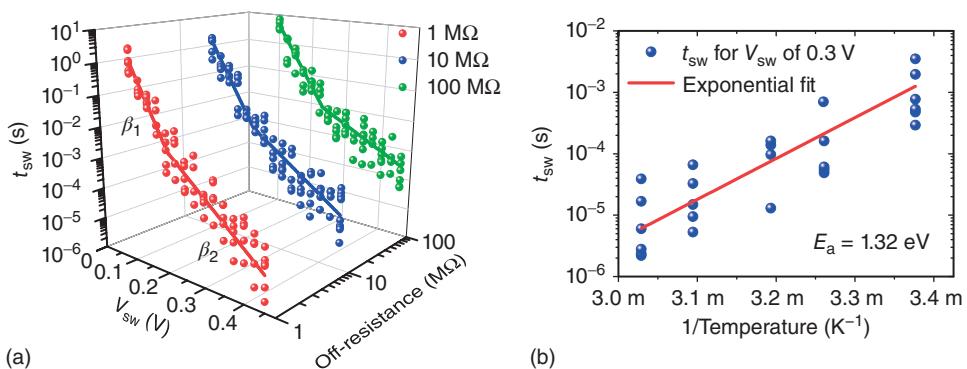


Figure 18.5 (a) Switching time t_{sw} as a function of bias voltage V_{sw} measured at room temperature for initial OFF-resistances of 1, 10, and 100 MΩ. The exponential fits are shown by solid lines and the exponents β_1 and β_2 belong to the lower and higher

bias ranges, respectively. (b) Temperature dependence of t_{sw} for an initial OFF-resistance of 10 MΩ and a V_{sw} of 0.3 V. The activation energy (E_a) is determined from the slope of the exponential fit [14].

Upon increasing the temperature of the Cu₂S material, t_{sw} decreased exponentially, following an Arrhenius relation ($t_{sw} \propto \exp^{-E_a/k_B T}$), where E_a is the activation energy for switching, k_B is Boltzmann's constant, and T is absolute temperature. Even a small increase in temperature of $\sim 10^\circ\text{C}$ decreased t_{sw} by about an order of magnitude. Figure 18.5b plots the temperature dependence of t_{sw} for a V_{sw} of 0.3 V and an initial OFF-resistance of 10 MΩ [14]. The slope of the Arrhenius plot yielded an E_a value of 1.32 eV. Similar measurements showed that E_a was independent of the initial OFF-resistance and V_{sw} .

18.2.2 Electrochemical Process

The switching process of a Cu₂S gap-type atomic switch is schematically illustrated in Figure 18.6 [14]. In the equilibrium condition without any bias application, the Cu substrate, on which the Cu₂S is formed, makes the electrochemical potentials of the Cu⁺ ions in the Cu₂S and the Cu atoms on the surface equal. Consequently, the activation energies for reduction, E_r (from Cu⁺ to Cu), and for oxidation, E_o (from Cu to Cu⁺), are equal (Figure 18.6a). Application of a positive bias voltage to the Cu₂S causes Cu⁺ ions in the Cu₂S to diffuse toward the surface, which increases the concentration of Cu⁺ ions at the subsurface. Since the electrochemical potential is described as $\mu = \mu_0 + F\varphi + RT \ln \gamma C$ (μ_0 is the chemical potential at $\gamma C = 1$, F , φ , R , T , γ , and C are the Faraday constant, electrical potential, molar gas constant, temperature, activity coefficient, and concentration of the metal cation, respectively), the electrochemical potential of the Cu⁺ ions at the subsurface becomes higher (Figure 18.6b). Consequently, E_r becomes smaller than E_o , promoting the precipitation of Cu atoms, and the switch is turned on. On the contrary, application of a negative bias voltage to the Cu₂S electrode makes Cu⁺

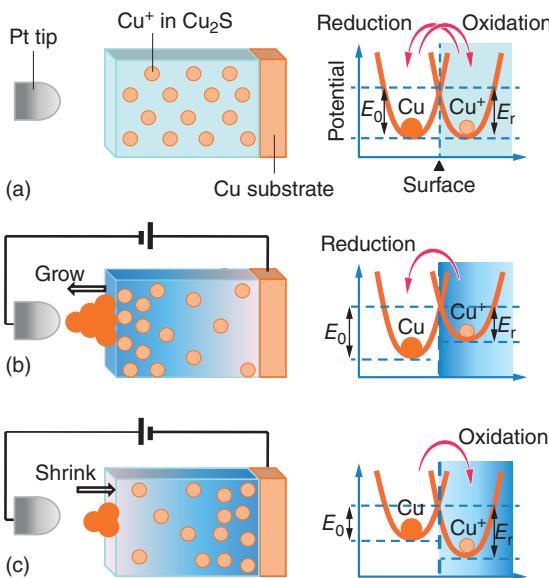


Figure 18.6 Schematic illustration of the operation in a Cu₂S-based gap-type atomic switch. (a) The equilibrium condition without applying a bias voltage. An electrochemical potential diagram for Cu⁺ ions in the Cu₂S and Cu on the surface is also

shown. (b) Application of positive bias voltage to turn the switch ON. (c) Application of negative bias voltage to turn the switch OFF. E₀ and E_r are the activation energies for the oxidation (Cu to Cu⁺) and reduction (Cu⁺ to Cu) processes [14].

ions diffuse toward the Cu substrate, resulting in a decrease in the concentration of Cu⁺ ions at the subsurface, and E₀ becomes smaller than that of E_r. This causes the precipitated Cu atoms to dissolve back into the Cu₂S material (Figure 18.6c), and hence, the switch is turned off.

The switching time consists of the initial nucleation time and nucleus growth time. The very first step of nucleation is the formation of a critical nucleus, which usually consists of only a few atoms, and hence, its thermodynamics and kinetics can be described in terms of the atomistic theory of nucleation [19]. After the critical nucleus forms, growth proceeds depending on the chemical reaction, the diffusion of metal cations toward the surface of the electrolyte, and the electric field in the nanogap. The rate-limiting step of the overall switching kinetics might depend on the ionic conductor material. For instance, Ag critical nucleus formation was found to be rate limiting for the RbAg₄I₅ system, which is a superionic solid electrolyte [19]. For the Ag₂S system, the chemical reaction that precipitates Ag atoms and diffusion of Ag⁺ ions toward the surface are considered to be two competitive rate-limiting steps [17]. However, the studies on the Cu₂S system emphasized the effect of the electric field in the nanogap, which is extremely large ($\sim 10^6$ V cm⁻¹) directly under the STM tip and causes the precipitated metal atoms to migrate preferentially toward the tip [15]. Thus, in the Cu₂S system, the

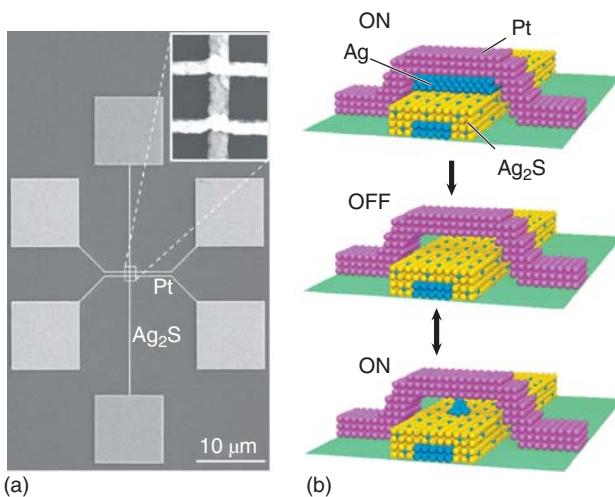


Figure 18.7 (a) Scanning electron microscope image of the gap-type atomic switch. The switches are formed at crossing points of the 150-nm-wide Ag_2S wire and the two 100-nm-wide Pt wires. (b)

Schematic diagram of the switch operation: switched-ON state (top), switched-OFF state (middle), and switched-ON state after the initial switching-OFF process (bottom).

electric-field-enhanced growth and the chemical reaction are considered to be probable rate-limiting steps [14].

18.2.3

Cross-Bar Structure

At the beginning of their development and in basic research, atomic switch operations using ionic/electronic mixed conductor materials, such as Ag_2S and Cu_2S , were demonstrated using STM [1, 9–11, 14–17]. Later, atomic switches could be fabricated at each crossing point by using a cross-bar structure [2]. The cross-bar structure enables integrated switches to be used in actual devices. An atomic switch consisting of the cross-bar structure with nanogaps, through which electrons can tunnel, were fabricated using a miniaturization technique based on semiconductor processing. Figure 18.7 shows this prototype with a cross-bar Pt/ $\text{Ag}_2\text{S}/\text{Ag}$ structure fabricated with electron-beam lithography and related techniques such as metal deposition and lift-off processes. Nanogaps of about 1 nm at the crossing points between an Ag_2S -coated Ag nanowire and Pt nanowires were formed in a solid electrochemical treatment after the lithography step formed the cross-bar structure. The overall process is as follows. First, the Ag nanowire was deposited on an insulator substrate, and the surface of the Ag nanowire was sulfurized to form an Ag_2S layer at 80 °C for 5 min in an ultrahigh vacuum by introducing sulfur vapor using a sulfur-valved cracker cell. A 1-nm-thick Ag layer was then deposited on the Ag_2S -covered Ag nanowire.

Subsequently, Pt was deposited to form the crossing nanowires, as shown in Figure 18.7b. The electrochemical treatment with the reaction $\text{Ag} \rightleftharpoons \text{Ag}^+ + e^-$ involved applying a bias voltage between the top Pt and bottom Ag nanowires. The application of a certain positive bias voltage to the Pt nanowires oxidized the Ag atoms of the 1-nm-thick Ag layer into Ag^+ ions, and the Ag^+ ions were incorporated into the Ag_2S layers. As a result, 1-nm-wide gaps were formed between the Pt nanowires and the Ag_2S -coated Ag nanowire, and the switches were in the OFF states because of the high resistances of the nanogaps. The switches could be turned on by applying a certain negative bias voltage to the Pt nanowire, wherein electrons tunneling from the Pt nanowires reduced Ag^+ ions in the Ag_2S layer, thereby precipitating Ag atoms in the nanogaps. The precipitated atoms bridged the Pt nanowires and the Ag_2S layer in the nanogaps, and the switches changed to the low-resistance ON state.

18.2.4

Quantized Conductance

Atomic switches enable not only ON/OFF operation merely through resistive switching, but also quantized conduction [2]. Here, precise control of local ion transport at the nanoscale level and the electrochemical reaction in the atomic switch yields an interesting conductance quantization phenomenon that appears even at room temperature [20]. This phenomenon was first found in a gap-type switch with a Pt/ Ag_2S /Ag structure. Quantized conductance appears, as shown in Figure 18.8a, when the Ag bridge between the Ag_2S -coated Ag layer and the Pt layer is slowly formed or dissolved by applying a relatively small bias voltage near the threshold voltage to start the electrochemical reaction $\text{Ag} \rightleftharpoons \text{Ag}^+_{(\text{Ag}_2\text{S})} + e^-$. Note that the time on the horizontal axis of Figure 18.8a starts immediately

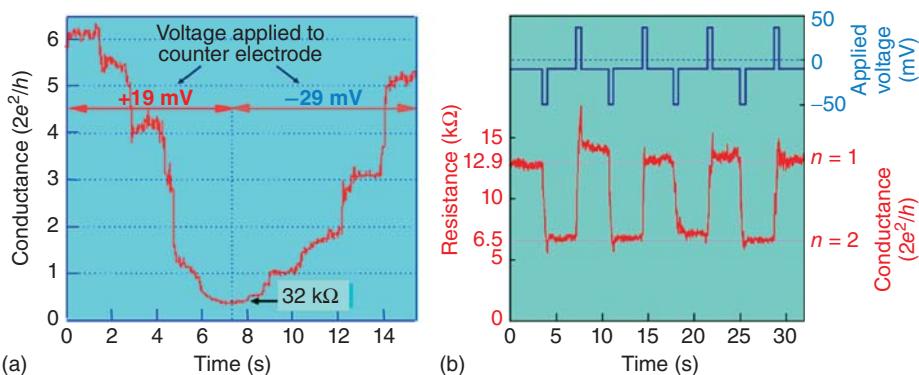


Figure 18.8 (a) Generation of quantized conductance during formation and dissolution point contact between the Ag filament and Pt substrate Ref. [20].

(b) Switching between the quantized conductance channels ($n = 1$ and 2) by alternately applying positive and negative voltage pulses.

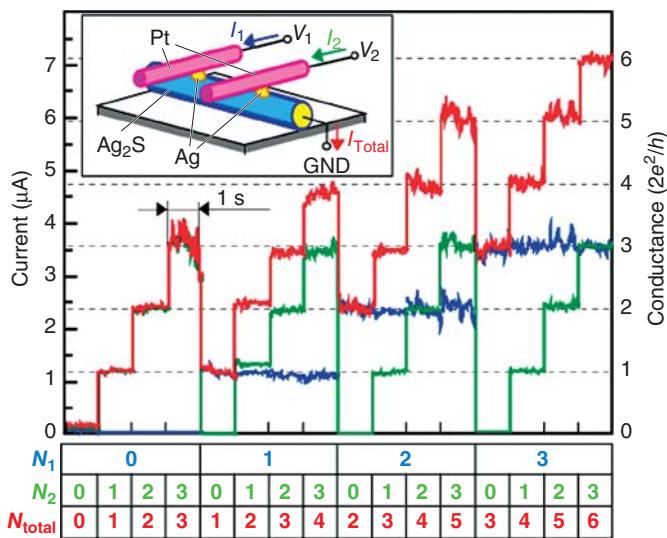


Figure 18.9 A 1 × 2 array of gap-type atomic switches. The quantized conductances of each channel were changed independently from $N=0$ to $N=3$; **50-ms-long pulsed bias voltages of 200 mV (from 0 to 1), 100 mV (from 1 to 2), 80 mV (from 2 to 3), and -260 mV (from 3 to 0)** were used [2].

before the point contact of the Ag bridge separated. The vertical axis indicates the numbers of quantized units, as calculated by dividing the electrical conductance by $2e^2/h$ (where e is the charge of the electron and h is Plank's constant). Applying a bias voltage of 19 mV to the Pt layer for a period of time decreased the quantum number from 6 to 1 in steps, and the quantum point contact of the Ag bridge became separated. After that, the polarity of the bias voltage switched to negative, and -29 mV was applied. The Ag atoms precipitated and the Ag bridge reformed into the quantum point contact. The quantum number then increased from 1 to 6 with increasing area of the point contact.

Repeatable switching between quantum numbers of 1 and 2 could be done by applying alternating pulse bias voltages of positive and negative polarities, as shown in Figure 18.8b. The applied pulse bias voltages were larger than the threshold bias voltage of the electrochemical reaction ($\text{Ag} \rightleftharpoons \text{Ag}^{+}_{(\text{Ag}_2\text{S})} + e^-$). A bias voltage smaller than the threshold one was applied in order to read the quantum number between these pulse biases. To demonstrate this operation, 1 × 2 array atomic switches with cross-bar structures consisting of Ag₂S-coated Ag nanowire and two Pt nanowires were fabricated. Figure 18.9 depicts experimental results showing that the quantum numbers of two point contacts in the 1 × 2 arrays are independently controllable. We could change the quantum number from 0 to 3 at each point contact, which means $4 \times 4 = 16$ states could be created. The array of two atomic switches can thus function as a multistate memory storing 16 states.

18.2.5

Logic-Gate Operation

Basic circuits for computing, such as logic gates, can be configured using gap-type atomic switches [2]. For instance, Figure 18.10a–c shows AND, OR, and NOT gates. The AND gate consists of a resistor and two atomic switches with crossing Ag₂S-coated Ag nanowire and Pt nanowires. Two input signals (V_1 and V_2) were applied to the Pt wires, and the output signal (V_{out}) was measured at one of the ends of the Ag₂S-coated Ag nanowire. The formation and annihilation of the Ag bridges between the Pt and Ag₂S–Ag nanowires were controlled according to the change in the two input signals. V_{out} became high only when both inputs signals were high. The OR gate was fabricated using a resistor and two atomic switches with a cross-bar structure. In this circuit, V_{out} became low only when both input signals were low. The NOT gate consists of a single atomic switch with two resistors and a capacitor. Here, V_{out} became low when both input signals were

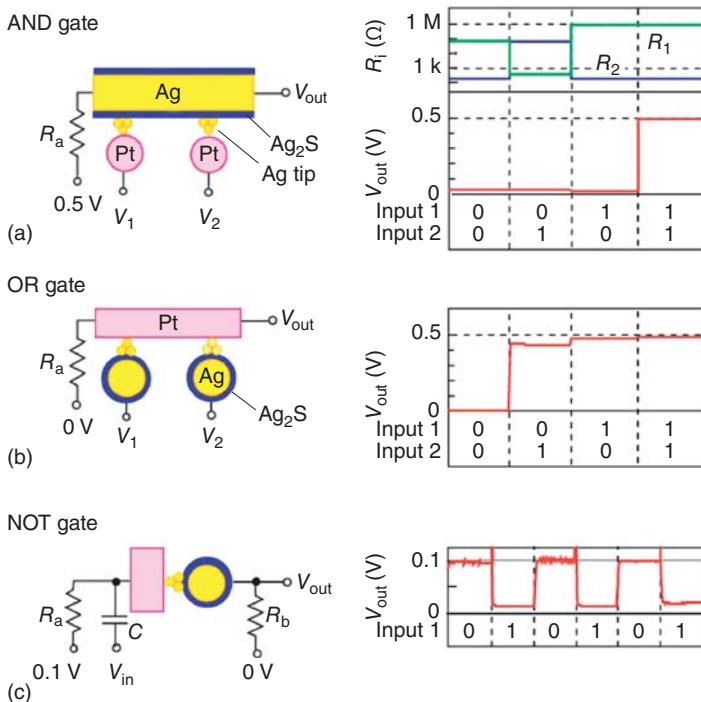


Figure 18.10 (a) Schematic diagram of AND gate built using gap-type atomic switches (left) and its operation (right). (b) Schematic diagram of OR gate (left) and its operation (right). (c) Schematic diagram of NOT gate (left) and its operation (right). Resistors R_a (10 kΩ) and R_b (1 kΩ) and a capacitor C

(100 pF) were used. V_1 and V_2 were applied as input bias voltages. Input level 1 was 0.5 V for the AND and OR gates, and 1.5 V for the NOT gate. Input level 0 was 0 V for all gates. Input levels were changed every second [2].

high. This operation was achieved by charging and discharging the current in the capacitor, which turned off and on the atomic switch, respectively. These results mean that any kind of logic circuit can be realized using atomic switches.

18.2.6

Synaptic Behavior

Today's computers are of the von Neumann type, which has separate memories and processors and is suited to performing logical operations and mathematical calculations. However, this sort of computer has a hard time handling image recognition and open-ended problems that are easily dealt with by the human brain. In the human brain, incoming analog data can be stored and processed in the same place at the same time, enabling it to quickly process massive amounts of information in parallel in a small area without consuming too much energy. In addition, the brain has the ability to learn, as a result of the synaptic plasticity of its neurons. For these reasons, the advent of artificial neural networks based on synaptic electronics, which emulate the information processing in brains, has been eagerly anticipated.

A synapse in the brain acts as a two-terminal element connecting two neurons. The connecting weight (i.e., strength) is modified in response to learning and forgetting. Here, the gap-type atomic switch shows artificial synaptic functions depending on the history of the input signals. In practice, the characteristics can be realized by controlling the formation and annihilation of the metal filament in the nanogap through the application of a bias voltage [7, 8, 21, 22].

Information is believed to be memorized in the brain as result of the two synaptic actions of short-term plasticity (STP) and long-term potentiation (LTP). Figure 18.11 illustrates experimental demonstration of the STP and LTP behaviors emulated in a gap-type atomic switch. Here, enhancement of the conductance in an atomic switch with an Ag/Ag₂S/Pt structure is related to STP; this enhancement occurs before the Ag bridge is completely formed, and it subsequently decays if the bridge is incomplete and dissolves back into the Ag₂S. However, the enhancement can be maintained for a long period once a stable bridge is formed. This situation corresponds to LTP. These STP and LTP behaviors are schematically illustrated in Figure 18.11a. STP and LTP are measured by controlling the conditions under which the bias voltage generates the electrochemical reaction of $\text{Ag} \rightleftharpoons \text{Ag}^+_{(\text{Ag}_2\text{S})} + e^-$. The STP behavior arises as a result of applying pulses with a bias (V) of 80 mV, a width (W) of 0.5 s, and a repetition interval (T) of 2 s, as shown in Figure 18.11b. Although the conductance rapidly increases, the high conductance (77.5 μS) of approximately one quantized channel immediately decreases back to a low value. On the contrary, the LTP behavior appears by applying the same V and W but a different T of 20 s, as shown in Figure 18.11c. In this case, the high-conductance state was maintained and did not return to a low-conductance state. A permanent transition from a low-conductance state to a high-conductance state was obtained with repeated application of input pulses, thereby emulating the LTP in a biological synapse. The LTP phenomenon in the atomic switch results from

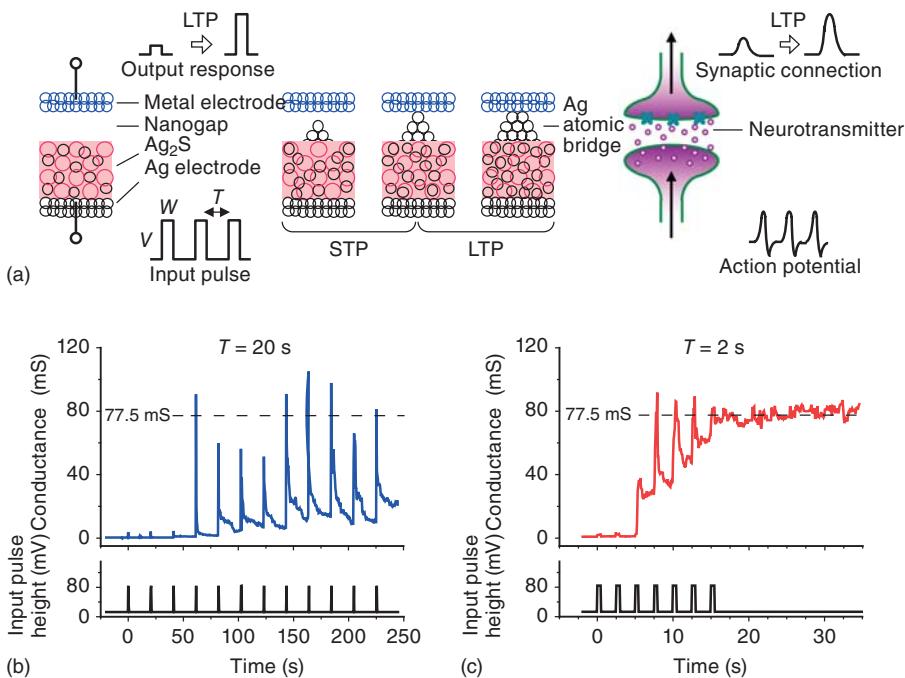


Figure 18.11 (a) Schematic illustrations of an inorganic synapse and the signal transmission of a biological synapse. Application of input pulses causes the precipitation of Ag atoms from the Ag₂S electrode, resulting in the formation of an Ag atomic bridge between the Ag₂S electrode and counter metal electrode. When the precipitated Ag atoms do not form a stable bridge, the inorganic synapse works as an STP. After a stable bridge is formed, it works as an LTP. In the case of a biological synapse, the

release of neurotransmitters is caused by the arrival of action potentials generated by firing, and then a signal is transmitted as a synaptic potential. Frequent stimulation causes long-term enhancement in the strength of the synaptic connection. (b and c) Change in the conductance of the inorganic synapse when the input pulses ($V = 80$ mV, $W = 0.5$ s) were applied with intervals of (b) $T = 20$ s and (c) 2 s, which shows the STP and LTP, respectively [8].

the formation of a stable and strong Ag bridge in the nanogap. This phenomenon is similar to the one in which there is a persistent increase in the synaptic connection following high-repetition stimulations by action potentials in the brain.

18.2.7

Photo-Assisted Switch

The atomic switch works by controlling the local ion migration and the electrochemical reaction. A gap-type atomic switch works by using tunneling electrons created by applying an external bias voltage. A photo-assisted atomic switch has also been developed; it works with a photocurrent instead of tunneling electrons [23, 24]. Here, a photoconductive layer is put in a wider nanogap between the

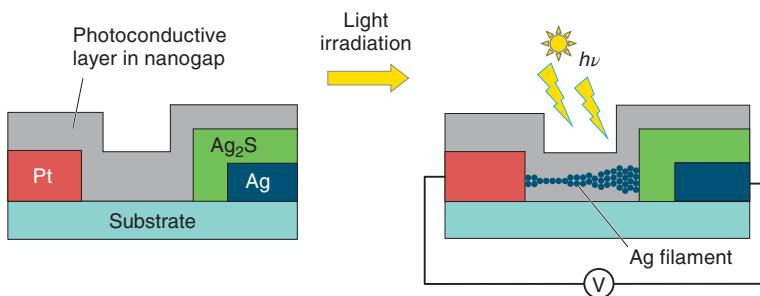


Figure 18.12 Schematic diagram of a photo-assisted atomic switch with photoconductive organic layer in the gap between the electrode and the ionic conductor.

ionic conductor and the electrode, as shown in Figure 18.12. Light generates a photocurrent between two biased electrodes, resulting in the formation of a metal filament in the photoconductive layer. The photo-assisted atomic switch was experimentally operated using an Ag₂S mixed ionic and electronic conductor and *N,N'*-diheptylperylene-tetracarboxylicdiimide (PTCDI) as the photoconductive molecule. Switching was repeated through formation and annihilation of the Ag filament in the photoconductive layer between the Pt electrode and the Ag₂S-coated Ag electrode. The switching time for the first demonstration was very long (about 1000 s) because of the large gap of the photo-conductive layer (80 nm). It takes a long time for the Ag atoms to nucleate and to grow a filament toward the opposite electrode in the photo-conductive gap. However, the switching time could be dramatically shortened by narrowing the gap. These characteristics will be useful for creating unique photosensing circuits such as intelligent artificial retinas.

18.3 Gapless-Type Atomic Switches

The metal/ionic conductor/metal (MIM) cell, in which a metal-ion conductor thin film is sandwiched between an electrochemically active electrode (usually formed from Ag or Cu) and an inert metal electrode (formed from, e.g., Pt or Au), also shows a resistive switching behavior with unique characteristics. The MIM cell is referred to as a “gapless-type atomic switch,” and similar to the gap type, its switching mechanism is based on the migration of metal ions and solid-state electrochemical reactions.

18.3.1 Sulfide-Based Switch

The gapless-type atomic switch was first reported using Cu₂S as the ionic conductor film [25]. An Au/Pt/Ti/Cu₂S/Cu cell was fabricated by electrochemically

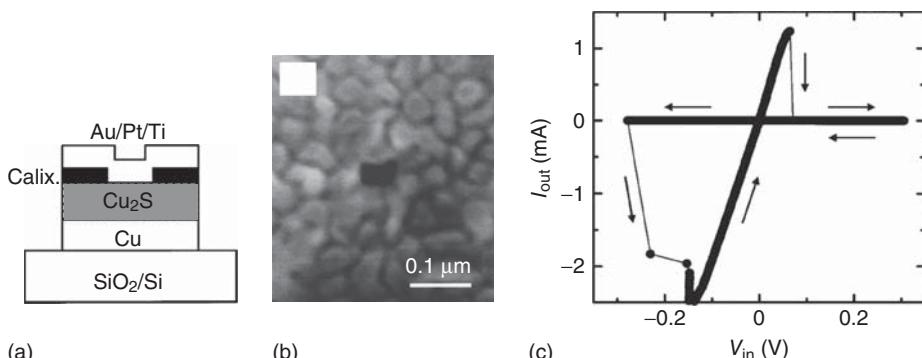


Figure 18.13 (a) Schematic view of a gapless-type atomic switch using a Cu_2S film sandwiched between Cu film and a top electrode ($\text{Au}/\text{Pt}/\text{Ti}$). (b) Plane view of top

electrode with a hole in the center. (c) I - V characteristic of the cell with a $0.03 \mu\text{m}$ hole. (Reproduced with permission from Ref. [25]. © (2003), AIP Publishing LLC.)

sulfurizing Cu film on SiO_2/Si in an Na_2S solution under anodic polarization. Figure 18.13a,b shows a schematic cross section and top view of a cell with $0.03 \mu\text{m}$ hole. Figure 18.13c shows typical current–voltage (I - V) characteristics while sweeping the bias voltage relative to the top ($\text{Au}/\text{Pt}/\text{Ti}$) electrode. The cell was turned on at $\sim -0.3 \text{ V}$ and turned off at $\sim 0.1 \text{ V}$, with a high ON/OFF-resistance ratio larger than 10^6 . The switch could also be operated by applying $\pm 0.3 \text{ V}$ pulse voltages with a pulse width of 1 ms, and it worked for more than 3×10^3 cycles. The simple structure, scalability, and low voltage operation of the nanometer-scale switch are advantageous characteristics for nonvolatile memory elements.

The effects of varying the sulfurization conditions and postannealing treatment on the structural and electrical properties of Ag_2S [26] and Cu_2S were investigated [27]. Increasing the sulfurization time resulted in a composition variation from $\text{Ag}(\text{Cu})$ -rich to S-rich through the stoichiometric composition, and $\text{Ag}(\text{Cu})$ ions existed in the most polarizable environment in the $\text{Ag}(\text{Cu})$ -rich films. Postannealing of the sulfurized films resulted in $\text{Ag}(\text{Cu})$ -rich films in all cases, and clear switching of the resistance could be observed for all the annealed films. This result indicates that excess $\text{Ag}(\text{Cu})$ in the sulfide films is required for the observation of resistive switching behavior. The atomic switches using sulfides are usually operated at bias voltages lower than 0.3 V.

18.3.2

Oxide-Based Switch

Another type of gapless-type atomic switch, in which a thin oxide film is used as the ionic conductor, has been proposed as a way of increasing threshold voltages for resistive changes above the operating voltage of complementary metal-oxide-semiconductor (CMOS) devices. A $\text{Cu}/\text{Ta}_2\text{O}_5/\text{Pt}$ cell was fabricated and bipolar

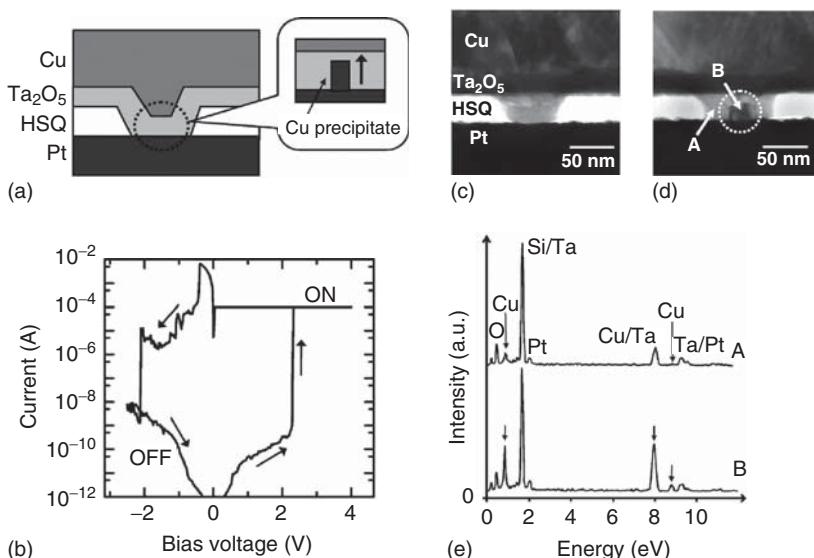


Figure 18.14 (a) Switch structure and (b) I - V characteristics of a switch with 15-nm-thick Ta₂O₅ film sandwiched between Cu and Pt electrodes. The inset is a cross section of Cu metal precipitated in Ta₂O₅. Cross-sectional transmission-electron-microscope images of switch (c) before turning it ON and (d)

after turning it OFF. (e) Spectra of energy dispersive X-ray spectrometer (EDX). EDX signals were measured for points A and B in (d). The arrows in (e) show Cu-related signals. (Reproduced with permission from Ref. [28]. © (2007), AIP Publishing LLC.)

resistive switching behavior was observed while sweeping of the bias voltage relative to the Cu electrode, as shown in Figure 18.14a,b [28]. The cell was turned on at ~ 2 V and turned off around -2 V, and the ON/OFF-resistance ratio was more than 5 orders of magnitude. Figure 18.14c,d shows cross-sectional transmission-electron-microscope images of the cell measured before and after turning it on. From energy dispersive X-ray spectra (Figure 18.14e) obtained at points A and B in Figure 18.14d, it was found that a dark region (B) consists of precipitated Cu metal in the Ta₂O₅ film. This result indicates that the observed resistive switching can be attributed to the formation and dissolution of the metal filament between the two electrodes as the bias voltage is being swept. Similar switching phenomena were also reported using other oxides such as SiO₂ [29, 30], HfO₂ [31], and WO₃ [32]. However, their resistive switching mechanism is not fully understood.

The switching mechanism of oxide-based atomic switches was investigated by carefully comparing the I - V and current-time (I - t) characteristics of Cu/Ta₂O₅/Pt and Pt/Ta₂O₅/Pt cells [33]. It was found that repeatable resistive switching occurred only when a positive bias was initially applied to the Cu electrode of the Cu/Ta₂O₅/Pt cell. The first turn-ON operation corresponds to the first formation (forming process) of a metal filament between the electrodes through inhomogeneous nucleation and subsequent growth of Cu on the Pt electrode, as a result of migration of Cu ions from the Cu electrode into the

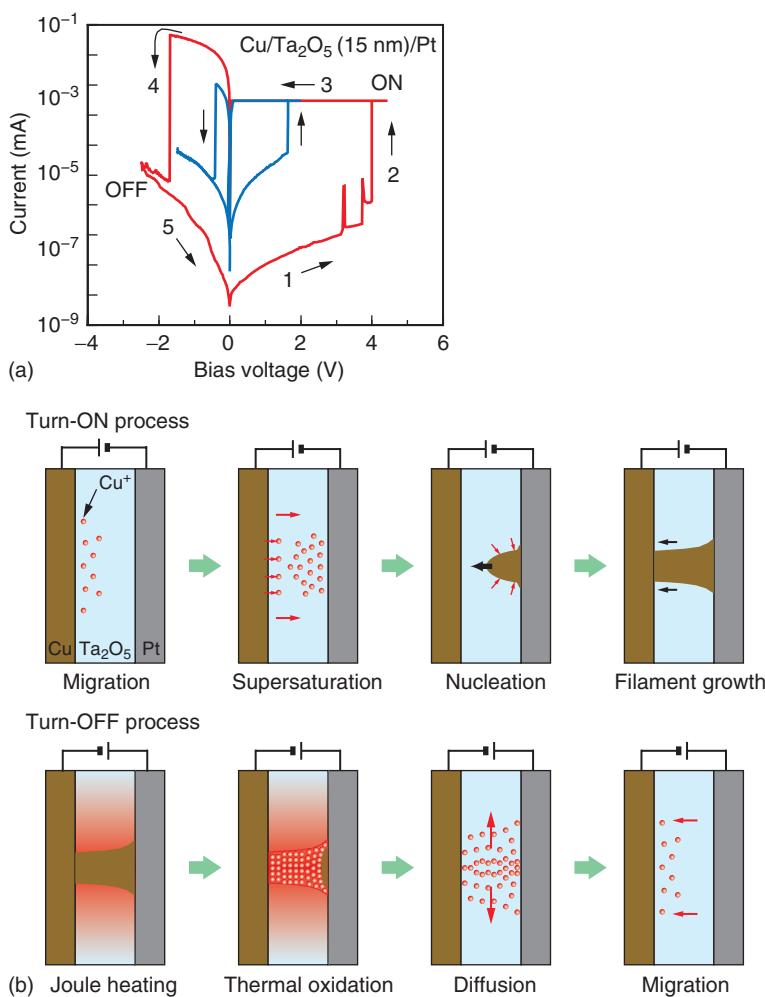


Figure 18.15 (a) Typical forming (red curve) and subsequent switching (blue curve) characteristics of the Cu/Ta₂O₅/Pt atomic switch. The turn-ON and turn-OFF mechanisms are schematically illustrated in (b).

Ta₂O₅ film. The forming process occurred at lower bias voltages (~ 2 V) but with longer times ($> 10^3$ s) under the application of a constant bias voltage, which is different from the forming of transition-metal-oxide (TMO)-based valence change memory (VCM). In the case of TMO-based VCM, the forming generally induces an electrical breakdown that destroys a small part of the TMO film. Indeed, the Pt/Ta₂O₅/Pt cell showed only electrical breakdown behavior and no resistive switching after the breakdown.

After the forming process, the Cu/Ta₂O₅/Pt cell exhibited bipolar switching behavior with the repetition of turn-ON and turn-OFF operations, as shown in

Figure 18.15a. The bias voltages needed to turn-ON and turn-OFF the switch gradually decreased and eventually became almost constant. From the measurements of the $I-t$ characteristics and thermal stability of the ON state, it was concluded that the turn-ON process corresponds to the reformation of the metal filament through nucleation, whereas the turn-OFF process is due to thermal dissolution of the metal filament from Joule-heating-assisted oxidation followed by diffusion of (oxidized) Cu ions under the concentration gradient and applied electric field, as illustrated in Figure 18.15b.

To determine the validity of the switching model of Figure 18.15b, the effect of varying the temperature on the switching behavior of the Cu/Ta₂O₅/Pt cell was examined [34]. The turn-ON and turn-OFF voltages decreased in magnitude as the temperature increased. From the nucleation rate of Cu nuclei calculated based on the basis of classical nucleation theory, it was found that the observed temperature variation of the turn-ON voltage is determined by supersaturation in the vicinity of the Pt electrode and migration of Cu ions in the Ta₂O₅ film. The supersaturation required for spontaneous growth of a Cu nucleus decreases with increasing temperature, resulting in lower turn-ON voltages at high temperature. The turn-OFF voltage is determined by the size of the metal filament formed at a given temperature and its thermal stability. Moreover, the growth speed of the Cu nucleus after the nucleation decreases with increasing temperature. These results are consistent with the switching model of Figure 18.15b.

18.3.3

Effect of Moisture

It is well known that oxide materials can absorb moisture from the ambient air, which causes shifts in the characteristics of CMOS devices. The impact of moisture on the switching characteristics was investigated for Cu/Ta₂O₅/Pt and Cu/SiO₂/Pt cells [35]. The turn-ON and turn-OFF voltages of the Cu/Ta₂O₅/Pt cell were nearly constant, whereas those of the Cu/SiO₂/Pt cell exhibited an increase in magnitude as the ambient H₂O pressure decreased. Fourier-transform infrared absorption measurements showed the presence of a certain amount of residual water in both the Ta₂O₅ and SiO₂ films, which had been absorbed from the air ambient. The residual water remained very stable in the Ta₂O₅ film (Figure 18.16b) but easily desorbed from the SiO₂ film (Figure 18.16c) as the ambient H₂O pressure decreased. The ionization of Cu at the anode interface was attributed to chemical oxidation assisted by residual water, as illustrated in Figure 18.16a. Moisture absorption also resulted in a hydrogen-bond network forming at grain boundaries in the oxide film, and metal ions were likely to migrate along the grain boundaries. Because of the strong hydrogen bonds in the Ta₂O₅ film, the operation voltages of the Cu/Ta₂O₅/Pt cell were independent of the ambient H₂O pressure. In contrast, the switching behavior of the Cu/SiO₂/Pt cell was significantly influenced by the ambient conditions because residual water easily desorbed from the SiO₂ film.

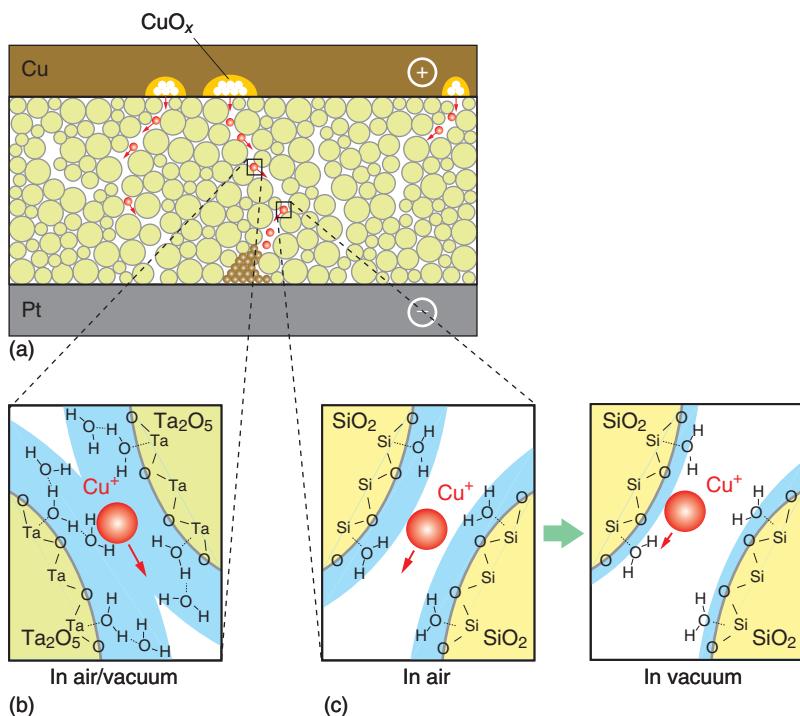


Figure 18.16 Potential mechanism for the moisture effects on the ionization of Cu and the migration of Cu ions in an oxide-based atomic switch (a). Schema (b) and

(c) illustrate the bonding states of residual water molecules at grain boundaries in Ta_2O_5 and SiO_2 films, respectively, under air and vacuum conditions [35].

18.3.4

Switching Time

Switching-time measurements are very important not only for demonstrating the fast switching speed, but also for gaining an understanding of the high-speed dynamics of the switching mechanism. Torrezan *et al.* measured a switching time of ~ 0.1 ns in a Ta_2O_5 -based VCM integrated into a coplanar waveguide (CPW) structure [36]. The integration into the CPW structure enabled impedance matching between the measurement system and the device under investigation. Measuring the transient current during the switching operation of the oxide-based atomic switch is very difficult because of the high ON/OFF-resistance ratio, which leads to a large impedance mismatch. This problem was overcome by integrating the cell into the CPW structure and by using a specially designed measurement system [37]. The results showed that a turn-ON time as low as 1 ns could be achieved using moderate pulse amplitudes, and this suggests that oxide-based atomic switches would be promising for fast switching memory devices. From a comparison with calculations based on atomistic nucleation and

hopping conduction theories, nucleation on the Pt electrode was found to be likely the rate-limiting process determining the turn-ON time. On the contrary, a turn-OFF time decreases from a few microseconds to several nanoseconds with increasing OFF-resistance from ~ 200 to $\sim 600 \Omega$. The electrical power needed to turn off the cell decreases with the increase of ON-resistance. This suggests that the turn-OFF process may involve thermal effects to dissolve the metal filament.

18.3.5

Quantized Conductance and Synaptic Behavior

Besides having the ability of bistable switching with a high ON/OFF-resistance ratio, gapless-type atomic switches also show quantized conductance and synaptic behavior, similar to the gap type [38]. Figure 18.17a is an example of the changes in the conductance of an Ag/Ta₂O₅/Pt cell when positive voltage pulses with a pulse width of 20 ms are applied to it at intervals of 2 s. The conductance

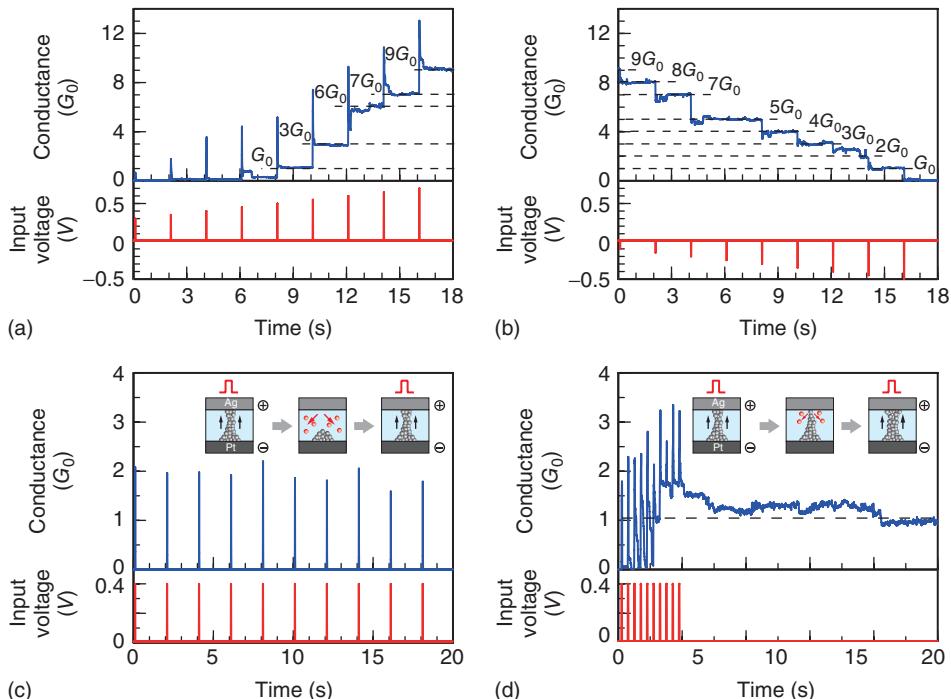


Figure 18.17 Increase (a) and decrease (b) in the quantized conductance of the Cu/Ta₂O₅/Pt cell observed by applying positive (a) and negative (b) voltage pulses with widths of 20 ms at intervals of 2 s, respectively. Changes in cell conductance under

the application of 10 consecutive input pulses at intervals of 2 s (c) and 0.2 s (d). The insets illustrate the formation of an atomic contact under consecutive input pulses with the corresponding interval times [38].

increased in a stepwise fashion as the pulse amplitude increased, and each step height corresponds to the conductance of a single atomic point contact G_0 . Figure 18.17b shows the changes in the cell conductance after those of Figure 18.17a, this time with negative voltage pulses. The conductance decreased in a stepwise fashion with increasing negative amplitude and eventually reached zero. The observed conductance changes can be attributed to the formation and dissolution of a metal filament with an atomic point contact of different integer multiples in the Ta_2O_5 film. Similar phenomena were also observed in MIM cells using GeS_2 [39] and AgI [40].

The conductance behavior changed drastically depending on the interval of the input pulses. Figure 18.17c shows the conductance variation when 10 input pulses were applied at intervals of 2 s. Each pulse enhanced the conductance of the cell by up to $\sim 2 G_0$. However, the conductance state decayed immediately after the pulses because of the unstable tiny metal filament. When the interval was decreased to 0.2 s (Figure 18.17d), the cell initially exhibited a temporary increase for the first few input. However, as the input pulses were increased, the conductance gradually increased with accompanying temporary increases and reached integer multiples of G_0 . This conductance state was maintained for more than 60 s. This behavior can be explained by the formation of a larger metal filament with a stable atomic contact, and it is analogous to the LTP behavior of biological synapses, in which the strength of synaptic weight depends upon the time interval between the stimulating pulses.

18.3.6

Polymer-Based Switch

An atomic switch can be realized with solid polymer electrolyte (SPE) materials as well as inorganic materials. Bipolar resistive switching of an $\text{Ag}/\text{SPE}/\text{Pt}$ cell was demonstrated using a film composed of polyethylene oxide–silver perchlorate complexes [41], as shown in Figure 18.18b. The switching behavior varied from nonvolatile to volatile with increasing silver salt concentration. The thermal, transport, and electrochemical properties of the SPE film revealed that the switching behavior results from the formation and dissolution of an Ag filament inside the SPE film, which is similar to the case of inorganic material-based atomic switches. The cell also exhibited a high ON/OFF-resistance ratio ($>10^5$), high programming speed ($<1 \mu\text{s}$), and long retention time (>1 week).

Recently, the $\text{Ag}/\text{SPE}/\text{Pt}$ cells were successfully fabricated on a plastic substrate by drop-on-demand inkjet printing (IJP) of a polymer electrolyte solution [42]. The advantage of IJP is that the required amount of the target solution can be deposited onto specific positions when needed. Thus, it is a more economical method of ink delivery than deposition methods such as spin coating. By exploiting the large difference between the surface energies of the metal electrode and the substrate, a thin SPE film could be formed intentionally over the electrode, as shown in Figure 18.18a. The fabricated $\text{Ag}/\text{SPE}/\text{Pt}$ cells operated for more than 100 cycles of sweeping the bias voltage in vacuum and in air. The cells also exhibited

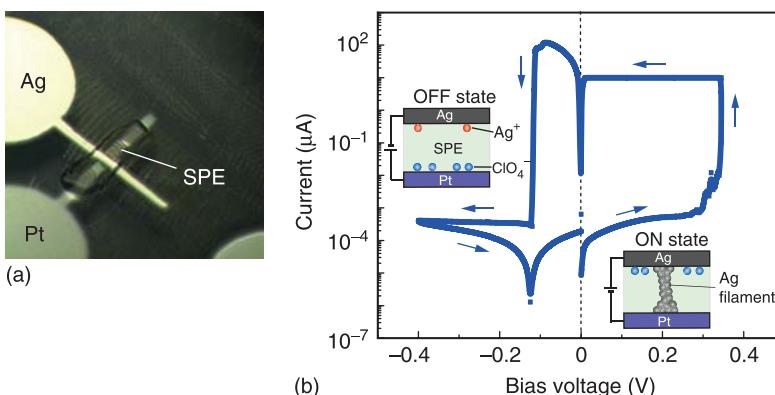


Figure 18.18 (a) Optical micrograph of a cross-point Ag/SPE/Pt cell fabricated on a plastic substrate. (b) Typical I - V characteristics of the Ag/SPE/Pt atomic switch.

The insets illustrate the ON and OFF states.
(Reproduced with permission from Ref. [42].
© (2012), AIP Publishing LLC.)

stable switching even when the substrate was bent. These results indicate that SPE-printed atomic switches are promising for flexible switch/memory applications.

18.4

Three-Terminal Atomic Switches

In three-terminal atomic switches, the formation and annihilation of a conductive path between the source electrode and the drain electrode is controlled by applying a bias to the gate electrode. Three-terminal atomic switches have an advantage in that the signal and control lines are separated, as in a semiconductor transistor, which dramatically reduces their power consumption. The high level of controllability widens the application range as well. For instance, nonvolatile atomic switches offer a solution to issues inherent in conventional semiconductor transistors. In particular, the recent development of the redox-based three-terminal nonvolatile switch, which is operated with a polarity opposite to that of the three-terminal atomic switches, shows promise for nonvolatile versions of CMOS functions.

There are two types of three-terminal atomic switch: one based on filament growth and the other based on the nucleation of metal clusters. Each type has advantages in different applications.

18.4.1

Filament-Growth-Controlled Type

The formation and annihilation of the metal filaments between the source and drain electrodes can be controlled with a gate electrode. For instance, the

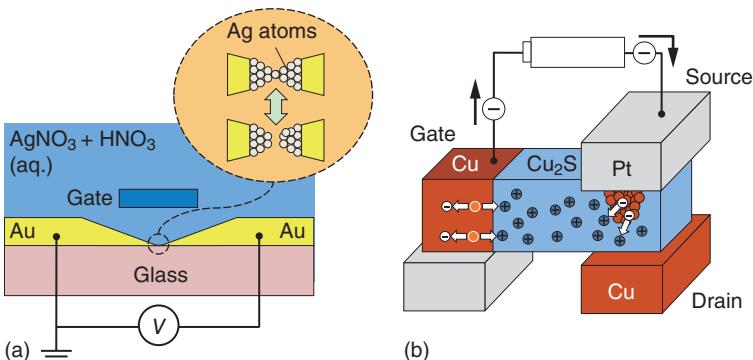


Figure 18.19 Three-terminal type atomic switches in which filament growth is controlled using an electrochemical reaction. (a) Schematic diagram of three-terminal device using (a) an electrolyte [43] and (b) a solid electrolyte [44].

deposition/dissolution of metal atoms in an electrolyte can be controlled using a reference electrode in electrochemistry. Cyclic voltammetry is a typical measurement using the electrochemical technique. The electrochemical technique has been used to control the atomic scale bridge formation/dissolution between two Au electrodes immersed in an electrolyte of $\text{AgNO}_3 + \text{HNO}_3$, as shown in Figure 18.19a [43]. It has been shown that switching among the quantized conductances can be controlled by varying the gate voltage [44]. Switching like this in an electrolyte has also been demonstrated using other materials, such as CuSO_4 [45].

Filament growth is also possible in a solid electrolyte, as shown in Figure 18.19b [46]. In the first study of this kind, Cu_2S was used as a solid electrolyte in which Cu^{2+} cations migrate in the direction of the electrical field. Filament formation/dissolution was achieved by supplying/extracting Cu^{2+} cations to/from the channel region. The filament initially grows with Cu^{2+} cations supplied from a drain electrode made of Cu, which reduces a switching area controlled by the gate electrode, resulting in high reliability and short switching time.

Although the first three-terminal atomic switch required initialization, a later version used Cu for the gate electrode and Pt for both the drain and source electrodes (Figure 18.20a,b) [47] and work without initialization. Figure 18.20c plots the I - V characteristics as a function of the gate voltage. This structure has an ON/OFF ratio higher than 10^4 and a cyclic endurance of 10^2 times (Figure 18.20d). However, because it uses a solid electrolyte (Cu_2S) that also has electrical conductivity, the gate current is on the order of a microampere, which is much larger than that of semiconductor transistors. Consequently, three-terminal atomic switches based on filament formation are potentially reconfigurable and would have uses in programmable devices that are not turned on/off so often. Moreover, the new structure can be easily formed in metal wiring layers. The other characteristics of this device, such as its very low ON-resistance and ON/OFF ratio, can meet a range of application requirements.

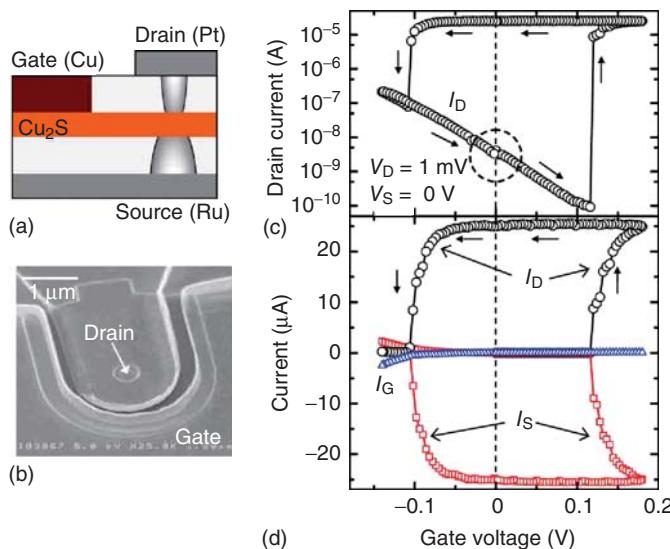


Figure 18.20 (a) Schematic cross section and (b) scanning electron microscopic image of three-terminal atomic switch. (c) Change in drain current on a log scale

and (d) change in drain, source, and gate currents while sweeping the gate voltage. (Reproduced with permission from Ref. [47]. © (2010), AIP Publishing LLC.)

18.4.2

Nucleation-Controlled Type

Another type of three-terminal atomic switch is based on the control of the nucleation of metal clusters [46]. Figure 18.21a shows the concept of the three-terminal atomic switch, where metal cations brought from the gate electrode form a conductive channel between the source and drain electrodes. The key here is the use of Ta₂O₅, an insulator in terms of electrical conductivity, as the ion-transport material. When the concentration of metal cations, which is a function of the gate electric field, reaches a certain threshold value, nucleation of metal clusters takes place, resulting in a dramatic change in conductivity, that is, an insulator to conductor transition. As the gate electric field becomes smaller, nuclei dissolve into the metal cations moving toward the gate electrode and the three-terminal atomic switch turns OFF. Since this nucleation type of switch works in a similar way to that of a silicon transistor, that is, the state variable is gate voltage (V_G), it is referred to as an *Atom Transistor*.

Atom transistors possess novel characteristics, such as being able to select between volatile and nonvolatile operations. They also consume very little power (pW), have high ON/OFF ratios (10^6 (volatile operation) to 10^8 (nonvolatile operation)), and are compatible with CMOS processes. This enables them to be used in computing systems that fully utilize highly integrated CMOS technology. A prototype showed good endurance of 10^4 switching cycles.

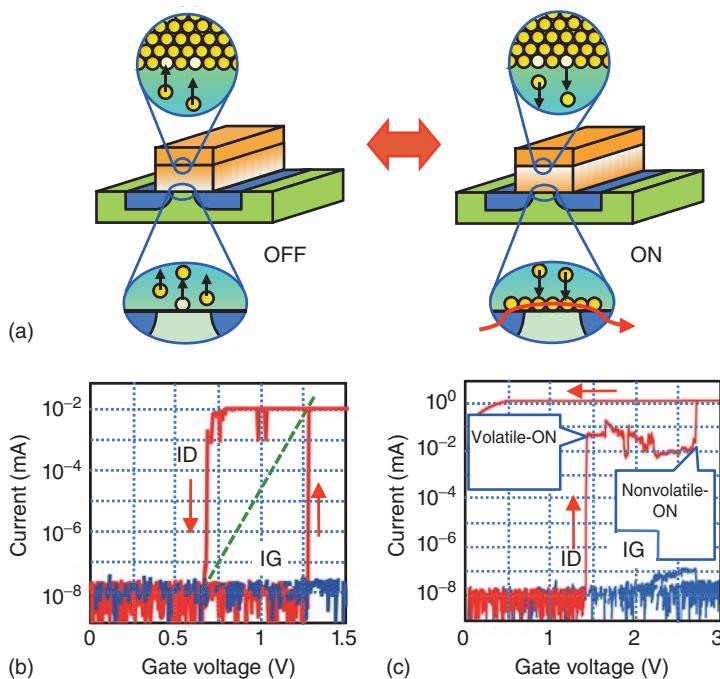


Figure 18.21 Three-terminal type atomic switch controlling supersaturation-induced nucleation. (a) Schematic illustration of the operating [46] mechanism, (b) change

in drain (red) and gate (blue) currents in volatile operation, and (c) those in nonvolatile operation. (© (2011) The Japan Society of Applied Physics.)

Figure 18.21b,c shows the changes in the drain current (I_D) and gate current (I_G) in the atom transistor (Ta_2O_5 : 20 nm, Cu (gate): 30 nm) when the gate voltage (V_G) was swept. By sweeping V_G between 0 and 1.5 V, the atom transistor showed volatile operation; namely, it turned on when V_G reached 1.25 V and turned off at $V_G = 0.65$ V when the gate voltage was swept back to 0 V. Thus, I_D changed by 6 orders of magnitude in the V_G range of 0.65–1.25 V, corresponding to a sub-threshold slope of 100 mV per decade.

Figure 18.21c shows two different switching behaviors when V_G was swept to 3 V. The first switching occurred at $V_G = 1.4$ V, when I_D was raised to about $10 \mu A$; this result is similar to the one shown in Figure 18.21b. The second switching occurred at $V_G = 2.65$ V, and it increased I_D by another 2 orders of magnitude. Once the second switching occurred, the ON state was kept at $V_G = 0$ V, requiring the application of a negative V_G to turn off the atom transistor.

In volatile and nonvolatile operations, I_G remains small because of the use of Ta_2O_5 . Although further experiments will be required, the existence of volatile and nonvolatile modes depending on the gate voltage range relies on the stability of the nuclei formed while increasing the gate voltage.

Since supersaturation-induced nucleation is used in this type of three-terminal atomic switch, there is a clear threshold voltage in operation; namely, the state variable is the gate voltage, which means that such a device is more suitable for logic circuits rather than memories. In this sense, the recent development of redox-based nonvolatile three-terminal switches [48], with a gate bias polarity opposite to that of other three-terminal atomic switches, is a promising step toward electrochemically controlled nonvolatile versions of CMOS functions.

18.5 Summary

In this chapter, we have discussed various types of atomic switches. All types of atomic switches, including some not discussed in this chapter, are summarized in Figure 18.22. The atomic switches are classified into *gap-type* and *gapless-type*, as shown in Figure 18.22. In both types, we can design *nonvolatile*, *volatile*, and *synapse-like* atomic switches (volatile atomic switches were not discussed in this chapter). Not only the original *two-terminal* atomic switches but also *three-terminal* atomic switches have been realized. Due to the wide variety of types and structures, atomic switches can have various functionalities, including simple *ON/OFF switching*, *quantization of conductivity*, *learning and forgetting*,

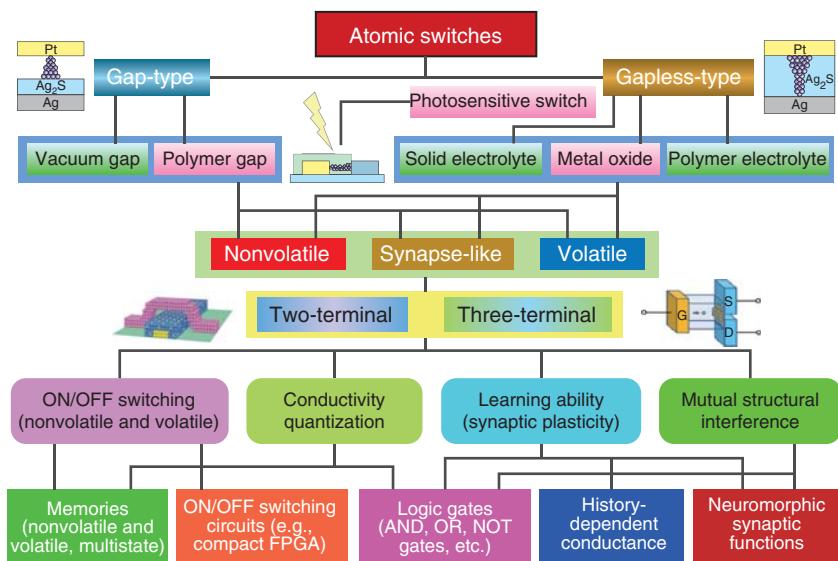


Figure 18.22 Overview of wide variety of types and functionalities of the atomic switch. In the top half, various types of the atomic switch are shown, indicating their mutual relationships. The atomic switch

exhibits various functionalities depending on its types and can be used to realize various novel nanoscale devices, as shown in the bottom half.

and *mutual structural interference* (the last one was not discussed in this chapter). Because of this, we can realize nanoscale devices such as *memories*, *ON/OFF switches*, *logic gates*, *neuromorphic network circuits*, and *brain-type-material-based computers*, as listed in Figure 18.22. The last one is one of the most interesting future applications of the atomic switch, but its discussion was omitted in this chapter due to space limitations.

We would like to finish with a final important remark. If any material is sandwiched by two electrodes and a voltage is applied between the two electrodes, some change will occur in the material, often accompanied by a change in the electrical resistance of the material. In that case, a system consisting of the material and two electrodes can be used as an electrical switch or a memory bit, called *resistive switch* and *resistive memory*, respectively. If the system's *I-V* characteristic exhibits a pinched hysteresis, it is often called a *memristor* [49]. With this definition, the atomic switch is sometimes categorized as the resistive switch or resistive memory depending on the functionality used, and it is also called the *memristor* when showing the learning ability.

The essential mechanism of the atomic switch is the growth and shrinkage of a cluster (or an aggregate) of metal atoms (ions) formed on one of the two electrodes. The electrical conductance between the two electrodes is essentially determined by the nanoscale or atomic-scale electrical contact of the metal atom cluster with the counter electrode. Therefore, the atomic switch usually exhibits quantized values of conductivity. Recently, some resistive switches based on the formation and migration of oxygen vacancies in metal oxides have shown quantized conductance. The observed conductivity quantization is due to the formation of a metal filament or a metal-rich new phase (such as a Magnéli phase in TiO_2 [50]) similar to that in the atomic switch. Finally, because of the unique mechanism based on the reversible nanoscale or atomic-scale electrical contact, the atomic switch can have various functionalities that cover those of the resistive switch, resistive memory, and memristor, as shown in Figure 18.22.

References

- Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2001) Quantum point contact switch realized by solid electrochemical reaction. *Riken Rev.*, **37**, 7.
- Terabe, K., Haegawa, T., Nakayama, N., and Aono, M. (2005) Quantized conductance atomic switch. *Nature*, **433**, 47.
- Hasegawa, T., Terabe, K., Sakamoto, T., and Aono, M. (2009) Nanoionics switching device: atomic switches. *MRS Bull.*, **34**, 929.
- Hino, T., Hasegawa, T., Terabe, K., Tsuruoka, T., Nayak, A., Ohno, T., and Aono, M. (2011) Atomic switches: atomic-movement-controlled nanodevices for new types of computing. *Sci. Technol. Adv. Mater.*, **12**, 013003.
- Hasegawa, T., Terabe, K., Tsuruoka, T., and Aono, M. (2012) Atomic switch: atom/ion movement controlled devices for beyond Von-Neumann computers. *Adv. Mater.*, **24**, 252.
- Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833.
- Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J., and Aono, M.

- (2011) Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.*, **10**, 591.
8. Sillin, H.O., Shieh, H.H., Aguilera, R., Avizienis, A.V., Aono, M., Stieg, A.Z., and Gimzewski, K. (2013) A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology*, **24**, 384004.
 9. Terabe, K., Nakayama, T., Hasegawa, T., and Aono, M. (2002) Formation and disappearance of a nanoscale silver cluster realized by solid electrochemical reaction. *J. Appl. Phys.*, **91**, 10110.
 10. Terabe, K., Nakayama, T., Iyi, N., and Aono, M. (1999) in *Proceedings of the 9th International Conference on Production Engineering* (eds Y. Furukawa, Y. Mori, and T. Kataoka), The Japan Society for Precision Engineering, Osaka, p. 711.
 11. Terabe, K., Nakayama, T., Hasegawa, T., and Aono, M. (2002) Ionic/electronic mixed conductor tip of a scanning tunneling microscope as a metal atom source for nanostructuring. *Appl. Phys. Lett.*, **80**, 4009–4011.
 12. Rickert, H. (1982) *Electrochemistry of Solids – An Introduction*, Springer-Verlag, Berlin.
 13. Ohachi, T. and Taniguchi, I. (1972) Controlled filamentary growth of silver from silver compounds. *J. Cryst. Growth*, **13–14**, 191.
 14. Nayak, A., Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) Switching kinetics of a Cu₂S-based gap-type atomic switch. *Nanotechnology*, **22**, 235201.
 15. Nayak, A., Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) Theoretical investigation of kinetics of a Cu₂S-based gap-type atomic switch. *Appl. Phys. Lett.*, **98**, 233501.
 16. Tamura, T., Hasegawa, T., Terabe, K., Nakayama, T., Sakamoto, T., Sunamura, H., Kawaura, H., Hosaka, S., and Aono, M. (2007) Material dependence of switching speed of atomic switches made from silver sulfide and from copper sulfide. *J. Phys. Conf. Ser.*, **61**, 1157.
 17. Nayak, A., Tamura, T., Tsuruoka, T., Terabe, K., Hosaka, S., Hasegawa, T., and Aono, M. (2010) Rate-limiting processes determining the switching time in a Ag₂S atomic switch. *J. Phys. Chem. Lett.*, **1**, 604–608.
 18. Morales-Masis, M., van der Molen, S.J., Hasegawa, T., and van Ruitenbeek, J.M. (2011) Bulk and surface nucleation processes in Ag₂S conductance switches. *Phys. Rev. B*, **84**, 115310.
 19. Valov, I., Sapezanskaia, I., Nayak, A., Tsuruoka, T., Bredow, T., Hasegawa, T., Staikov, G., Aono, M., and Waser, R. (2012) Atomically controlled electrochemical nucleation at superionic solid electrolyte surfaces. *Nat. Mater.*, **11**, 530.
 20. Terabe, K., Hasegawa, T., Liang, C.H., and Aono, M. (2007) Control of local ion transport to create unique functional nanodevice based ion ionic conductors. *Sci. Technol. Adv. Mater.*, **8**, 536.
 21. Hasegawa, T., Ohno, T., Terabe, K., Tsuruoka, T., Nakayama, T., Gimzewski, J.K., and Aono, M. (2010) Learning abilities achieved by a single-solid-state atomic switch. *Adv. Mater.*, **22**, 1831.
 22. Ohno, T., Hasegawa, T., Nayak, A., Tsuruoka, T., Gimzewski, J.K., and Aono, M. (2011) Sensory and short-term memory formations observed in a Ag₂S gap-type atomic switch. *Appl. Phys. Lett.*, **99**, 203108.
 23. Hino, T., Tanaka, H., Haegawa, T., Aono, M., and Ogawa, M. (2010) Photoassisted formation of an atomic switch. *Small*, **6**, 1745.
 24. Hino, T., Haegawa, T., Tanaka, H., Tsuruoka, T., Terabe, K., Ogawa, M., and Aono, M. (2013) Volatile and nonvolatile selective switching of a photo-assisted initialized atomic switch. *Nanotechnology*, **24**, 384006.
 25. Sakamoto, T., Sunamura, H., Kawaura, H., Hasegawa, T., Nakayama, T., and Aono, M. (2003) *Appl. Phys. Lett.*, **82**, 3032.
 26. Kund, M., Terabe, K., Haegawa, T., and Aono, M. (2006) Effects of sulfurization conditions and post-deposition annealing treatment on structural and electrical properties of silver sulfide films. *J. Appl. Phys.*, **99**, 103501.
 27. Kund, M., Hasegawa, T., Terabe, K., and Aono, M. (2008) Effects of sulfurization conditions on structural and electrical

- properties of copper sulfide films. *J. Appl. Phys.*, **103**, 073523.
28. Sakamoto, T., Lister, K., Banno, N., Hasegawa, T., Terabe, K., and Aono, M. (2007) Nanometer-scale switches using copper sulfide. *Appl. Phys. Lett.*, **91**, 092110.
29. Schindler, C., Weides, M., Kozicki, M.N., and Waser, R. (2008) Low current resistive switching in Cu-SiO₂ cells. *Appl. Phys. Lett.*, **92**, 122910.
30. Schindler, C., Staikov, G., and Waser, R. (2009) Electrode kinetics of Cu-SiO₂-based resistive switching cells: overcoming the voltage-time dilemma of electrochemical metallization memories. *Appl. Phys. Lett.*, **94**, 023504.
31. Haemori, M., Nagata, T., and Chikyo, T. (2009) Impact of Cu electrode on switching behavior in a Cu/HfO₂/Pt structure and resultant Cu ion diffusion. *Appl. Phys. Express*, **2**, 061401.
32. Kozicki, M.N., Gapalan, C., Balakrishnan, M., and Mitokova, M. (2008) A low-power nonvolatile switching element based on copper-tungsten oxide solid electrolyte. *IEEE Trans. Nanotechnol.*, **5**, 535.
33. Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2011) Forming and switching mechanisms of a cation-migration-based oxide resistive memory. *Nanotechnology*, **21**, 254013.
34. Tsuruoka, T., Terabe, K., Hasegawa, T., and Aono, M. (2012) Temperature effects on the switching kinetics of a Cu-Ta₂O₅-based atomic switch. *Nanotechnology*, **22**, 254013.
35. Tsuruoka, T., Terabe, K., Hasegawa, T., Valov, I., Waser, R., and Aono, M. (2012) Effects of moisture on the switching characteristics of oxide-based, gapless-type atomic switches. *Adv. Funct. Mater.*, **22**, 70.
36. Torrezan, A.C., Strachan, J.P., Modeiros-Ribeiro, G., and Williams, R.S. (2011) Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, **22**, 485203.
37. Tsuruoka, T., Hasegawa, T., Valov, I., Waser, R., and Aono, M. (2013) Rate-limiting processes in the fast SET operation of a gapless-type Cu-Ta₂O₅ atomic switch. *AIP Adv.*, **3**, 032114.
38. Tsuruoka, T., Hasegawa, T., Terabe, K., and Aono, M. (2012) Conductance quantization and synaptic behavior in a Ta₂O₅-based atomic switch. *Nanotechnology*, **23**, 435705.
39. Jameson, J.R., Gilbert, N., Koushan, F., Saenz, J., Wang, J., Holmer, S., and Kozicki, M.N. (2011) One-dimensional model of the programming kinetics of conductive-bridge memory cells. *Appl. Phys. Lett.*, **99**, 063506.
40. Tapperzhofen, S., Valov, I., and Waser, R. (2012) Quantum conductance and switching kinetics of AgI-based microcrossbar cells. *Nanotechnology*, **23**, 145703.
41. Wu, S., Tsuruoka, T., Terabe, K., Hasegawa, T., Hill, J.P., Ariga, K., and Aono, M. (2011) A polymer-electrolyte-based atomic switch. *Adv. Funct. Mater.*, **21**, 93.
42. Mohapatra, S., Tsuruoka, T., Hasegawa, T., Terabe, K., and Aono, M. (2012) Flexible resistive switching memory using inkjet printing of a solid polymer electrolyte. *AIP Adv.*, **2**, 022144.
43. Xie, F.-Q., Nittler, L., Obermair, C., and Schimmel, T. (2004) Gate-controlled atomic quantum switch. *Phys. Rev. Lett.*, **93**, 128303.
44. Xie, F.-Q., Maul, R., Obermair, C., Wenzel, W., Schon, G., and Schimmel, T. (2010) Multilevel atomic-scale transistors based on metallic quantum point contacts. *Adv. Mater.*, **22**, 2033.
45. Banno, N., Sakamoto, T., Iguchi, N., Kawaura, H., Kaeriyama, S., Mizuno, M., Terabe, K., Hasegawa, T., and Aono, M. (2006) Solid-electrolyte nanometer switch. *IEICE Trans. Electron.*, **E89-C** (11), 1492.
46. Hasegawa, T., Itoh, Y., Tanaka, H., Hino, T., Tsuruoka, T., Terabe, K., Miyazaki, H., Tsukagoshi, K., Ogawa, T., Yamaguchi, S., and Aono, M. (2011) Volatile/nonvolatile dual-functional atom transistor. *Appl. Phys. Express*, **4**, 015204.
47. Sakamoto, T., Iguchi, N., and Aono, M. (2010) Nonvolatile triode switch using electrochemical reaction in copper sulfide. *Appl. Phys. Lett.*, **96**, 252104.
48. Wang, Q., Itoh, Y., Tsuruoka, T., Hasegawa, T., Watanabe, S., Yamaguchi, S., Hiramoto, T., and Aono, M. (2014)

- Two types of on-state observed in the operation of a redox-based three-terminal device. *Key Eng. Mater.*, **596**, 111.
49. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) Missing memristor found. *Nature*, **453**, 80.
50. Hu, G., McDaniel, M.D., Posadas, A., Demkov, A.A., Ekerdt, J.G., and Yu, E.T. (2014) High controllable and stable quantized conductance and resistive switching mechanism in single-crystal TiO₂ resistive memory on silicon. *Nano Lett.*, **14**, 4360.

19

Scaling Limits of Nanoionic Devices

Victor Zhirnov and Gurtej Sandhu

19.1

Introduction

For decades, the economics of the semiconductor industry has been driven by scaling the dimensions of complementary metal-oxide-semiconductor (CMOS) transistors. Although this strategy has been successful, the limits to scaling have prompted some to look for technologies “beyond CMOS” [1]. The industry has expressed particular interest in technologies that do not use electrons as information-bearing particles [2]. In response, researchers have proposed numerous systems based on alternative mechanisms. Nanoionic devices such as redox-based resistive random access memory (ReRAM) is one such alternative.

This chapter is intended to provide an expository, physics-based, framework for the estimation of the performance potential and physical scaling limits for nanoionic devices, with an emphasis on ReRAM. The approach taken seeks to provide physical insight into those parameters and those physical effects that define device performance and scaling properties.

19.2

Basic Operations of ICT Devices

Information can be defined as a quantitative measure of distinguishability of a physical subsystem from its environment [3]. Information of arbitrary kind and amount can be represented by the combination of just two distinguishable states (known as *binary states* and often marked as state 0 and state 1). A system with two distinguishable states forms the basic Information and Communication Technologies (ICT) element (*binary switch or memory cell*) shown in Figure 19.1. It consists of the following: (i) two states 0 and 1, which are equally attainable and distinguishable; (ii) a means to control the change of the state (WRITE operation);

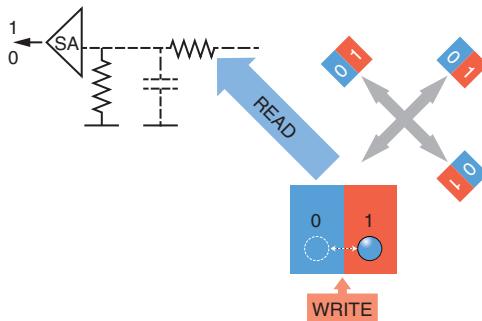


Figure 19.1 An abstract binary ICT element.

(iii) a means to detect the state (READ operation); and/or (iv) a means to communicate with other binary switches (TALK operation).

Information-processing systems represent system states in terms of physical variables. One way to create physically distinguishable states is by the *presence or absence* of material particles in a given location. Examples are electrons in transistor gates, or atoms/ions in nanoionic devices. Essential requirements for the implementation of an ICT device are the ability to *move* the particles from *state 0* to *state 1* and from *state 1* to *state 0*, when an external WRITE signal is applied, and conversely, the particles must remain in their position for a sufficiently long time. For example, a typical practical requirement for a nonvolatile memory element is the state lifetime (the retention time) $t_r \sim 10$ years $\sim 3 \times 10^8$ s.

In low-energy ICT systems, the operating voltage has to be decreased, and the most advanced logic circuits currently operate at $V = 0.8 - 1$ V. The operating frequency of the current baseline logic circuits is in the gigahertz range, which equates to switching speed of < 1 ns. It is desirable that the operating voltage and the speed of operation of memory be in the same range as logic (for the reasons of energy saving and operational compatibility), and, therefore, one needs to *detect* the presence/absence of the particles in, for example, the *state 1* by a fast (\sim ns) READ operation. In a typical READ operation, a read voltage V_r is applied to a communication line, and a voltage V_{sense} is sensed at the load resistor R_L (see Figure 19.2) and compared with a reference voltage. The discrete nature of electrical charge sets a fundamental restriction on the minimal current needed for a fast and reliable reading. The read current passing through the load resistor in Figure 19.2 is

$$I_r = \frac{q}{t_r} = \frac{Ne}{t_r}, \quad (19.1a)$$

and the minimal current, when only one electron passes through the resistor during the read interval is

$$I_{r\min} = \frac{e}{t_r} \quad (19.1b)$$

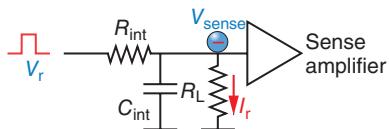


Figure 19.2 A typical READ operation in an ICT element.

For example, if the specified read time is $t_r \sim 1$ ns, the minimal read current is ~ 100 pA, which corresponds to only one electron (on average) passing through the load resistor R_L during the read interval. For such small numbers, the inevitable statistical fluctuations can easily lead to an erroneous result. The limits of *READ operation*, therefore, can be assessed based on the “margin of error” for each state: if N is the average number of electrons, injected in the READ circuit during the read interval t_r , the margin of error is given by the Poisson distribution as a standard deviation of $\approx \pm\sqrt{N}$, and the corresponding relative error can be estimated as

$$\delta \sim \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}} \quad (19.2)$$

For example, if $N = 1$, the error $\delta = 100\%$, then for a reliable reading, larger number of electrons is needed. From Eqs. (19.1b) and (19.2):

$$\delta \sim \sqrt{\frac{e}{I_r t_r}} \quad (19.3)$$

For example, if the specified read time and error margin are, respectively, $t_r \sim 1$ ns and $\delta \sim 1\%$, the read current should be > 1 μ A.

In summary, the essential properties of an advanced high-performance ICT device are as follows:

Operating voltage	~ 1 V
Read time	~ 1 ns
Read current	> 1 μ A
State lifetime	$\sim 3 \times 10^8$ s (nonvolatile memory applications)

In the following sections, the physics of nanoionic devices are analyzed and the scaling limits of nanoionic switching mechanisms that can satisfy the above requirements are explored.

19.3

Minimal Nanoionic ICT

19.3.1

Switching Mechanisms and the Material Systems

Examples of nanoionic memory devices [4] have been introduced in the literature with different names including conductance change memories [5], atomic/ionic

switches [6, 7], and redox-based resistive memories [8]. The operation of nanoionic devices is based on the *change in resistance* of a metal–insulator–metal (M-I-M) structure. The cell resistor has two stable values: high (OFF), R_{off} , and low (ON), R_{on} . Switching between R_{off} and R_{on} occurs by applying an electrical bias to the resistor.

Mechanisms for resistive switching are based on atomic re-arrangements in a material caused by ion (cation or anion) migration combined with redox processes involving the electrode material or the insulator material, or both [8, 9]. Resistive switching is due to formation of a conductive path within (semi)insulating matrix. In the context of this chapter, we cover the two most relevant classes of bipolar resistive switching: Firstly, it could be, in the simplest case, that an atomic filament (AF) is formed by metal atoms by an electrochemical metallization process (ECM, Figure 19.3a), such as growth of Ag dendrites in silver sulfide (Ag_2S) [6], sometimes also called conducting bridge (CB) in the literature. The second mechanism of resistive switching is modulation of the interface resistance between a metal electrode and the insulating/semiconducting matrix. This can be achieved by re-arrangement of charged defects/impurities near the interface between the semiconductor matrix and an electrode (Figure 19.3b). As explained in Chapter 1, this mechanism is typically found in valence change memories (VCMs). In this case, increased concentration of ionic interface charges change the width and the height of the interface (contact) barrier, and as result the contact resistance changes. The interface area may be limited to the contact point of a filament as reported for several materials, all of which are metal oxides, for example, TiO_2 or HfO_2 or the area may extend to the entire electrode area (see Chapter 11). Within this chapter, we call both variants the interface controlled resistance (ICR) switching.

Note that in nanoionic devices, the state is created by moving atoms/ions (WRITE operation), while electron transport is used for sensing the state (READ operation). In the following sections, the mechanisms of both atomic and

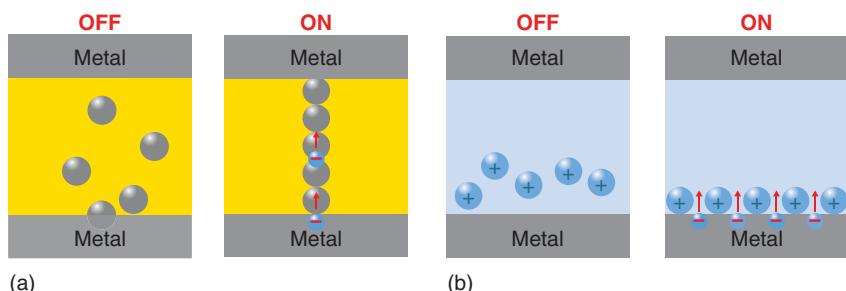


Figure 19.3 Two basic mechanisms of resistance switching due to atomic rearrangements are discussed in this chapter: (a) atomic filament (AF) as typically found in

electrochemical metallization memories (ECM) and (b) interface-controlled resistance (ICR) as typically found in valence change memories (VCMs).

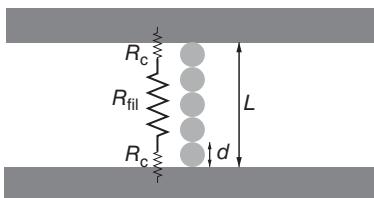


Figure 19.4 Single metal atom bridge.

electronic transport in solids are considered, and the scaling limits of resistive nanoionic devices are investigated based on the conductance and stability of the nanostructures consisting of a few atoms.

19.3.2

Atomic Filament: Classical and Quantum Resistance

19.3.2.1 Classical Resistance

Consider a single metal atom filament bridging two electrodes (Figure 19.4). The total resistance of the single-atom filament bridge can be obtained as

$$R_{\text{bridge}} = R_{\text{fil}} + 2R_C \quad (19.4)$$

where R_{fil} is the resistance of the “filament” and R_C is the contact (constriction/spreading) resistance of a contact between the filament and the electrodes. The filament resistance can be calculated as

$$R_{\text{fil}} = \rho^* \frac{L}{d^2} \quad (19.5)$$

where L is the length of the filament, d is the filament diameter (in our case close to the atomic diameter, s), and ρ^* is the resistivity of metal nanowire. The Fuchs–Sondheimer approximation can be used to calculate the resistance of nanoscale metal wires:

$$\rho^* = \rho_0 \frac{1 - \alpha}{1 + \alpha} \frac{\lambda_0}{d} \quad \text{for } d \ll \lambda_0, \quad (19.6)$$

where λ_0 is the bulk metal mean free path, derived from the electron concentration in metal using the Sommerfeld model [10–12], ρ_0 is the bulk resistivity of the metal, and α is specularity, that is, the probability of an electron being scattered elastically at the side surface of the wire.

The effective diameter of an atom, s , can be estimated based on the atomic density in a bulk solid metal, n_{at} :

$$s \approx n_{\text{at}}^{-\frac{1}{3}} \quad (19.7)$$

In general, the values of r_0 , λ_0 , and s are material-specific parameters. The corresponding values of these parameters for silver are given in Table 19.1. (Silver (Ag) is chosen in this chapter as an example model material.)

The length of the filament can be expressed as a function of the number of the constituting metal atoms, N :

$$L = Ns \quad (19.8)$$

Table 19.1 Parameters of bulk silver used for minimal atomic bridge estimates.

Parameter	Numerical value
Atomic density	$n_{\text{at}} = 5.83 \times 10^{22} \text{ cm}^{-3}$
Effective atomic diameter	$s = 0.258 \text{ nm}$
Bulk resistivity	$\rho_0 = 15.8 \text{ n}\Omega \text{ m}$
Electron mean free path (at 300 K)	$\lambda_0 = 54 \text{ nm}$

Substituting Eqs. (19.6–19.8) into Eq. (19.2) gives (assuming $\alpha = 0$)

$$R_{\text{fil}} = \rho_0 \frac{\lambda_0}{s} \frac{L}{s^2} = \rho_0 \frac{\lambda_0}{s^2} N \quad (19.9)$$

The constriction (spreading) resistance exists at points of contact between the filament and the electrodes, and is given by [13, 14]

Note that in Eq. (19.10) the bulk metal resistivity is used.

$$R_C = \frac{\rho_0}{d} \quad (19.10)$$

Combining Eqs. (19.9) and (19.10), we obtain

$$R_{\text{bridge}} = \rho_0 \left(\frac{\lambda_0}{d^2} N + \frac{2}{d} \right) \quad (19.11a)$$

The smallest possible bridge corresponds to $N = 1$, that is, single-atom contact. The “classical” resistance Eq. (19.11a), in this case, becomes

$$R_{\text{bridge}} = \rho_0 \left(\frac{\lambda_0}{s^2} + \frac{2}{s} \right) \quad (19.11b)$$

The numerical result from evaluating Eq. (19.11b) using parameters from Table 19.1 is $R_{\text{bridge}} = 12.94 \text{ k}\Omega$.

19.3.2.2 Quantum Resistance

The best-case conductance through a single atom bridge shown in Figure 19.5 can be estimated as follows. Assume that a monovalent atom such as Ag, Cu, or Au contributes one conductance electron and consider an elementary act of electrical conductance where one electron passes from electrode A to electrode B (A and B are not shown in Figure 19.5) with a potential difference between the electrodes of V_{AB} (the corresponding energy change $\Delta E = eV_{AB}$). The speed of the passage process is bounded by the Heisenberg energy–time relation

$$\Delta E \cdot \Delta t \geq \frac{\hbar}{2} \quad (19.12a)$$

from which the minimum passage time (the Heisenberg time) Δt_H is

$$\Delta t = \frac{\hbar}{2\Delta E} = \frac{\hbar}{2eV} \quad (19.12b)$$

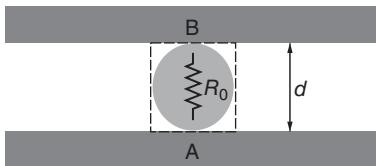


Figure 19.5 Single-atom bridge.

The current between two electrodes through the atom (involves only one electron at a time) is

$$I_{AB} = \frac{e}{\Delta t} \quad (19.13)$$

Substituting Eq. (19.12b) into Eq. (19.13), and taking into account Ohm's law, that is, $I = V/R$, we obtain

$$I_{AB} = \frac{2e^2}{h} \cdot V = \frac{V}{R_0}, \quad (19.14)$$

where

$$R_0 = \frac{h}{2e^2} = 12.95 \text{ k}\Omega \quad (19.15a)$$

is *quantum resistance*. A related parameter is *quantum conductance*:

$$G_0 = \frac{1}{R_0} = \frac{2e^2}{h} \quad (19.15b)$$

The quantum resistance/conductance sets the limit on electrical conductance in a one-electron channel *in the absence of barriers*.

$$I_0 = \frac{V}{R_0} = \frac{V}{12.95 \text{ k}\Omega} \quad (19.16)$$

Experiments with the single atom contacts indeed have demonstrated that the minimum contact resistance is approximately 12.9 kΩ [15].

Therefore, it has been shown that both classical and quantum models yield a very similar result for the minimal resistance of a single-atom bridge. In the following, Eq. (19.11a) is used for detailed calculations of electron transport in continuous atomic chains.

19.3.2.3 Conductance in the Presence of Barriers

If a barrier is present in the electron transport system, the conductance will be decreased due to the barrier transmission probability $p < 1$. The electrical conductance in the presence of barrier is obtained by multiplying the barrierless quantum conductance (Eq. (19.15b)) by the barrier transmission probability p :

$$G = \frac{1}{R} = G_0 \cdot p \quad (19.17)$$

Equation (19.17) is a form of the *Landauer formula* [16] for a one-electron conductive channel.

Assume that the single-atom bridge in Figure 19.3 has the minimum resistance R_0 (i.e., barrierless transport). Now, if the atom is removed, the two electrodes are

separated by the gap of length d , and thus a barrier is formed. The change in the interelectrode resistance with and without the bridging atom (R_{on} and R_{off}) can be obtained from Eq. (19.17) as

$$\frac{R_{\text{off}}}{R_{\text{on}}} = \frac{G_0}{G_0 P} = \frac{1}{p} \quad (19.18)$$

The Landauer formalism (Eq. (19.17)) allows for analysis of electron transport in the presence of a barrier in which all different mechanisms of electron transport (conductance) can be expressed through the transmission coefficient p , which is the probability that an electron can transmit through a medium. In the simplest case of one single-electron transmission channel, the electron current is expressed as

$$I_1 = \frac{2e^2}{h} \cdot V \cdot p, \quad (19.19)$$

In the case of transmission through many parallel electron channels, the total current is obtained as a product of I_1 and the number of parallel channels M :

$$I_M = M \cdot I_1 = \frac{2e^2}{h} \cdot M \cdot V \cdot p, \quad (19.20a)$$

which can also be rewritten as

$$I_M = \frac{2e^2}{h} \cdot M \cdot V \cdot p = e \cdot \frac{2e}{h} M V P = e f_0 \cdot p, \quad (19.20b)$$

where f_0 is the *attempt frequency* – the rate at which electrons available for transmission hit the barrier. For practical cases of multiple-channel transmission, the attempt frequency can be calculated based on electron distribution functions in solids.

The two mechanisms of the barrier transport are over-barrier thermal transition and through-barrier tunnel transition. The over-barrier transitions occur due to a permanent supply of thermal energy to the system that occurs via mechanical vibrations of atoms (phonons) and via the thermal electromagnetic field of photons (background radiation). The probability of over-barrier transition is

$$p_{\text{o-b}} = \exp \left(-\frac{E_b}{k_B T} \right) \quad (19.21)$$

Also, for sufficiently thin barriers, “through-barrier” tunneling transition can occur, which is a quantum mechanical effect. The probability of tunneling is

$$p_{\text{tun}} \sim \exp \left(-\frac{2\sqrt{2m}}{\hbar} \cdot a \cdot \sqrt{E_b} \right) \quad (19.22)$$

This equation also emphasizes the parameters controlling the tunneling process. They are the barrier height E_b and barrier width a as well as the mass m of the information-bearing particle. Other forms of the tunneling equation can be obtained from Eq. (19.22) by substituting a and E_b by the corresponding average values in nonuniform barrier structures. Useful analytical formulae to calculate barrier transitions are the Richardson equation for over-barrier transitions and Fowler–Nordheim equation for tunneling (discussed in Section 19.3.3).

19.3.2.4 Barriers in Atomic Gaps: Nonrectangular Barrier

Consider a vacuum gap of length a between two metal electrodes. There will be an energy barrier formed at the metal–vacuum interface on both electrodes, as discussed in Appendix A. For larger gaps, if no voltage is applied across the gap, $V_{\text{gap}} = 0$, the barrier can be approximated by a *rectangular* barrier with height equal to the metal work function, ϕ_0 (Figure 19.6a).

For smaller gaps, the shape of the barrier changes, by reducing barrier height and inducing corner rounding due to image forces as is discussed in Appendix A. The barrier thus becomes inherently nonrectangular. Extension of Eq. (A3) (see Appendix A) to two-sided barrier results in [17]

$$E_b = \phi_0 - \frac{e^2}{16\epsilon_0 Kx} - eFx - \frac{e^2}{8\pi\epsilon_0 K} \sum_{n=1}^{\infty} \left(\frac{na}{(na)^2 - x^2} - \frac{1}{na} \right), \quad (19.23)$$

where K is the dielectric constant of the matrix material.

The last term in Eq. (19.23) accounts for the *interface-to-interface interaction*. A useful analytical approximation of Eq. (19.23) was obtained by Simmons [17] in the form

$$E_b(x) = \phi(x) \approx \phi_0 - \frac{e^2 \ln 2}{16\pi\epsilon_0 K_{\text{opt}}} \cdot \frac{a}{x(a-x)} - \frac{eV_{\text{gap}}x}{a} \quad (19.24)$$

As shown in Figure 19.6a,b, for very small gaps, for example, $a < 2$ nm, the barrier height strongly depends on the gap voltage, V_{gap} . In addition, the effective barrier width for electrons, a_{eff} , is smaller than the interelectrode “metallurgical” gap, a . In the following, when a vacuum gap is considered, it is assumed that $\phi_0 = 4.7$ eV (work function of Ag) and $K = 1$, and when the gap is filled by an insulating material, $\phi_0 = 1$ eV and $K = 7$.

19.3.2.5 Transmission through Atomic Gaps

Assume that the single-atom bridge in Figure 19.4 has the minimum resistance R_0 (i.e., barrierless transport). If one atom is removed, the two electrodes are separated by an atomic gap, and thus a barrier is formed. If a barrier is present in

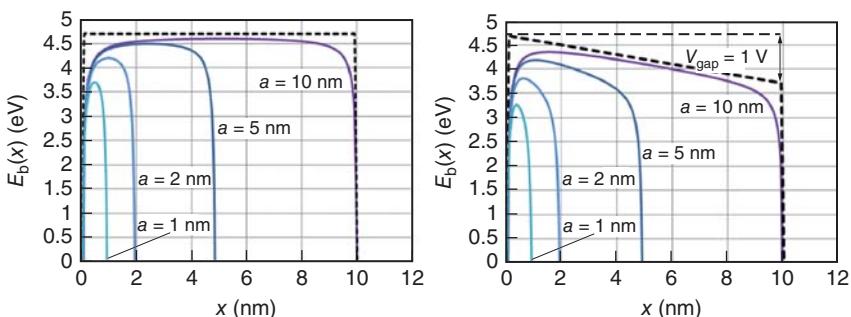


Figure 19.6 Barrier profiles in small *Me–vacuum–Me* gaps ($\phi_0 = 4.7$ eV, $K = 1$) for four different gap lengths (1, 2, 5, and 10 nm): (a) unbiased gap, $V_{\text{gap}} = 0$ and (b) biased gap, $V_{\text{gap}} = 1$ V.

the electron transport system, the conductance will decrease due to the barrier transmission probability $p < 1$.

The minimum number of atoms needed in a conductive chain to provide sensing margin can now be estimated. This can be done with following thought experiment: the atoms from the chain in Figure 19.1 are removed one by one, and the resulting change in conductance is calculated. The change in the conductance will be due to barriers formed in such subnanometer gaps as shown in Figure 19.6.

From the numerical values given in the table shown in Figure 19.7, a three-atom gap is sufficient to obtain both a sufficiently large ON current, a reasonably large resistance ON/OFF ratio to satisfactorily differentiate the state of a nanoionic device, for example, the ReRAM cell.

19.3.3

Interface Controlled Resistance (ICR)

Another mechanism of redox-based resistive switching is the re-arrangement of charged defects/impurities near the interface between the matrix and an electrode. In this case, increased concentration of interface charges reduces the width and the height of the interface (contact) barrier, resulting in contact resistance decrease. A simple analysis of this scenario can be performed based on the Mott–Schottky theory of contacts between a metal and a nonmetal [18].

The interface resistance mechanism of nanoionic devices was reported for several materials, all of which are metal oxides, such as TiO_2 [19, 20], HfO_2 [21], ZrO_2 [22], ZnO [23], and Ta_2O_5 [24]. For numerical estimates in this chapter, material parameters of titanium oxide TiO_x are used. It is important to note that the metal oxide films are usually nonstoichiometric due to an excess of metal ions or deficiency of oxygen ions [25]. This nonstoichiometry appears to play a key role

N_{at}	a (nm)	a_{eff} (nm)	E_{bmax} (eV)	E_{beff} (eV)	$I_{\text{on}}/I_{\text{off}}$
1	0.258	0.160	0.38	0.26	2.30
2	0.516	0.426	0.66	0.50	11
3	0.714	0.680	0.75	0.59	73

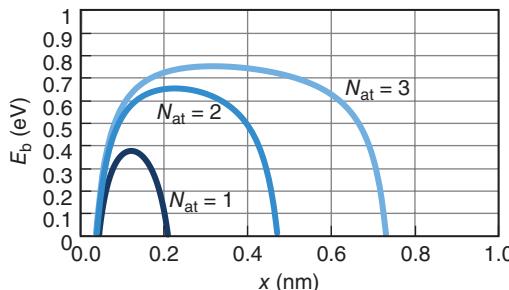


Figure 19.7 Barriers formed as result of removal of 1, 2, and 3 atoms from a single-atom metallic chain embedded in an insulating matrix ($\phi_0 = 1$ eV, $K = 7$). (The corresponding numerical barrier parameters are shown.)

Table 19.2 Material parameters determining the interface barrier properties.

	HfO_2	TiO_2
Relative static dielectric constant, K	14–34	7–114
Band gap, E_g	5.1 eV	3.2–3.8 eV
Donor concentration, N_d	10^{18} cm^{-3}	10^{20} cm^{-3}
Effective electron mass, m^*	$0.15m_0$	$1m_0$

in the electronic properties of the oxides, since the resulting lattice point defects (e.g., vacancies or interstitial atoms) can electrically act as donors or acceptors. Due to these nonstoichiometric defect levels, the materials often behave as doped semiconductors, and can be described in the framework of a classical semiconductor model. In the past, there were attempts to describe these materials as a special class, called *mixed ionic-electronic conductors* or *chemiconductors* [25, 26] to emphasize three important differences from classical semiconductors: (i) the ions forming donor and acceptor levels in chemiconductors are primarily due to composition variation, (ii) the ions can move under electrical fields, while in semiconductors the *dopants* (i.e., donors and acceptors) do not change their positions, and (iii) the distribution of the “dopants” (donors or acceptors) are *inherently nonuniform*, especially near the interfaces. Except for these caveats, chemiconductors (i.e., nonstoichiometric metal oxides) can be considered as classic semiconductors. For example, the electrical properties of their interfaces with metal electrodes can be described using the Mott–Schottky model [25, 26] suggesting formation of interface energy barriers (known as *Schottky barriers*). The Mott–Schottky theory (discussed later) describes the electrical properties of interfaces in terms of the space-charge formation in the interface region, which depends on macroscopic materials parameters such as dielectric constant K and donor concentration N_d . In addition, electron transport across the interface depends on the effective electron mass m^* . The material parameters determining the interface barrier properties are shown in Table 19.2.

19.3.3.1 Electrical Properties of Material Interfaces

When two different materials are brought in contact, an energy barrier is commonly formed at the interface. The origin of the interface energy barrier is different concentration and distribution of electrical charges in dissimilar materials. In an extreme case of an interface between a metal (maximum concentration of electrons) and vacuum (zero concentration of electrons), larger barriers are formed often referred to as the *work function*, φ . Now, if two solid materials are brought in contact, they will exchange electrons, and the resulting barrier height can be estimated as the difference between the work functions of the first and the second materials, that is, $\varphi_1 - \varphi_2$, called the contact potential difference (measured in volts). In the case of contacts between a metal and a semiconductor or an insulator, the interface barrier height, Φ , is the difference between the work function

of the metal, φ_M , and the *electron affinity* of the semiconductor/insulator, χ_s (if interface states are neglected):

$$\Phi = \varphi_M - \chi_s \quad (19.25)$$

Consider the contact between a metal and an n-type semiconductor with a concentration of dopants N_d . As a result of the electron exchange, an increased negative charge will accumulate on the metal side of the interface (in an infinitely thin layer for an ideal metal). Due to the charge neutrality requirement, this negative charge must be compensated by an equal positive charge on the semiconductor side (formed by the *ionized* dopants). Since the concentration of charge carriers in semiconductors is much lower than in metals, the positive charge is formed within some extended layer on the semiconductor side known as a *depletion layer* of width W . The potential profile in the semiconductor material near the interface can be obtained by solving the *Poisson equation*:

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\rho(x)}{\epsilon_0 K} \quad (19.26)$$

For the simplest model scenario, the Poisson equation (19.26) can be solved assuming uniform distribution of the ionized dopants in the interface layer of width W , and zero net charge outside the interface layer:

$$\rho(x) = \text{const} = eN_d^+ \quad \text{for } 0 < x < W$$

$$\rho(x) = 0 \quad \text{otherwise}$$

We thus can write

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\rho(x)}{\epsilon_0 K} = \frac{eN_d^+}{\epsilon_0 K} \quad (19.27)$$

The boundary conditions for integration are $eV(0) = \Phi$, $V(W) = 0$, and $dV/dx|_{x=W} = 0$ (zero electric field outside the interface depletion layer). The first integration of Eq. (19.27) gives the interface electric field distribution $F(x)$:

$$-\frac{\partial V}{\partial x} = -F(x) = \frac{eN_d^+}{\epsilon_0 K} \cdot x + C_1 \quad (19.28a)$$

The integration constant C_1 can be found from the boundary condition $F(W) = 0$:

$$\frac{eN_d^+}{\epsilon_0 K} W + C_1 = 0 \quad (19.28b)$$

$$C_1 = -\frac{eN_d^+}{\epsilon_0 K} W \quad (19.28c)$$

The second integration results in the interface potential profile:

$$V(x) = -\frac{eN_d^+}{\epsilon_0 K} \frac{x^2}{2} + \frac{eN_d^+}{\epsilon_0 K} Wx - C_2 \quad (19.28d)$$

The integration constant C_2 can be found from the boundary condition $eV(0) = \Phi$, which results in

$$C_2 = -\frac{\Phi}{e} \quad (19.28e)$$

The resulting potential distribution near the interface is

$$V(x) = -\frac{eN_d^+}{\epsilon_0 K} \left(\frac{x^2}{2} - Wx \right) - \frac{\Phi}{e} \quad (19.29)$$

The potential distribution $V(x)$ near the interface (Eq. (19.29)) is plotted in Figure 19.8. The zero-bias depletion width W_0 is straightforward to derive from Eq. (19.29) using the condition $V(W) = 0$, which results in

$$W_0 = \sqrt{\frac{2\epsilon_0 K \Phi}{e^2 N_d^+}} \quad (19.30a)$$

If an external bias V is applied to the interface, the depletion width of the biased interface from Eq. (19.29) is

$$W = \sqrt{\frac{2\epsilon_0 K (\Phi \pm eV)}{e^2 N_d^+}} \quad (19.30b)$$

where plus corresponds to the “reverse” bias and minus to the “forward” bias.

Equations (19.29) and (19.30a,b) represent the parabolic approximation of the interface potential profile, which is most commonly used [18]. For a simple qualitative analysis, a linear approximation can also be used [27].

In the linear model, the barrier-forming potential bending is entirely given by the constant interface electric field F , which, in turn, depends only on the potential change along distance W (interface depletion width):

$$eF \sim \frac{\Phi \pm eV}{W} \quad (19.31)$$

The interface electric field (Eq. (19.31)) acts both on the electrons and the ionized donors: the negatively charged electrons are repelled from the interface (thus resulting in *depletion*); on the contrary, the positively charged donors are attracted closer to the interface.

From Eq. (19.30b), the depletion width at the metal–semiconductor (M–S) interface depends on the concentration of ionized impurities, near interface. The current is then modulated by changing the impurity concentration and, therefore, the depletion width at the M–S interface.

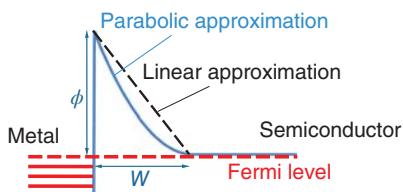


Figure 19.8 Potential distribution near metal–semiconductor interface.

If *finite* lateral dimensions of a three-dimensional semiconductor structure are considered, the side interfaces can also affect the current flow. In simplest model case, these side interfaces are formed between the semiconductor surface and vacuum (S-V interface). In practice, interfaces with passivation insulator, for example, $\text{TiO}_2/\text{SiO}_2$, are representative. Band bending/barrier formation usually occurs at these interfaces, and they need to be taken into account. The band bending results in either depletion (bent up) or accumulation (bent down), and correspondingly, a layer with lower (depletion) or higher (accumulation) conductivity of width W_{SV} is formed as shown in Figure 19.9. Therefore, in addition to the depletion W_{MS} layer aligned with the direction of current (“active” interface, modulated by external stimulus), there is a lateral depletion layer W_{SV} perpendicular to the current flow (“passive” interface, which remains more or less stable during device operation). This “passive” side interface may also affect the total current. If a depletion high-resistive layer of width W_{SV} is formed, the effective cross-sectional area for modulated current flow is decreased. In the case of an accumulation low-resistive layer, a parasitic surface resistor will be formed in parallel with the resistive memory element. In the treatment discussed in the following text, a depletion layer of width W is considered, which is a typical case of n-type semiconductor. Side depletion effectively reduces the conductive cross-sectional area of the materials system (the blue-colored central region in Figure 19.9).

19.3.3.2 Contact Resistance in a M–S (M–I) Structure

The barrier height and width determine electron transport through the barrier and thus the contact resistance. In the following, it is assumed that the contact resistance dominates the total resistance of the structure.

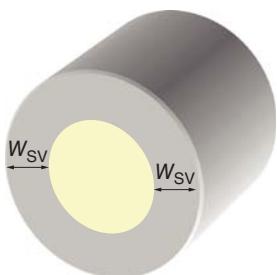
A dominant mechanism of the ohmic conduction through a typical M–S interface is tunneling [18]. For a simple model of a triangular barrier in Figure 19.8 (dashed line), the charge flow through the barrier can be calculated using the Fowler–Nordheim (FN) equation:

$$J = a_{FN} \frac{F^2}{\Phi} \exp\left(-b_{FN} \frac{\Phi^{3/2}}{F}\right), \quad (19.32)$$

where a and b are constants, and

$$a_{FN} = \frac{e^2}{8\pi h} \quad (19.33a)$$

Figure 19.9 Reduction of the effective conduction area due to side depletion.



and

$$b_{FN} = \frac{4}{3} \frac{\sqrt{2em^*}}{\hbar} \quad (19.33b)$$

Note that the barrier height in Eq. (19.32), Φ^* , is reduced compared with Φ due to image force effect in high electric fields: $\Phi^* = \Phi - \Delta\Phi$, where $\Delta\Phi$ is the calculated using Eq. (A5) (see Appendix A).

Equation (19.32) was used to calculate the interface current as a function of the dopant concentration. A characteristic parameter is interface-specific contact resistance:

$$R_c = \frac{V}{J} (\Omega \text{cm}) \quad (19.34)$$

Calculations using Eq. (19.34) for parameters of TiO_2 (listed in Table 19.2) are shown in Figure 19.10. To take into account the effect of side depletion in the case of extreme scaling, the current injected from the metal contact through the interface was calculated as

$$I = J \cdot A = J \cdot (L - 2W_0)^2 \quad (19.35)$$

where A is the effective conductive area of a semiconductor structure with total spatial dimensions L . The effective conduction area is reduced due to side depletion, as discussed in previous section (Figure 19.9). Figure 19.10 shows calculated interface contact resistance with and without the side depletion effects (data for silicon are also shown for reference). From Eq. (19.35), in order to satisfy the condition of constant minimal read current of $1 \mu\text{A}$, the current density J must increase with scaling, that is, decrease of spatial dimensions L . This implies that the interface critical dopant concentration must increase with scaling, as is shown in Figure 19.11.

As shown in Figure 19.11, for memory cell sizes less than 10 nm , very high concentrations of the interface doping in the order of $N_d > 10^{21} \text{ cm}^{-3}$ are needed. The limit on the maximum dopant concentration for the case of TiO_2 can be estimated based on the formation of the Magnéli phases $\text{Ti}_n\text{O}_{2n-1}$ ($4 \leq n \leq 10$). A concentration of $N_d = 3 \times 10^{21} \text{ cm}^{-3}$ corresponds to $n = 10$. Therefore, the formation of the

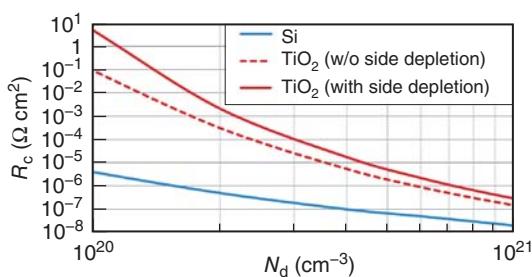


Figure 19.10 Calculated interface contact resistance of Me– TiO_2 barriers ($\Phi = 0.85 \text{ eV}$).

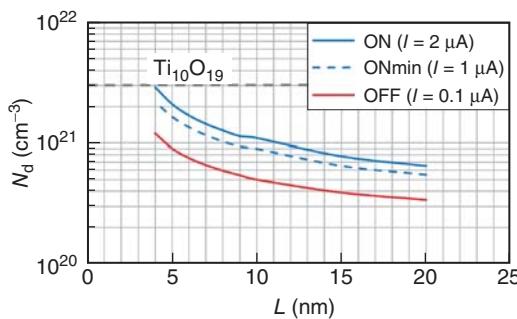


Figure 19.11 Interface critical dopant concentration as a function of the ICR memory cell size.

Magnéli phase $\text{Ti}_{10}\text{O}_{19}$ can be considered as the limit for operation of the modulated Schottky barrier interface resistance mechanism. This limit is reached for cell size around $L = 4 \text{ nm}$.

Equation (19.35) was also used to investigate changes of the interface current as a function of the number of interface dopants. For retention estimates, the question to be answered is how many dopant atoms, Δn , does need to move from the interface before the ON state is lost. To address this question, “operational” parameters were set as $I_{\text{ON}} \sim 2 \mu\text{A}$ and $\text{ON}/\text{OFF} \sim 20$, that is, twice as much as the minimum read current and minimum On/Off ratio specified in Section 19.2. With such settings, the state is considered “lost,” when the current through the structure $I < I_{\text{ONmin}} = 1 \mu\text{A}$.

First, the ON current was set to be $I_{\text{ON}} \sim 2 \mu\text{A}$ and the required (critical) concentration of the interface dopants $N_{d,\text{ON}}$ was calculated from Eqs. (19.32) to (19.35). Next, according to the “benchmark” specifications discussed at the end of Section 19.2, the resistance ratio was set to be $R_{\text{off}}/R_{\text{on}} > 10$, and therefore the OFF current $I_{\text{OFF}} < 0.1 I_{\text{ON}} \sim 100 \text{ nA}$. The corresponding concentration of the interface dopants $N_{d,\text{OFF}}$ was then calculated from (19.32) to (19.35).

Finally, the total number of the dopant atoms/defects (or oxygen vacancies in the case of TiO_2) can be calculated for ON and OFF states as follows:

$$n_{\text{on}} = N_{d,\text{ON}} \cdot L^2 \cdot W_{\text{on}} \quad (19.36a)$$

$$n_{\text{off}} = N_{d,\text{OFF}} \cdot L^2 \cdot W_{\text{on}} \quad (19.36b)$$

Thus, the number of atoms/defects, which need to be moved to/from the interface to enable the specified minimum ON/OFF ratio, is

$$\Delta n_{\text{on-off}} = n_{\text{on}} - n_{\text{off}} = L^2 W_{\text{on}} (N_{d,\text{ON}} - N_{d,\text{OFF}}) \quad (19.36c)$$

The calculation results are summarized in Table 19.3.

Table 19.3 Material parameters for TiO_2 , used for the interface contact resistance calculations.

Dielectric constant	
Static	$K_{\text{stat}} = 40$
Optical	$K_{\text{opt}} = 7$
Me/TiO ₂ barrier height	$\Phi = 0.85 \text{ eV}$
Tunneling effective mass	$1m_0$

19.3.4

Stability of the Minimal Nanoionic State

Formation of an AF implies alignment of atoms (e.g., metal atoms) between the electrodes to promote electron conduction. At equilibrium, atoms in solids are kept in their position because they are confined in a “well” between the barriers formed by chemical bonds. The barrier height E_a is often referred to as activation energy (e.g., for diffusion). Dimensions of both the well and the barrier can, to first order, be approximated by the interatomic distance s (Eq. (19.7)). As a result of thermal excitation, a confined atom vibrates around its equilibrium position with an average energy of $1/2k_B T$ per degree of freedom and it strikes the barrier with a frequency f_0 (thermal attempt frequency). In some cases, when its instantaneous energy exceeds E_a , the atom will jump over the barrier to another site. The rate of such transition f_{tr} is given by the Boltzmann probability (Eq. (19.21)):

$$f_{\text{tr}} = f_0 \cdot P = f_0 \exp\left(-\frac{E_a}{k_B T}\right) \quad (19.37)$$

where f_0 is the attempt frequency, determined by an average time τ_0 between barrier strikes, which is determined by the average thermal velocity, u , of the atom and the interatomic separation s (Eq. (19.7)):

The velocity u can be found from the kinetic energy relation:

$$\tau_0 \sim \frac{s}{u} \quad (19.38)$$

$$\frac{k_B T}{2} = \frac{mu^2}{2} \quad (19.39)$$

$$\tau_0 \sim s \sqrt{\frac{m_{\text{at}}}{k_B T}} \quad (19.40)$$

Assuming that the movable atoms are of silver ($s = 0.258 \text{ nm}$, $m_{\text{at}} = 108 \text{ a.u.m.} = 1.79 \times 10^{-25} \text{ kg}$) and $T = 400 \text{ K}$, Eq. (19.40) results in $\tau_0 \sim 1.5 \text{ ps}$ (or $f_0 \sim 7 \times 10^{11} \text{ Hz}$).

The state lifetime Δt can be defined through the probability that n atoms move out from the filament. Let p be the probability of “success” in one trial. The number of trials k during time interval Δt is

$$k = \Delta t \cdot f_0. \quad (19.41)$$

The probability that at least one atom will pass over the barrier E_a during a sampling time Δt is

$$\pi_k = 1 - (1 - p)^k \quad (19.42a)$$

And the probability for n atoms to escape the filament during the interval Δt is

$$\pi_{kn} = (1 - (1 - p)^k)^n \quad (19.42b)$$

Let, $\pi_{kn} = 1/2$, then from Eq. (19.42b):

$$\left(1 - \exp\left(-\frac{E_a}{k_B T}\right)\right)^{f_0 \cdot \Delta t} = 1 - 2^{-\frac{1}{n}}, \quad (19.43a)$$

or

$$\Delta t = \frac{1}{f_0} \frac{\ln\left(1 - 2^{-\frac{1}{n}}\right)}{\ln\left(1 - \exp\left(-\frac{E_a}{k_B T}\right)\right)} \quad (19.43b)$$

Now, consider the resistive switch in ON state (Figure 19.12a), with conductive filament formed with a single-atom chain. Let $E_a = 1$ eV, then the lifetime of one atom ($n = 1$) in the filament from Eq. (19.43b) will be about 4 s at $T = 400$ K and $\Delta t \sim 1800$ s at $T = 330$ K (Figure 19.12b). As discussed in Section 19.3.2.5, a three-atom gap is sufficient to obtain a reasonably large resistance ON/OFF ratio (>10). Repeating calculations for $n = 3$, one obtains $\Delta t \sim 9$ s at $T = 400$ K and $\Delta t \sim 4000$ s at $T = 330$ K. The state lifetime can be increased by adding parallel single-atom chains, to increase the redundancy. For example, if there are two touching parallel conductive chains, the conductance will be broken only if both atoms in the same level jump out of the conductive filament. To further increase redundancy, assume that minimal conductive bridge forms a 4 atoms \times 4 atoms \times 4 atoms cube, which is approximately 1 nm in size. According to Eq. (19.43b), such a structure can have a lifetime of $\sim 3 \times 10^8$ s (nonvolatile memory applications), if $E_a = 1.3$ eV ($T = 330$ K) or $E_a = 1.56$ eV ($T = 400$ K). A more detailed stability analysis of AFs can be found in Ref. [28].

This approach can be extended to the stability analysis of ICR structures of Figure 19.3b. In this case, it is convenient to use the time-dependent diffusion equation for a fixed number of ions [29]:

$$N(x, t) = \frac{S_0}{\sqrt{\pi D t}} \exp\left(-\frac{x^2}{\sqrt{4 D t}}\right), \quad (19.44a)$$

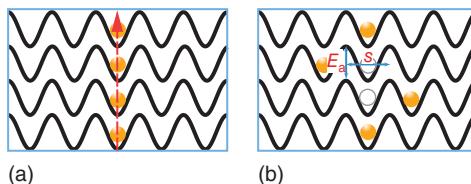


Figure 19.12 A schematic representation of an atomic conductive filament (AF) depicting barriers for moving atoms within the matrix: (a) OFF state and (b) ON state.

Table 19.4 Scaling-dependent parameters of an ICR cell: the interface dopant concentrations and the corresponding number of dopants that need to be moved to/from the interface to enable the ON/OFF switching.

(nm)	$n_{\text{on}} (\text{cm}^{-3})$ $= 2 \mu\text{A}$	$n_{\text{on min}} (\text{cm}^{-3})$ $= 1 \mu\text{A}$	$n_{\text{off}} (\text{cm}^{-3})$ $= 0.1 \mu\text{A}$	on	on min	off	Δ	$\Delta_{\text{on-off}}$	$\Delta_{\text{On-On min}}$	$V_{\text{on}} (1 \text{ V})$
20	6.5×10^{20}	5.5×10^{20}	3.4×10^{20}	567	480	296	271	87	2.18	
15	7.8×10^{20}	6.5×10^{20}	3.9×10^{20}	349	291	175	174	58	1.99	
10	1.1×10^{21}	9.0×10^{20}	5.0×10^{20}	185	151	84	101	34	1.68	
5	2.1×10^{21}	1.7×10^{20}	9.0×10^{20}	64	52	42	22	12	1.23	
4	2.9×10^{21}	2.2×10^{21}	1.2×10^{21}	49	37	20	29	12	1.05	

where D is the diffusion coefficient, and

$$D = \Delta x^2 f_0 \exp\left(-\frac{E_a}{k_B T}\right) \quad (19.44b)$$

In Eq. (19.44a), $N(x,t)$ is the ion concentration and S_0 is the initial “planar” concentration at the surface. When the initial state is ON, governed by the number of ions n_{on} within the interface depletion width W (see Table 19.4), then as a simplification, one can assume that initially all of the ions within one atomic jump distance ($\Delta x = s$) of the surface reside at the surface, thus $S_0 = n_{\text{on}}/L^2$.

The solution for the final ion count in the interface region is [28]

$$n_{\text{final}} = n_{\text{on}} \cdot \text{Erf}\left(\frac{W}{\sqrt{4D \cdot \Delta t}}\right) \quad (19.45)$$

In order to estimate the state lifetime (data retention time), Eq. (19.45) can be solved for Δt assuming the final ion count is known. The ON state is considered “lost” when $n_{\text{final}} = n_{\text{on min}}$ (see Table 19.4). For example, for a cell with $L = 10 \text{ nm}$, $n_{\text{on}} = 185$ and $n_{\text{on min}} = 151$. The retention time of an ICR cell is shown in Figure 19.13 as a function of cell size. Note that retention decreases for smaller cells. A more detailed stability analysis of AFs can be found in Ref. [28].

Figure 19.14 compares the state lifetimes in AF and ICR models as a function of the critical number of atoms to create the ON state. Note that both models, that is, as given by Eq. (19.43b) for AF and Eq. (19.45) for ICR are in a reasonable agreement.

19.4 Energetics of Nanoionic Devices

19.4.1 Switching Speed and Energy

Consider the minimal atomic bridge cell in ON state, formed by a 4 atoms \times 4 atoms \times 4 atoms cube. (64 metal atoms, approximately 1 nm in size). Then, to

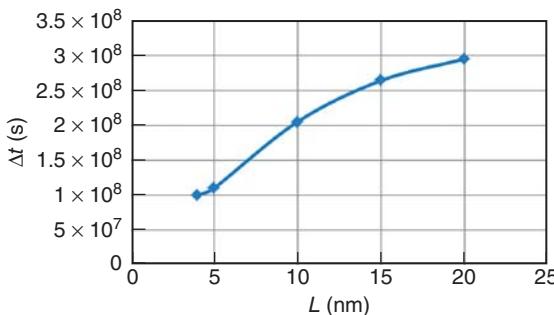


Figure 19.13 Retention time of an ICR cell as a function of cell size for $E_a = 1.25$ eV and $T = 330$ K.

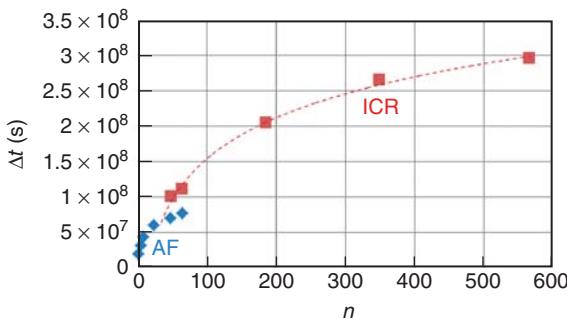


Figure 19.14 State lifetimes in AF and ICR models as a function of the critical number of atoms to create the ON state ($E_a = 1.25$ eV and $T = 330$ K).

achieve nonconductive OFF state, we partially “dissolve” the cube by applying external stimulus, such as electrical signal (specific mechanisms can be different for different materials; however, they all include moving atoms). As discussed in Section 19.3.2.5, a three-atom gap is sufficient to obtain a reasonably large resistance ON/OFF ratio (>10); therefore, it is assumed that “dissolving” of a 3 atoms \times 4 atoms \times 4 atoms fragment (48 atoms) represents ON–OFF switching. Assume that the external stimulus decreases the activation energy E_a in Eq. (19.37) by ΔE for each atom in the bridge cube and enables atomic drift and diffusion in the direction of the applied field. The time to dissolve the bridge can be estimated using Eq. (19.43b)

$$t_{SW} = \frac{1}{f_0} \frac{\ln \left(1 - 2^{-\frac{1}{n}} \right)}{\ln \left(1 - \exp \left(-\frac{E_a - \Delta E}{k_B T} \right) \right)} \quad (19.46)$$

For $E_a = 1.25$ eV and $\Delta E = 1$ eV, $t_{SW} \approx 40$ ns.

Assuming $\Delta E = 1$ eV = 1.6×10^{-19} J, the switching energy can be estimated as

$$E_{SW} \sim 48\Delta E \sim 10^{-17} \text{ J} \quad (19.47)$$

It is instructive to compare the result (Eq. (19.47)) with the energy for fusion (melting) of a $v = 1 \text{ nm}^3$ silver:

$$E_{\text{fusion}} = c_{\text{Ag}} \cdot m \cdot (T_m - T_{\text{op}}) + H_{\text{fus}} \cdot m = \gamma v \cdot (c_m(T_m - T_{\text{op}}) + H_{\text{fuse}}) \approx 3.3 \cdot 10^{-18} \text{ J} \quad (19.48)$$

where $c_{\text{Ag}} = 0.233 \text{ J g}^{-1} \text{ K}^{-1}$ is the specific heat capacity of silver, $T_m = 1235 \text{ K}$ is the melting point of silver, $T_{\text{op}} = 330 \text{ K}$ is the device operating temperature, $H_{\text{fus}} = 104.4 \text{ J g}^{-1}$ is the latent heat of fusion of silver, and $\gamma = 10.49 \text{ g cm}^{-3}$ is the density of silver.

Note that Eqs. (19.47) and (19.48) represent an estimate of a lower bound of switching energy of an nanoionic device, as it assumes a 100% conversion efficiency of the external stimulus into the activation barrier lowering ΔE , which will be difficult to achieve in practical devices. If, for example, the conversion efficiency is 10%, the resulting switching energy is $E_{\text{SW}} \sim 10^{-16} \text{ J}$.

19.4.2

Heat Dissipation and Transfer in a Minimal Nanoionic Device

As discussed in Section 19.3.2, the minimal resistance achievable in an atomic bridge is $\sim 13 \text{ k}\Omega$, and thus the theoretical limit on maximum current at operating voltage of 1 V is $\sim 80 \mu\text{A}$. This is almost two orders of magnitude larger than the read current requirement ($1 \mu\text{A}$) for a high-performance ICT device suggesting that the currents $\sim 1 \mu\text{A}$ could be achievable in atomic-scale devices. However, Joule heat will be released as the current passes through the structure, for example, during read operation, which can potentially destroy the few-atom conductive bridge structure. In this section, a simple analysis of thermal properties of nanoionic switch during READ operation is conducted, assuming the read current $I_r = 1 \mu\text{A}$, the read time $t_r = 1 \text{ ns}$, and the conductive filament formed by a 4 atoms \times 4 atoms \times 4 atoms cube, approximately 1 nm in size. The corresponding bridge resistance from Eqs. (19.4) to (19.6) (using parameters in Table 19.1 and assuming the specularity $\alpha = 0$ in Eq. (19.6)) is $\sim 10^3 \Omega$. The power and energy dissipation in the ON state are, respectively,

$$P = I^2 R \sim (10^{-6} \text{ A})^2 \cdot 10^3 \Omega = 10^{-9} \text{ W} \quad (19.49\text{a})$$

$$E = Pt = 10^{-9} \text{ W} \cdot 10^{-9} \text{ s} = 10^{-18} \text{ J} \quad (19.49\text{b})$$

The direct analogy between classical formulas of heat and electrical transport, Fourier's law of heat flow Q is reminiscent to Ohm's law for electrical current:

$$Q = \frac{\Delta T}{R_\Theta} \quad (19.50)$$

where the temperature difference ΔT is analogous to the potential difference V and R_Θ is the *thermal* resistance. For the case of AF bridging two metal plates, similar to Eq. (19.4),

$$R_{\Theta\text{bridge}} = R_{\Theta\text{fil}} + 2R_{\Theta C}, \quad (19.51)$$

where $R_{\Theta\text{fil}}$ is the thermal resistance of the “filament” and $R_{\Theta C}$ is the contact (spreading) resistance of a thermal contact between the filament and the metal plates. The spreading thermal resistance, analogous to Eq. (19.10), is given by

$$R_{\Theta C} = \frac{\theta_0}{d} = \frac{1}{\sigma_0 d}, \quad (19.52)$$

where θ_0 is the *bulk* thermal resistivity of metal reservoir, $\sigma_0 = 1/\theta_0$ is the bulk thermal conductivity and d is the filament diameter. The filament thermal resistance can be calculated as

$$R_{\Theta\text{fil}} = \theta^* \frac{L}{d^2}, \quad (19.53)$$

where L is the length of the filament, d is the filament diameter, and θ^* is the thermal resistivity of metal nanowire, which is expected to be larger than the bulk thermal resistivity θ_0 . To estimate θ^* , the Wiedeman–Franz law can be used, which states that at a given temperature, the thermal and electrical conductivities of metals are proportional:

$$\frac{\rho}{\theta} = \rho \cdot \sigma = \text{const} \cdot T. \quad (19.54a)$$

Therefore, it is assumed that

$$\rho_0 \cdot \sigma_0 = \rho^* \cdot \sigma^*, \quad (19.54b)$$

or

$$\sigma^* = \frac{\rho_0}{\rho^*} \sigma_0. \quad (19.54c)$$

Using the values of $\sigma_0 = 429 \text{ W m}^{-1} \text{ K}^{-1}$ (bulk thermal conductivity of silver), $\rho_0 = 15.8 \text{ n}\Omega \text{ m}$ (Table 19.1), and $\rho^* = 853 \text{ n}\Omega \text{ m}$ (calculated using Eq. (19.6) for $d = 1 \text{ nm}$) results in $\sigma^* = 7.9 \text{ W m}^{-1} \text{ K}^{-1}$.

To calculate the total thermal resistance, it is further assumed that the heat source is located in the middle of the 1 nm cube (i.e., in Eq. (19.53) $L = d/2$) and that no lateral heat conductance into insulator, that is, most of the heat is flowing in the top and bottom metal contacts. From Eqs. (19.51) to (19.53),

$$R_{\Theta\text{bridge}} = \frac{1}{\sigma^*} \frac{L}{2d^2} + \frac{1}{\sigma_0} \frac{1}{2d} \quad (19.55)$$

Assuming that the metal plates contacting the bridge are maintained at constant temperature (e.g., $T_c = 330 \text{ K}$) and moderate $\Delta T = 10 \text{ K}$, Eqs. (19.50) and (19.55) result in the heat removal rate from the conductive bridge $Q = 3 \times 10^{-7} \text{ W}$. This is more than two orders of magnitude greater than the heat generation rate given by Eq. (19.53). The expected temperature increase can be estimated by balancing P (Eq. (19.49a)) and Q (Eq. (19.50)) and is very small:

$$\Delta T = P \cdot R_{\Theta\text{bridge}} \sim 0.06 \text{ K}$$

These estimates suggest that in principle, the 1-nm atomic bridge structure could operate at realistic conditions. Of course, more detailed calculations that take into account exact geometry, materials system, and so on, can be done.

Such an analysis would provide additional insight into the scaling properties of nanoionic devices.

19.5

Summary

In this chapter, two mechanisms for bipolar resistive switching based on the rearrangement of atoms have been considered: conductive AF as typically formed by the ECM process and ICR as typically arises from valence change processes in oxides. For both mechanisms, the minimum critical size of the conductive element has been estimated, based on the advanced high-performance ICT device constraints, such as operating voltage ~ 1 V, read time ~ 1 ns (read current $> 1 \mu\text{A}$), state lifetime $\sim 3 \times 10^8$ s (for nonvolatile memory applications). The AF cell can, in principle, be scaled to 1 nm structure, while for the ICR cell, the critical dimensions $L_{\min} \sim 4$ nm, as determined by the minimum read current requirement and material limits of doping concentration. The ICR cell has a clear scaling limit due to doping limitations, and the ultimately scaled nanoionic devices are likely to be of the conductive filament type.

Estimates of energetics and speed of nanoionic device cells at the limits of scaling suggest that the write speed of the order of 10 ns can be achieved at a write energy of 10^{-16} to 10^{-17} J/bit. The read energy is determined by the read current and can be as low as 10^{-17} J/bit for 1 nm AF device.

Acknowledgment

The authors would like to thank Roy Meade (Micron Technology, Inc.) for stimulating discussions of heat dissipation and transfer in a minimal nanoionic device.

APPENDIX A Physical Origin of the Barrier Potential

Consider a vacuum–metal interface. Clearly, a surface barrier must prevent the easy escape of electrons from the metal “reservoir,” and thus make it possible for the existence of a stable solid state. A simple (but rather accurate) model for the surface barrier was first proposed by Schottky, based on classical electrostatics [22, 23]. He suggested that electrons leaving a conducting surface and at a given moment located at distance x above the surface must create a positively charged surface layer which attracts the electron (Figure A.1a). As a result an attractive force appears between the electron and the surface, and thus a barrier exists preventing the escape. As is shown in the theory of electrostatics, this is equivalent to the force due to a fictitious positive charge located behind the surface at the equal distance ($-x$) as the original charge, that is, the mirror image of the electron. The attractive force between the electron above the surface and its image below the surface is expressed by the Coulomb law:

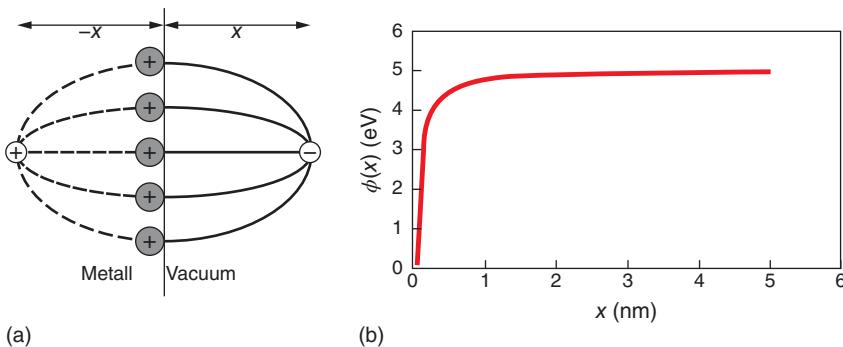


Figure A.1 Image charge model of barrier formation at metal–vacuum interface.

The attractive force acting on the escaping electron is equivalent to a presence of a barrier preventing electron escape. The barrier height, E_b , is equal to the total work to move the electron from a point x near the surface to infinity:

$$F = \frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{e^2}{(2x)^2} = \frac{e^2}{16\pi\epsilon_0 x} \quad (\text{A1})$$

$$E_b = \int_x^\infty F dx = \int_x^\infty \frac{e^2}{16\pi\epsilon_0 x^2} dx = -\frac{e^2}{16\pi\epsilon_0 x} + \text{const} = \phi_0 - \frac{e^2}{16\pi\epsilon_0 x} \quad (\text{A2})$$

The integration constant ϕ_0 in Eq. (A2) is called the work function and is the characteristic property of a given material. For example, the work function of silver is $\phi_{\text{Ag}} = 4.7$ eV.

This simple model predicts the behavior of the surface barrier with surprising accuracy. In fact, models based on accurate quantum mechanical calculations give only small corrections to the Schottky potential at distances >1 nm from the surface. At very short distances from the surface, the simple Schottky model is insufficient, because factors other than pure electrostatics contribute to the potential.

The profile of the surface barrier $E_b(x)$ is shown in Figure A.1. At large x , the barrier can be considered rectangular, and the rectangular barrier shape is often assumed in simplified analyses.

Until now, we have considered the surface barrier profile, when no external electric fields have been applied to the solid. The effect of an external field F is to change the slope of the barrier profile curve corresponding to the potential energy change of eFx :

$$E_b(x) = \phi_0 - \frac{e^2}{16\pi\epsilon_0 x} - eFx \quad (\text{A3})$$

Figure A.2 shows surface barrier profiles for different external electric fields. It can be observed that the barrier height is lower in higher fields, an effect known as Schottky lowering of the surface barrier. This effect can be quantified by taking

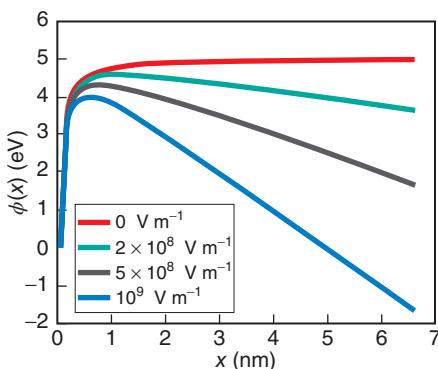


Figure A.2 Schottky lowering of barrier due to external fields.

the derivative of Eq. (A3) and setting $(dE_b/dx) = 0$,

$$\frac{dE_b}{dx} = \frac{e^2}{16\pi\epsilon_0 x^2} - eF = 0, \quad (\text{A4})$$

from which the barrier height reduction is

$$\Delta E_b = \phi_0 - E_{b\max} = \sqrt{\frac{eF}{4\pi\epsilon_0}} \quad (\text{A5})$$

and the position of the barrier maximum relative to the surface is

$$x_{\max} = \sqrt{\frac{e}{16\pi\epsilon_0 F}} \quad (\text{A6})$$

References

1. Welser, J. (2009) The semiconductor industry's nanoelectronics research initiative: motivation and challenges. Design Automation Conference, pp. 298–300.
2. Bourianoff, G.I., Gargini, P.A., and Nikonorov, D.E. (2007) Research directions in beyond CMOS computing. *Solid-State Electron.*, **51**, 1426–1431.
3. Ayres, R.U. (1994) *Information, Entropy and Progress*, AIP Press, New York.
4. Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833–840.
5. Beck, A., Bednorz, J.G., Gerber, C., Rossel, C., and Widmer, D. (2000) Reproducible switching effect in thin oxide films for memory applications. *Appl. Phys. Lett.*, **77**, 139–141.
6. Terabe, K., Hasegawa, T., Nakayama, T., and Aono, M. (2005) Quantized conductance atomic switch. *Nature*, **433**, 47–50.
7. Hasegawa, T., Terabe, K., Sakamoto, T., and Aono, M. (2009) Nanoionics switching devices: atomic switches. *MRS Bull.*, **34**, 929–934.
8. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) Redox-based resistive switching memories – Nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21**, 2632.
9. Akinaga, H. and Shima, H. (2010) Resistive Random Access Memory (ReRAM) based on metal oxides. *Proc. IEEE*, **98**, 2237–2251.

10. Blatt, F.J. (1968) *Physics of Electron Conduction in Solids*, McGraw-Hill, New York.
11. Dingle, R.B. (1950) The electrical conductivity of thin wires. *Proc. R. Soc. London*, **47**, 545.
12. Sondheimer, E.H. (1952) The mean free path of electrons in metals. *Adv. Phys.*, **1**, 1–42.
13. Jansen, A.G.M. *et al.* (1980) Point-contact spectroscopy of metals. *J. Phys. C: Solid State Phys.*, **13**, 6073.
14. Norberg, G. *et al.* (2006) Contact resistance of thin metal film contacts. *IEEE Trans. Compon. Packag. Technol.*, **29**, 371.
15. Scheer, E., Agrait, N., Cuevas, J.C., Yeyati, A.L., Ludoph, B., Martin-Rodero, A., Bollinger, G.R., van Ruitenbeek, J.M., and Urbina, C. (1998) The signature of chemical valence in the electrical conduction through a single-atom contact. *Nature*, **394**, 154–157.
16. Imry, Y. and Landauer, R. (1999) Conductance viewed as transmission. *Rev. Mod. Phys.*, **71**, S306–S312.
17. Simmons, J.G. (1963) Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.*, **34**, 1793–1803.
18. Sze, S.M. (1981) *Physics of Semiconductor Devices*, John Wiley & Sons, Inc..
19. Yang, J.J., Pickett, M.D., Li, X.M., Ohlberg, D.A.A., Stewart, D.R., and Williams, R.S. (2008) Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat. Nanotechnol.*, **3**, 429–433.
20. Jeong, D.S., Schroeder, H., and Waser, R. (2009) Mechanism for bipolar switching in a Pt/TiO₂/Pt resistive switching cell. *Phys. Rev. B*, **79**, 195317.
21. Gonon, P. *et al.* (2010) Resistance switching in HfO₂ metal-insulator-metal devices. *J. Appl. Phys.*, **107**, 074507.
22. Sun, B., Liu, L.F., Xu, N. *et al.* (2009) The effect of current compliance on the resistive switching behaviors in TiN/ZrO₂/Pt memory device. *J. Appl. Phys.*, **48**, 04C061.
23. Kukreja, L.M., Das, A.K., and Misra, P. (2009) Studies on nonvolatile resistance memory switching in ZnO thin films. *Bull. Mater. Sci.*, **32**, 247–252.
24. Nishi, Y., Menzel, S., Fleck, K., Bottger, U., and Waser, R. (2014) Origin of the SET kinetics of the resistive switching in tantalum oxide thin films. *IEEE Electron Device Lett.*, **35**, 259–261.
25. Schultze, J.W. and Lohrengel, M.M. (2000) Stability, reactivity and breakdown of passive films. Problems of recent and future research. *Electrochim. Acta*, **45**, 2499–2513.
26. Stimming, U. and Schultze, J.W. (1979) A semiconductor model of the passive layer on iron electrodes and its application to electrochemical reactions. *Electrochim. Acta*, **45**, 859–869.
27. Schroeder, H., Zhirnov, V.V., Cavin, R.K., and Waser, R. (2010) Voltage-time dilemma of pure electronic mechanisms in resistive switching memory cells. *J. Appl. Phys.*, **107**, 054517.
28. Zhirnov, V.V., Meade, R., Cavin, R.K., and Sandhu, G. (2011) Scaling limits of resistive memories. *Nanotechnology*, **22**, 254027.
29. Runyan, W.R. and Bean, K.E. (1990) *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley Publishing Company, p. 390.

20

Integration Technology and Cell Design

Fred Chen, Jun Y. Seok, and Cheol S. Hwang

20.1

Materials

The categories of materials that can constitute a RRAM cell are presented as components of the basic RRAM cell shown in Figure 20.1.

The structure shown in Figure 20.1 represents the planar stack structure, which is the simplest way to fabricate a RRAM structure, since all that is required is a single etch mask pattern that defines the etch all the way through the stack. The top and bottom electrodes are, by definition, conducting media that interface with nonconducting or semiconducting media, which comprise the layers in between. The actual resistance-switching layer may be in contact with a reductant layer, which can enhance ion exchange with the resistance-switching layer as a key part of the switching process. The choices of the particular materials that may be used in each of these cases are discussed in the following sections.

20.1.1

Resistance Switching (RS) Materials

Resistance switching (RS) materials for RRAM generally have been based on four categories: (i) insulating oxides (as often used for capacitors), such as Al_2O_3 , Ta_2O_5 , HfO_2 , and ZrO_2 ; (ii) semiconducting oxides, such as TiO_x , NiO , and WO_x ; (iii) ion-conducting electrolytes, such as the chalcogenides GeSe and GeS ; and (iv) phase-change materials, such as Ge-Sb-Te alloys. There are relative advantages and disadvantages of using each category.

20.1.1.1 Insulating Oxides

Insulating oxides, typically binary oxides such as HfO_2 and Al_2O_3 , are extremely good electrical insulators, with extremely low leakage current densities. Also, with breakdown fields on the order of $5-10 \text{ MV cm}^{-1}$, several nanometers of oxide would require several volts to initiate breakdown through the filament-forming process. The combination of low voltage, low thickness, and low current readily makes these commonly encountered oxides an extremely attractive choice for

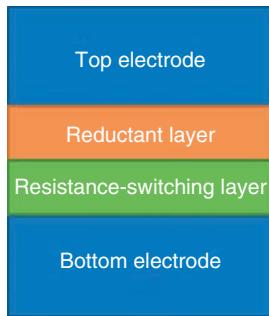


Figure 20.1 RRAM cell constituent material categories, corresponding to their role in a conventional planar stack structure. The reductant acts to reduce the resistance-switching layer by receiving oxygen or diffusing metal into the layer.

RRAM. However, there is a fundamental drawback, namely the tradeoff between leakage current, for too low a thickness, versus higher forming voltage, for too high a thickness. In addition, some oxides, particularly Al_2O_3 , display nonlinearly increasing leakage current near the breakdown voltage, due to Fowler–Nordheim tunneling.

20.1.1.2 Semiconducting Oxides

Semiconducting oxides, such as TiO_x , NiO , or WO_x , pass larger currents than insulating oxides. Hence, as resistive switching materials, they are at an inherent disadvantage compared with the insulating oxides. Furthermore, they are often included as layers adjacent to the resistance-switching insulating oxide layer, in certain RRAM cells, particularly those based on Ta_2O_5 , since the semiconducting form of tantalum oxide can act as a current limiter during forming or the SET operation [1]. This can possibly allow the self-compliant operation, which, coupled with a nonlinear resistance, is highly desirable for high-density transistorless RRAM cells.

20.1.1.3 Electrolyte Chalcogenides

Chalcogenides, which can act as ion-conducting electrolytes, such as germanium selenide or germanium sulfide, can be combined with electrodes made of a diffusing metal, such as copper or silver, to form programmable metallization cells (PMCs), also known as conducting bridge random access memory (CBRAM) cells [2]. The electrolytes have naturally very high resistance until penetrated by the conducting bridge formed from diffusing metal collecting on the inert counter electrode. Potentially, the electrolyte has the same traits as the insulating oxides, but there are additional material concerns. First, selenium and sulfur are highly toxic, so bringing them into a conventional fabrication environment may be difficult. Second, it is necessary to have the electrolyte medium between the electrodes already containing ions that can respond to an electric field. It means the electrolyte requires an additional step to drive in the ions from the diffusing electrode. This could involve photodoping, annealing, *in situ* doping during deposition, and so on. Third, the processing of copper or silver in high-density arrays

can be quite difficult, as the inclusion of diffusion barriers is now made more complicated.

20.1.1.4 Phase-Change Materials

Phase-change materials gained popularity first with the use of optical drives and then as the basis of ovonic unified memory (OUM). With the standard material being $\text{Ge}_2\text{Sb}_2\text{Te}_5$, the fabrication of phase-change memory cells was developed to an advanced stage by Samsung and Numonyx [3]. However, the strong thermal dependence of the phase-change mechanism itself, as well as the required sublithographic processing needed to guarantee sufficiently high current density for the large Joule heating energy required to drive the phase change, led to its failure to keep up with the power reduction offered by other emerging memory candidates, including MRAM, CBRAM, and RRAM.

20.1.2

Electrode Materials, Including Reductants

The electrode material has also played a key role in the functioning of RRAM cells, though its role is more subtle than that of the resistance-switching material. Some key characteristics of the electrode material are its work function, oxidation potential, and tendency to diffuse.

The work function determines whether the electrode–RS material interface will form a Schottky barrier for the given RS material. This, in turn, affects the operation voltage window, as the Schottky barrier needs to be lowered to effect a switching transition in the RS material. For example, the $\text{Pt}–\text{TaO}_x$ Schottky barrier is a key aspect of the $\text{TaO}_x/\text{TaO}_y$ bilayer RRAM devices demonstrated by Panasonic and Samsung [4].

The oxidation potential determines whether the electrode forms a compound that interfaces with the RS material. The formation of this compound could influence later switching of the RS material. For example, oxidation of a Ti layer in contact with HfO_2 could influence the oxygen vacancy concentration within the HfO_2 layer. In fact, due to the electrochemically active role of the Ti layer, it may not be rigorously correct to qualify it as an electrode, but more specifically as a reductant, that is, a layer that is to be oxidized.

A postdeposition annealing step could thermally drive a reaction between the RS material and the reductant. This is the basis for, for example, forming oxygen vacancies in HfO_2 when reacting with a strong oxygen getter material, such as Ti, at temperatures up to 450°C [5]. The higher the temperature, the more complete the vacancy generation within a specified period of time, for example, 5 min. Conversely, a lower temperature would require a longer heating duration for the same effect.

The diffusivity of the electrode material is a deciding factor of the functionality of CBRAMs, as in the case of Cu or Ag. However, high metal diffusivity poses serious fabrication considerations in that the interface between the electrolyte and the diffusing metal cannot be a barrier, but all other material interfaces

with the diffusing metal must have a barrier. This constraint means the diffusing metal should preferably be a bottom electrode material deposited into a trench with diffusion barrier liners. In turn, this often means that P-MOSFET rather than N-MOSFET is required as the selector to apply a positive voltage to the bottom electrode for the electroforming and SET operations. As with the previous Ti case, the diffusing metal also serves as the reducing agent.

20.2 Structures

RRAM structures can essentially be fabricated in three ways: as planar stack structures, lateral channel structures, or sidewall-conforming stack structures.

20.2.1

Planar Stack

The planar stack structure, as shown in Figure 20.1, represents the easiest type of RRAM structure to fabricate. A single photolithography step is used to pattern an etch mask, which is used to define the width of the cell stack. In fact, this fabrication process flow is similar to that being used for MRAM magnetic tunnel junction (MTJ) fabrication today, but with fewer and simpler layers to be etched [6]. It is expected that as feature sizes reach sub-30 nm levels, patterning would result in greater lithographic difficulty.

The stack etch can stop on either a wider bottom electrode or on a dielectric layer, which includes a smaller bottom electrode in a pore, similar to the mushroom structure commonly used in early PCM demonstrations. The latter is more compatible with industry practice because it is easier to find a metal etch that stops on dielectric than on metal.

For high-density applications, the planar stack is often used to pattern cross-bar structures, where the top and bottom electrodes are essentially intersecting orthogonal lines. In this case, the cell or array explicitly needs at least two masks for patterning, one to define lines of x -direction and another to define lines of y -direction.

Some complications of etching the planar RRAM including the top electrode, reductant, and switching layer can result in damage to the cell, resulting in loss of reliable performance. The first possible source of damage is radiation or ions from the etch plasma directly impacting the switching layer. This can result in a tendency for filaments to form along the perimeter of the stack [7]. Another source of damage is oxidation of the reductant during the fabrication process [8]. This would cause the reductant to be less effective in electrochemically reducing the RS layer, that is, generating oxygen vacancies or cations diffusing in. A third source of

damage comes from the antenna effect from plasma current being collected from the top electrode into the much smaller bottom electrode. This can be addressed by deliberately designing alternative current conduction paths that bypass the RS material.

Besides etching, the planar stack may be more subtly damaged by being exposed to air after stack etching, leading again to oxidation of the reductant. This can be avoided in industrial fab settings by the use of SMIF pods. Also the careful choice of encapsulation, for example, silicon nitride instead of silicon dioxide, can eliminate the possibility of an extra oxygen source that continues to oxidize the reductant even after the fabrication is completed [8].

The sidewall oxidation of the reductant is suspected to result in a variety of reliability issues, including forming failure [8], and retention failure [9]. Since this issue becomes more severe with smaller size, due to smaller distance between sidewalls, it could pose a scalability risk for the planar stack structure.

In terms of device operation parameters, the most significant impact of shrinking the width of the planar stack is to increase the forming voltage or forming time. The oxide thickness can be decreased to address this trend, but this has to be traded off against higher leakage current density.

20.2.2

Sidewall-Conforming Stack

The alternative to first depositing the stack of cell layers and then patterning directly is first patterning a topography, for example, a trench or via, and then depositing the cell layers over the topography. As we discuss shortly, this patterning methodology offers the opportunity to avoid many process-induced reliability risks.

An early example of a deposition-over-topography approach is the so-called concave structure (so named due to its shape similarity to a concave lens) [10] (Figure 20.2). In this approach, a dielectric layer is first deposited over a bottom electrode. A via opening is etched in the dielectric layer, landing on the bottom electrode. Then, the resistive switching material, reductant, and top electrode layers are deposited into and over the via opening. The deposited layers are etched to stop on the dielectric top surface. Therefore, the path from the top sidewall to the bottom active area is blocked or at least frustrated to a degree, mitigating a potential retention degradation mechanism. The interface between the bottom electrode and the resistive material is shielded from any plasma and radiation damage, mitigating those degradation mechanisms as well (Figure 20.2).

The sidewall-conforming structure has some advantages in avoiding process damage, but also fundamentally has some scaling limits. The sum of the sidewall layer thicknesses (excluding the center layer), when doubled, gives a minimum possible dimension of the structure. For example, a 5 nm HfO₂ outer sidewall layer

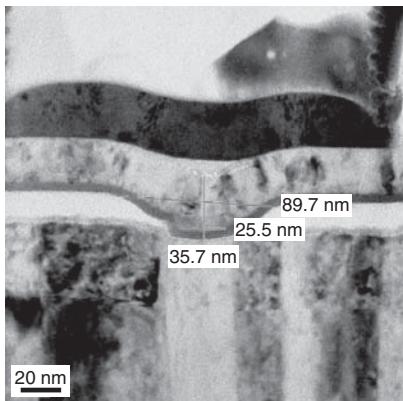


Figure 20.2 Cross section showing the concave structure with sidewall-conforming stack and sub-30 nm active area. (From Ref. [11].)

enclosing a 10 nm Ti inner reducing layer and a 10 nm TiN center layer (top electrode), gives 40 nm minimum trench or via width. In contrast, in a planar stack, the thicknesses do not influence the minimum width at all, and the limit would be the resolution of the patterning method instead. However, 3D array patterning can overcome the resolution limits of sidewall-conforming patterning. Also, sidewall-conforming structures effectively decouple the electroforming parameters from area scaling.

20.2.3

Lateral Structure

For completeness, we must also mention the lateral structure, where the two electrodes lie in the same horizontal plane, and the current is also conducted horizontally in between them (Figure 20.3). In some cases, it presents an excellent opportunity to view filament formation and disruption in RRAMs, particularly CBRAMs [12]. The gap between electrodes needs to be tightly controlled in the lithography as it is directly related to the electroforming voltage. Furthermore, if the electrodes are of different materials, an extra lithography step would be required. Since the electrodes necessarily occupy different lateral locations, the

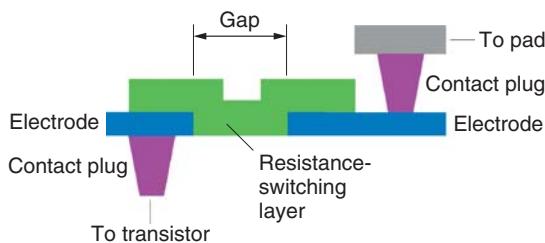


Figure 20.3 Lateral structure cross-sectional view.

cell size also cannot be minimized very conveniently. For these reasons, the lateral structure remains an interesting tool for viewing the basic operation of a RRAM but not for mass production.

20.3 Integration Architectures

The integration architecture considers the individual device as part of an entire memory array, with specific access strategies for reading and writing data. Besides the RRAM device, an isolation device may also be placed in the cell, to electrically isolate the cell from sneak currents from other cells when the cell is not selected for operation. The isolation device can be a two-terminal device such as a diode, or a three-terminal device, that is, a transistor. Complicating the selection is whether the RRAM device uses unipolar operation (set and reset in same direction of voltage) or bipolar operation (set and reset are opposite directions of voltage). Bipolar operation is particularly challenging in that it would require a special diode to operate in both voltage polarities, or operating an N-MOSFET with ungrounded source.

Besides basic operation parameters, the different styles of integration architecture have significantly different implications for performance, scalability, density, and power consumption. We discuss each style in detail.

20.3.1

Transistor in Series with RRAM (1T1R)

The most widely demonstrated RRAM array architecture to date is the 1T1R style (Figure 20.4). Each cell contains a transistor in series with the RRAM. The transistor performs several duties very reliably: (i) electrical isolation, preventing

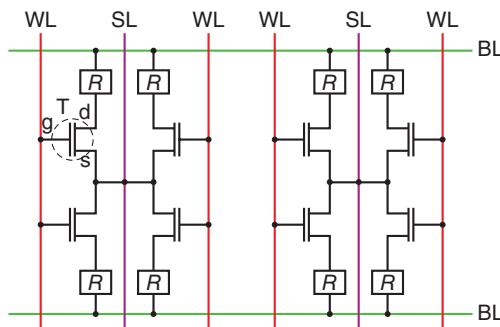


Figure 20.4 Schematic of 1T1R architecture. WL = word line connecting to transistor gates. BL = bit line connecting to resistive switching elements. SL = source line connecting to transistors' common source.

sneak currents through unselected cells; (ii) current compliance, preventing the RRAM from permanent breakdown during the forming and set operations; and (iii) voltage regulation, allowing sufficient voltage on the RRAM during reset and read operations [13].

Since the power consumption during the set operation can be reduced by reducing the current compliance through lowering the gate voltage, and the power consumption during RESET at a given current level can be reduced by increasing the gate voltage overdrive ($V_{gs} - V_{th}$) sufficiently to lower the source–drain voltage; the path to power reduction is straightforward in the 1T1R scheme. The caveat here is to avoid having the gate voltage overdrive being too excessive, to the point of causing gate oxide breakdown.

1T1R operation is also tricky in a number of aspects. First, during forming or set operations, the voltage on the RRAM (often up to 2–3 V) is suddenly shifted (at least in part) to the transistor as it exerts compliance. Obviously, this poses a strong reliability risk to the transistor. Drain-induced barrier lowering (DIBL) may also weaken the current compliance imposed by the transistor. To avoid all the risks associated with excessive voltage on the cell transistor, the circuit design could include an additional transistor along the same circuit path, for example, an access transistor on the bit line, which can share the large voltage, allowing the cell transistor to meet its voltage tolerance specification.

An even trickier operation to handle is the reset operation. In this case, the cell transistor is an N-MOSFET, with the RRAM effectively being between the source and ground. This arrangement leads to strong feedback between the gate voltage and the source voltage, which would require a load line model (Figure 20.5) to correctly determine the sequence of voltage inputs to the gate and drain. A larger voltage on the RRAM increases the source voltage, which in turn, drives

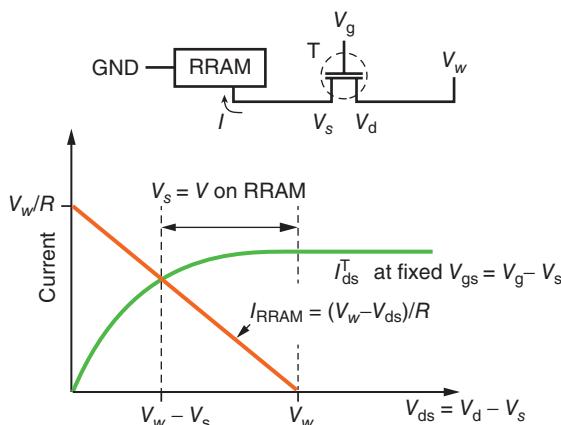


Figure 20.5 Load line model for the 1T1R reset operation, with the RRAM treated as a load with resistance R .

the gate voltage upward. Caution must be taken not to put the entire target voltage on the transistor all at once, in order not to exceed its specification. When the voltage input sequence is done properly, as the RRAM resistance increases, the source–drain voltage becomes negligible, while the gate-source voltage is near the maximum operation limit.

The scalability of the 1T1R architecture is defined by the transistor scaling and RRAM operation capability. The RRAM element size itself can be made very small, much smaller than the transistor could possibly be [14]. Physical size scaling of the transistor results in both current and voltage scaling. Current scaling requires the lowest resistance state of the RRAM to be able to reach very high values ($>1\text{ M}\cdot\Omega$). In turn, this entails an even higher reset resistance state (at least $1\text{ G}\cdot\Omega$). This could fundamentally limit the operation speed of the memory. Voltage scaling (lower I/O voltage, lower core voltage) requires a lower operation voltage (e.g., $<1\text{ V}$). This is possible, but could also entail much slower speeds due to the well-known voltage–time dilemma.

In fact, as far as the currently available transistor technology is concerned, the limiting feature for scaling is the transistor contact to diffusion. This is illustrated in more detail in Figure 20.6. Specifically, the contact requires some distance to the adjacent edges in the diffusion space, which itself is separated by some minimum spacing from other diffusion regions. The minimum half-pitch, therefore, can easily exceed twice the minimum feature size, for example, gate or contact critical dimension. Therefore, for logic-compatible embedded memory applications, the minimum 1T1R cell size will typically be expected to exceed $20F^2$, where F is the node designation (or minimum feature size).

Planar CMOS technology is expected to bottom out at around $F \sim 20\text{ nm}$, but with the current high cost of double patterning at this node, the practical limit is around $F \sim 28\text{ nm}$. Plainly, this cell size far exceeds typical Flash cell dimensions, although the flash memory fabrication process is itself not logic-compatible. For DRAM-level densities, the DRAM terminology is more likely to be used;

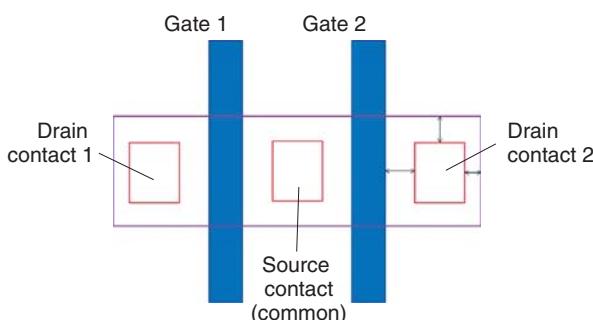


Figure 20.6 Layout top view of a typical 1T1R processed in last metal layer. The pitch can be several times the node designation.

the smallest 1T1R cell size is expected to be $6F^2$, with F being the half-pitch, as recently demonstrated by Micron/Sony at ISSCC 2014 [15].

20.3.2

Transistor in Parallel with RRAM (T||R)

NAND Flash as the mainstream nonvolatile storage memory has a cell footprint of $4F^2$. Rather than being a random-access memory, NAND Flash is organized into strings of bits connected together in series in a chain-like fashion [16]. It is possible to organize RRAM cells as shown in Figure 20.7.

Figure 20.8 shows a string of eight cells linked in series between ground and bit line. Each cell consists of a RRAM element in parallel with a transistor. In order to select a cell, say cell 5, the gate voltage to the corresponding cell is assigned to 0 V, while the other gate voltages are applied at successively higher values, with $V_{g4} > V_{g3} > V_{g2} > V_{g1} > V_t$ and $V_{g8} > V_{g7} > V_{g6} > V_t + V_{\text{RRAM}}$. It is also assumed that for each RRAM element, the LRS resistance far exceeds the source–drain resistance of each corresponding transistor when turned on. In fact, for successful RESET, the LRS resistance must exceed the sum of the source–drain ON resistances in series with the selected RRAM element. Otherwise, it would be rather difficult to guarantee sufficient RESET voltage on the element.

As with the lateral structure, a key process difficulty with this NAND-like arrangement is controlling the lateral dimensions of the contacting electrodes to avoid excessive electroforming voltage nonuniformity.

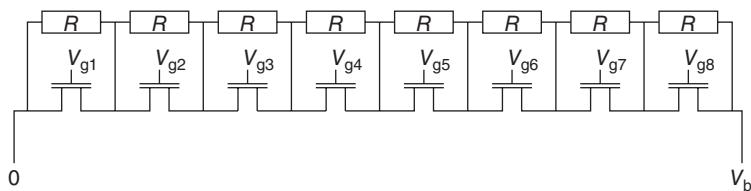


Figure 20.7 NAND-like RRAM bit string constructed from cells, each consisting of a resistive memory element in parallel with a transistor.

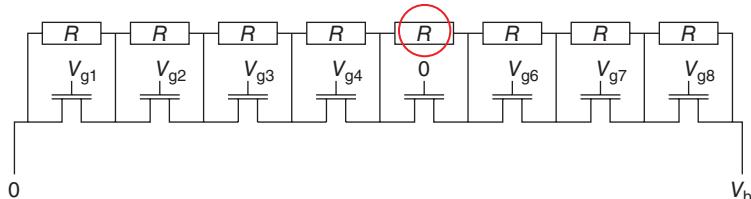


Figure 20.8 Cell selection in the chain RRAM arrangement of Figure 20.7.

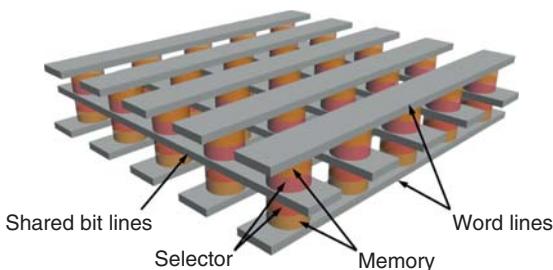


Figure 20.9 The stacked 1S1R crosspoint architecture. Each 1S1R cell is located at the intersection between crossed lines. (From Ref. [17].)

20.3.3

1S1R Stacked Crosspoint

Another way to achieve the $4F^2$ cell footprint is to use a crosspoint cell structure (Figure 20.9). In this case, the array consists of a set of parallel lines oriented in one direction, constituting one set of bottom electrodes, and another set of parallel lines oriented in an orthogonal direction, constituting a set of top electrodes, with the individual cells located at the intersection points. This array is inherently stackable, enabling the use of stacking in the third vertical dimension toward higher density integration.

20.3.3.1 The Selector Device

As each cell is located at the intersection between crossed lines, there is no room to place a conventional transistor. A vertical channel transistor would be an option, requiring a third line for the gate. However, it is inconvenient to mix the front- and back-end processes together in the fabrication. Furthermore, patterning the gate and gate oxide is particularly difficult. A more popular option is to fabricate a two-terminal selector in series with the RRAM element, to constitute a 1S1R (one-selector one-resistor) cell. The most conventional case is a p–n junction or Schottky diode. However, these devices pass current only in one direction, and block current in the other direction. Hence, they are only suitable for unipolar RRAM elements. Bipolar RRAM cells require selectors that can pass currents in both directions with comparable magnitude. However, to fulfill the isolation requirement, the current must decrease very significantly as voltage is decreased, for example, less than 1/1000 the current for half the voltage. This nonlinearity requirement is quite difficult to fulfill in practice. One device that has been particularly successful is the crested barrier, consisting of a strong TiO_2 barrier flanked by two weaker TaO_x barriers [18]. Another effective bipolar selector is the mixed ionic–electronic conductor (MIEC), which has had a very successful demonstration with PCM [19].

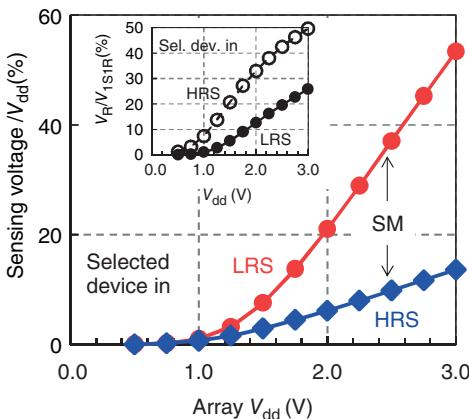


Figure 20.10 A higher read voltage increases the ratio of the RRAM relative to the selector. (From Ref. [20].)

20.3.3.2 Sensing Margin

The sensing margin of the crosspoint array is tied to the nonlinear behavior of the selector. It has been shown that larger operating voltages tend to improve performance [20].

A larger read voltage decreases the resistance of the selector relative to the RRAM, allowing for better sensing (Figure 20.10). However, this also increases the risk of state disturb by the read operation. A larger V_{read} also obviously increases power consumption.

The nonlinearity requirement for adequate cell isolation can be achieved through guaranteeing sufficiently low current under partial selection bias. Floating unselected lines (Figure 20.11a) leaves open the possibility of disturb when there are two HRS cells connected to ground or the applied voltage. The two most common biasing schemes for bipolar RRAMs are the half- and third-select schemes (Figure 20.11b,c). Unipolar RRAMs requiring only positive voltages may use p-n diodes as selectors, which can be reverse-biased to suppress leakage (Figure 20.11d). While the third-bias is safer against accidental disturb, the cumulative leakage over the whole array will be much larger than the half-bias scheme. In the half-bias scheme, most cells in the array are at zero bias, resulting in zero current. However, the cells in the same row and column are “half-selected”; the bias is $V_{dd}/2$. If the number of rows and columns in the array is N_{row} and N_{col} , respectively, then $I(V_{dd})/I(V_{dd}/2)$ should be much greater than $N_{row} + N_{col}$ to guarantee consistent current flow through the array under the same bias condition. For example, $N_{row} = N_{col} = 1000$ in a 1 Mb array. The selector should, therefore, satisfy $I(V_{dd})/I(V_{dd}/2) > 2000$.

The half- or third-select schemes just described assume that such reduced voltages would not be capable of disturb. Practically speaking, this could be dependent on the number of cycles over which such partial select voltages are applied.

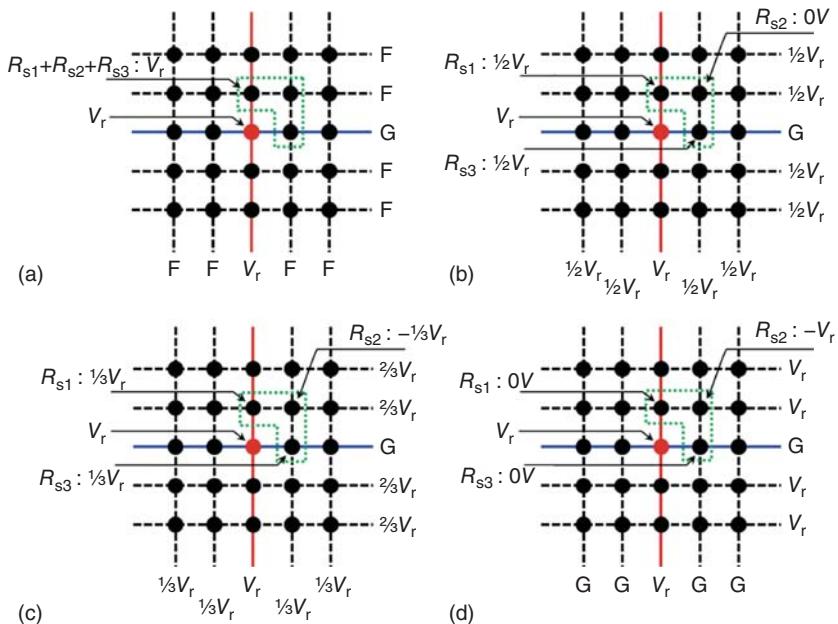


Figure 20.11 Four select schemes for voltage application to word and bit lines. (a) Floating unselected lines, (b) 1/2-select scheme, (c) 1/3-select scheme, and (d) reverse-biased unselected cells. G = Ground ($V = 0$). (From Ref. [17].)

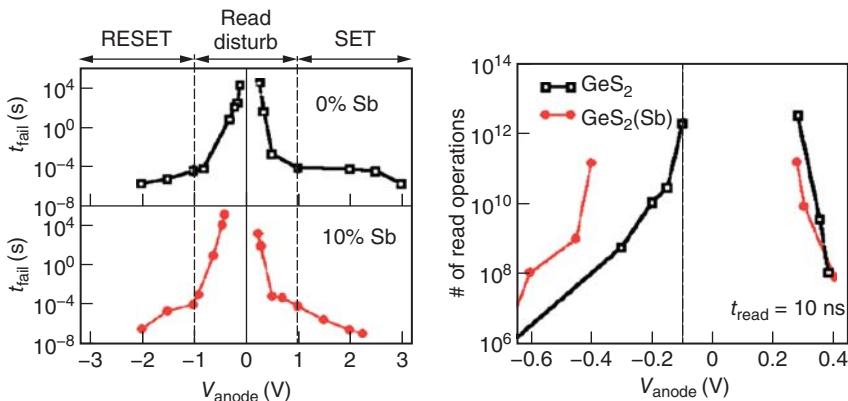


Figure 20.12 Number of read operations that can be performed before disturbing the cell, as a function of the anode voltage. (Based on data from Ref. [21].)

Figure 20.12 shows an example [21], where read voltages that are even less than a third of the switching voltage for an unselected cell could be high enough to disturb cells when applied over a sufficient number of cycles. The higher the number of write cycles, the more likely the partially selected cells on the same line will be affected. Therefore, the higher the endurance requirement, the tighter the

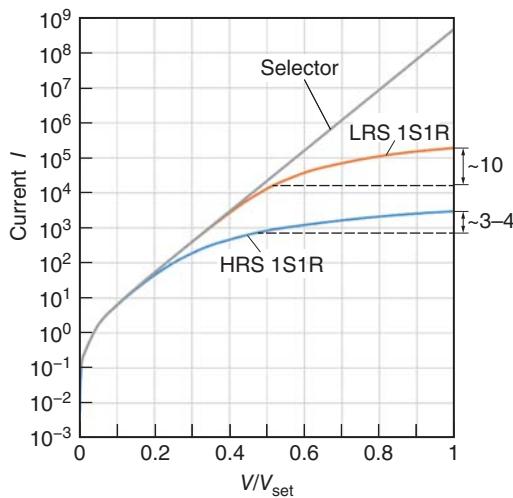


Figure 20.13 A selector with high nonlinearity does not guarantee current suppression at half the set voltage. Here, the HRS and LRS are both assumed to be linear with

a ratio of 100. At the half-set voltage, the benefit of the selector's nonlinearity of the selector has practically disappeared.

read disturb restriction. Or effectively, much less current could be allowed to pass through the cell switching element with sufficient nonlinearity.

20.3.3.3 Write Margin

Besides sensing, the writing operation must also minimize sneak currents as well as avoid cell disturbance. This is shown in Figure 20.13, where the selector device has a very large read sensing margin of more than 1000, but for the set operation, half-selected 1S1R LRS cells on the same line would still receive a substantial sneak current. The most serious concern is when the line is accessed by a current-limiting transistor with constrained dimensions. In that case, not enough current may go to the selected cell to reach the adequate set current level.

20.3.3.4 Cumulative Line Resistance

Another concern in crosspoint array design and fabrication is the cumulative resistance along the line joining a row of cells. Although on a per-cell basis, the line section resistance is negligible, over thousands of cells, the error of underestimating this resistance could lead to insufficient operating voltages delivered to a targeted cell. This error can be estimated systematically by considering Figure 20.14.

By applying practical values of the cell and line section resistances to calculate the overall resistance of the accumulated sections in parallel, one can estimate more accurately the load presented to an input signal. Figure 20.15 shows the error of neglecting the line section resistance in the resistance calculation, assuming a half-selected 1S1R cell resistance of $10\text{ M}\Omega$ and a line section resistance of 0.5Ω ,

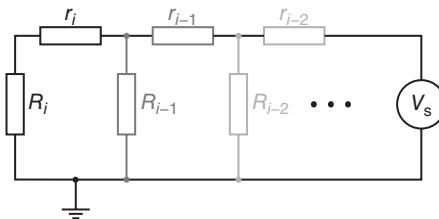


Figure 20.14 Schematic to calculate resistance of section of a row of cells, consisting of the cell resistance (R_i) and the lumped resistance of the section of the line between the indexed cell and its neighboring cell (r_i). Note that R_i and r_i in series are together in parallel with R_{i-1} .

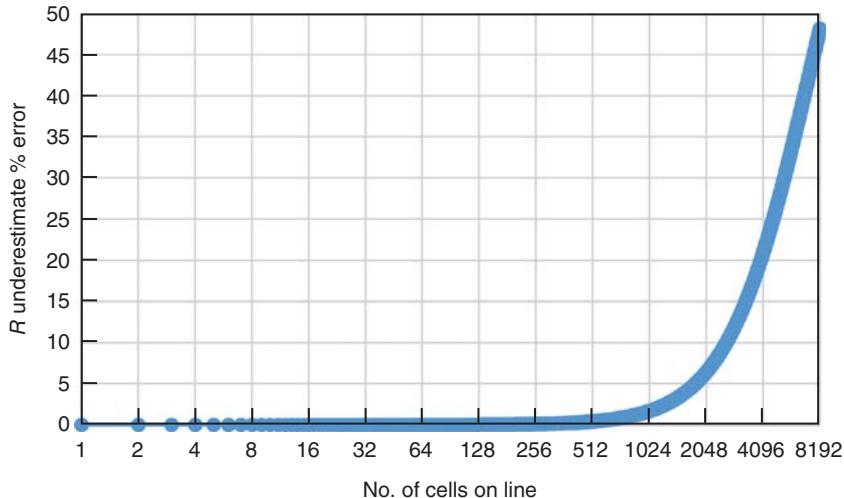


Figure 20.15 Calculated resistance error from neglecting line section resistance between cells (r_i) in Figure 20.11, for $r_i = 0.5 \Omega$ and $R_i = 10 \text{ M}\Omega$ (half-selected 10 K LRS cell), as a function of how many cells are on the line.

as a function of the number of cells on the line. For 8000 cells, the error is about 50% in this case. It implies that for this situation, the expected write current available to the selected cell could be half of the target value

Scaling the crosspoint array to higher density can be achieved by reducing array pitch or stacking additional vertical levels. Practically, the degree to which array pitch can be reduced will be limited by the aspect ratio of the 1S1R structure. For example, the selector and RRAM may each be effectively 20 nm thick, so scaling to 10 nm half-pitch would result in aspect ratios approaching or possibly even exceeding 4, making fabrication quite difficult. However, the vertically stacked crosspoint array fabrication cost per bit is independent of the number of vertical levels. Doubling the number of stacked levels doubles the number of deposition, lithography, and etching steps. Hence, from a manufacturing point of view, there is no economic advantage for 3D integration based on the stacked crosspoint

architecture, compared with 2D. A more cost-effective approach should have the cost of at least some key process steps be independent of the number of vertical levels. This is achieved with the through-multilayer via process, as described in the following section.

20.3.4

Through-Multilayer via Array

As sub-20 nm floating-gate NAND Flash has given way to 3D Vertical NAND based on vertical silicon channels or vertical word lines, RRAM technology can follow the same analogous path. In this case, the RRAM array is formed by depositing the memory layers inside vias formed through entire multilayers. Since the through-multilayer vias are patterned once for all the layers, there is improved cost per bit. However, there are also additional new process steps because necessary staircase-like connections to each of the layers are formed. There have been various integration schemes proposed in the literature [22, 23]. The principal elements specific to these schemes are presented individually in the following sections.

20.3.4.1 Through-Multilayer Vias

The key feature in the new 3D vertical RRAM concept is the vertical bit line constituting a via formed through multiple electrode layers in a single etch step (Figure 20.16). In some cases, such as Macronix's WO_x -based 3D RRAM, the electrode sidewalls within the via are oxidized to form the memory cells; however, this process may lead to extrusion [24].

Alternatively, the memory layer, any additional supporting layers, and finally the vertical electrode metal is deposited and then etched back if necessary inside the via successively [25]. The vertical electrode may be connected to an upper word line instead of a bottom contact, in which case the etch-back after deposition would not be necessary. The horizontal metal electrode lines may be formed by direct etching or by backfilling metal into nitride recesses formed by wet etching a oxide/nitride multilayer [22]. The metal electrode layers may be connected as

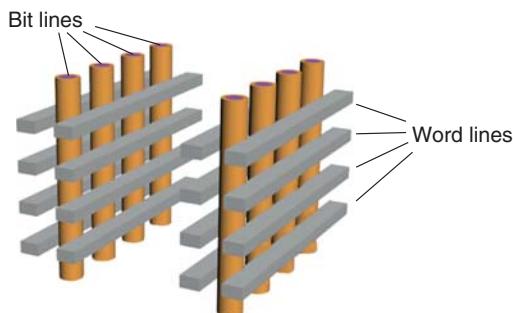


Figure 20.16 Vertical bit lines shared by horizontal word lines. (From Ref. [17].)

entire planes [23], separate lines connected similar to a fork [24], or electrically isolated lines [26].

20.3.4.2 Staircase Connections

The staircase connection is required to connect each of the horizontal electrodes at vertical level to lines at a common horizontal plane above the array. It consists of two difficult parts: (i) the fabrication of the actual staircase profile at the edge of the array, exposing the horizontal electrodes at different levels and (ii) the fabrication of vias of varying depths landing on the horizontal electrodes at different levels.

The fabrication of the staircase was first demonstrated for the BICS 3D NAND Flash by Toshiba in 2007 [27], with the use of repeated reactive ion etching and resist trimming. While inherently simple and free from the use of additional lithography, there is the inherently identical staircase fabrication at the four edges of the array, which may not be the desired outcome. Staircase profiles may be directly fabricated with grayscale lithography [28, 29]. Typically, this is performed with i-line (365 nm) light from an Hg lamp or even longer laser wavelength lithography using very thick ($>1\text{ }\mu\text{m}$) photoresists. As a result, it may not offer the sub-200 nm resolution needed to minimize the staircase area. Another alternative, commonly practiced for binary optics, is the alternate-period masking approach [30], as shown in Figure 20.17. In this approach, a square-wave profile is first fabricated on the surface. Then, every other square-wave period is masked for a subsequent etch. This leads to a repeating four-step staircase profile. Then, the same alternate period masking is applied again, with the period being the staircase period, to form an eight-step staircase profile. This can be repeated to form staircases with 16, 32, and 64 steps. In general, 2^N steps require N masks.

The fabrication of vias landing on the staircase steps can also be achieved with the alternate-period masking approach just described. Or, a single via etch approach may be preferred, where the via etch depth is tailored with the help of via width, taking advantage of etch microloading, or else simply very hard etch stops on the horizontal electrodes.

20.3.4.3 Horizontal Electrodes

The sneak current issue in the through-multilayer array varies from scheme to scheme. When the horizontal electrode is a plane or involves lines connected in a fork, the number of sneak paths can increase significantly with the number of vertical levels. Current can flow up one vertical electrode, then cross over through a horizontal path to another vertical electrode and flow down, then cross over back to the first vertical electrode through a horizontal path at another vertical level.

The horizontal path shape affects the severity of the sneak current issue. For example, the horizontal path may take the form of a plane electrode [23], or a fork or comb shape [31] or a grid-like mesh [24]. In addition, either the plane or the connected horizontal wires constitute a large area capacitor plate. These different architectures are presented in Figure 20.18.

The only way to minimize the sneak current paths as well as the capacitance is to electrically isolate the horizontal electrode lines (Figures 20.19 and 20.20)

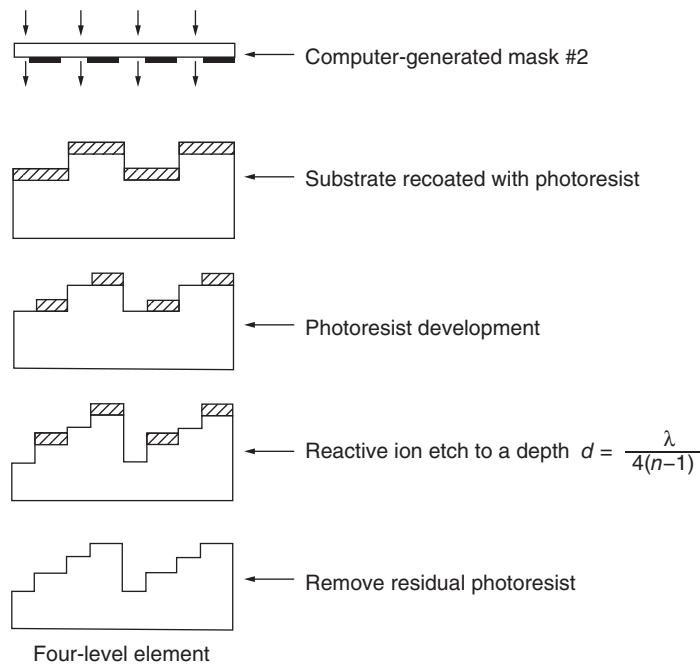


Figure 20.17 Binary-optic style of fabricating the staircase profile for 3D nonvolatile memory peripheral connections. Left: After first mask processing. Right: After second mask processing. (From Ref. [30].)

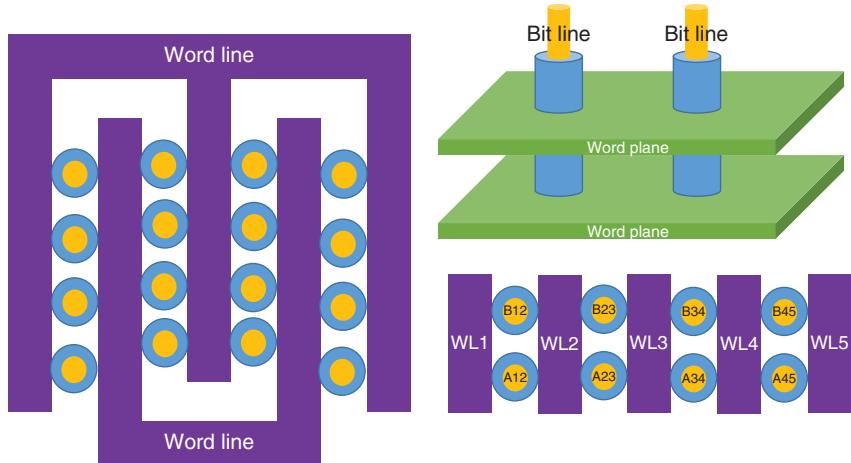


Figure 20.18 Upper right: 3D Vertical RRAM architecture based on horizontal plane electrode, as described in Ref. [23]. Left: Top view of fork-shaped word line, as described

in Ref. [31]. Lower right: Top view of mesh-type connection between cells, as described in Ref. [24]. The resistive switching layers are colored blue.

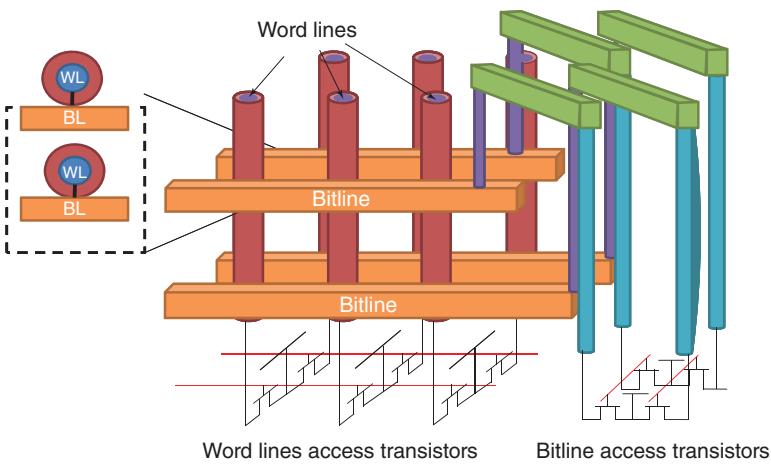


Figure 20.19 3D RRAM architecture using transistor-isolated word lines as described in Ref. [26].



Figure 20.20 3D RRAM architecture using crosspoint cells at sidewall locations, as described in Ref. [32].

[26, 32]. In the first example [26], the array becomes a set of vertical slice subarrays, constituting transistor-isolated pillars connecting to a stack of horizontal lines. These act as locally isolated crosspoint arrays, where the array extent in the vertical direction is defined by the number of vertical levels (N_z). From our previous crosspoint discussion, since N_z is much lower than N_x or N_y for a conventionally stacked crosspoint, that is, $N_x + N_z < N_x + N_y$, the half-select current reduction requirement is relaxed significantly. Furthermore, if all the cells on a given horizontal bit line in the x -direction are put into a very nonlinear high-resistance state, the sneak current among the N_x cells can be suppressed effectively. The sneak current through the N_z cells along the same vertical word line does not require a

strong nonlinearity. For $N_z = 8$, even the low-resistance state nonlinearity of the RRAM itself may be sufficient [26].

In the second example [32], multiple stacks horizontal lines are wrapped and linked by crossing lines. Since the individual bit line stacks are linked and not isolated, the $N_x * N_y * N_z$ 3D array is effectively mapped to an $N_x * N_y N_z$ crosspoint array. Furthermore, the transistor-isolated approach only deals with N_y -isolated $N_x * N_z$ crosspoint arrays for the same size 3D array. However, the decoding would be simpler than for the transistor-isolated architecture, as the latter would have two sets of transistors to be decoded [26].

20.3.4.4 Bathtub-Type Peripheral Connection

The architecture presented in Ref. [32] avoided the use of staircase-based electrical connections, instead making use of a bathtub-type connection, also used in the VSAT 3D-NAND architecture [33]. However, managing the peripheral area for decoding is not trivial, as shown in Figure 20.21. While the bathtub connection avoids the topography patterning challenges of staircase fabrication, the tub edge must be tilted with respect to the array lines, potentially resulting in the bulk portion of the chip area actually being consumed outside the array. This is because $N_y * N_z$ connections are required to be made where N_y connections were traditionally expected, and the tilt increases the extent laterally in both x and y directions, in proportion to $N_y * N_z$.

20.3.5

Array Area Efficiency

All the high-density array architectures considered here focused on reducing the array area, with the underlying assumption that the chip footprint would be reduced accordingly. In fact, the total chip area consists of both the array area and the area of the peripheral circuits surrounding the array. The ratio of the array area to the total chip area is the so-called area efficiency of the array. It can span a wide range, from more than 90% to less than 40%. It depends on the

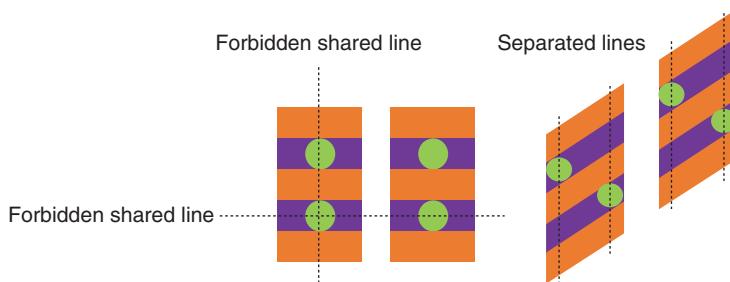


Figure 20.21 Top view of bathtub-type connection for 3D-RRAM. (a) In usually drawn patterns, there is no way for the individual bit lines to achieve their own connection

to outside the array. (b) This can be solved by tilting the tub edge relative to the orientation of the lines. The additional area consumed generally exceeds the array area.

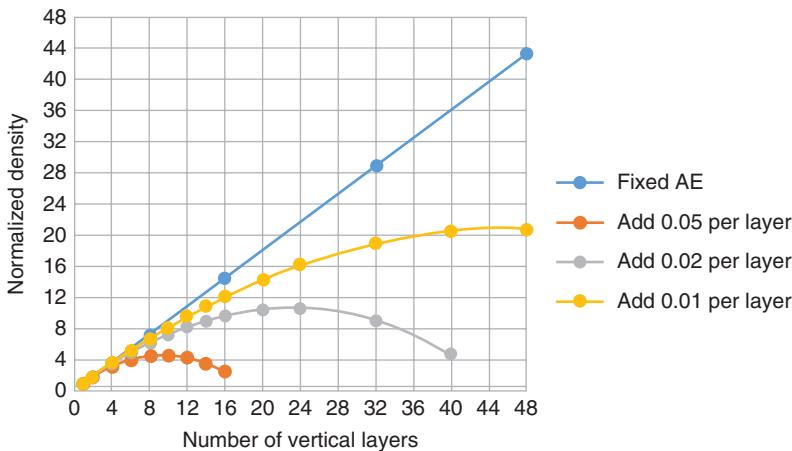


Figure 20.22 Normalized bit density versus number of vertical layers for a 3D NVM architecture. An area efficiency of 90% for the single-layer case is assumed.

memory-controlling functionality of the circuits surrounding the array (such as error correction or verification), as well as the design rule of the process for fabricating the circuitry.

The situation becomes more complicated when 3D architectures are considered. Each vertical device layer added requires the addition of another set of decoders and other control circuitry dedicated to that layer as well as connections to those devices. Even a small percentage added per layer could result in significant density penalties, as shown in Figure 20.22. However, this may be addressed by putting the peripheral circuitry under the array.

When the peripheral area is fixed, the density increases in proportion to the number of layers. But when the peripheral area increases with the number of layers, the increase in density will eventually saturate and then degrade with additional layers. Therefore, to continue 3D scaling effectiveness, conventional 2D scaling would have to be applied at least to the periphery devices, or possibly the periphery devices would be from another wafer connected by through-silicon via (TSV) [34].

20.4 Conclusions

The relative simplicity of the cell structure and process integration of oxide-based RRAM have driven the recent growth of its popularity among memory R&D groups. Now, it is also clear, from product design considerations, that the array architecture, which is directly related to the packing density and the fabrication cost and complexity, plays a key role in defining the device operation specifications as well. 1T1R RRAM for memories embedded in system-on-chip

(SOCs) can be targeted for foundry manufacturing in the near future, while transistorless 1S1R or through-multilayer 3D RRAM arrays are expected to emerge only after further improvements in device nonlinearity and fabrication cost and yield.

Acknowledgments

The authors thank the following individuals for extensive discussions and inspirations: Wei-Su (“George”) Chen, Pei-Yi (“Peggy”) Gu, Lijie Zhang (Samsung), Yi-Jen (“Ian”) Chan (Hermes), Ming-Jinn Tsai, Tzu-Kun Ku, Ming-Jer Kao, Heng-Yuan Lee, Yu-Sheng Chen, Tai-Yuan (“Dove”) Wu, Pang-Shiu Chen (Minghsin University of Science and Technology), Kan-Hsueh Tsai, Wen-Hsing Liu, Yen-Ya Hsu, Chen-Han (“Pierre”) Tsai, Sk. Ziaur Rahaman, Yu-De (“Kido”) Lin, Ming-Hung Lee (National Taiwan Normal University), Chrong Jung Lin (National Tsing-Hua University), Wen-Yueh Jang (Winbond), and Hubert Shih (UMC).

References

1. Lee, C.B., Lee, D.S., Benayad, A., Lee, S.R., Chang, M., Lee, M.-J., Hur, J., Kim, Y.B., Kim, C.J., and Chung, U.-I. (2011) *IEEE Electron Device Lett.*, **32**, 399.
2. Kozicki, M.N., Balakrishnan, M., Gopalan, C., Ratnakumar, C., and Mitkova, M. (2005) Non-Volatile Memory Technology Symposium.
3. Gleixner, B., Pellizzer, F., and Bez, R. (2009) European Phase Change and Ovonics Science Symposium (E/PCOS), p. 135.
4. Hur, J.H., Lee, M.-J., Lee, C.B., Kim, Y.-B., and Kim, C.-J. (2010) *Phys. Rev. B*, **82**, 155321.
5. Chen, P.-S., Chen, Y.-S., and Lee, H.-Y. (2013) *Semicond. Sci. Technol.*, **28**, 025016.
6. Wang, Y.-H., Huang, S.-H., Wang, D.-Y., Shen, K.-H., Chien, C.-W., Kuo, K.-M., Yang S.-Y., and Deng, D.-L. (2012) 2012 International Electron Devices Meeting (IEDM), p. 29.2.
7. Chen, F.T., Lee, H.-Y., Chen, Y.-S., Chen, P.-S., Gu, P., Chen, C.-W., Hsu, Y.-Y., Liu, W.-H., Chen, W.-S., Tsai, M.-J., Lo, S.-C., and Lai, M.-W. (2010) *Curr. Appl. Phys.*, **10**, e75.
8. Gu, P.-Y., Chen, Y.-S., Lee, H.-Y., Chen, P.-S., Liu, W.-H., Chen, W.-S., Hsu, Y.-Y., Chen, F., and Tsai, M.-J. (2010) VLSI Technology, Systems and Applications (VLSI-TSA).
9. Chen, Y.Y., Komura, M., Degraeve, R., Govoreanu, B., Goux, L., Fantini, A., Raghavan, N., Clima, S., Zhang, L., Belmonte, A., Redolfi, A., Kar, G.S., Groeseneken, G., Wouters, D.J., and Jurczak, M. (2013) IEEE International Electron Devices Meeting (IEDM), 2013, p. 10.1.
10. Chen, Y.S., Lee, H.Y., Chen, P.S., Gu, P.Y., Chen, C.W., Lin, W.P., Hsu, Y.Y., Sheu, S.S., Chiang, P.C., Chen, W.S., Chen, F.T., Lien, C.H., and Tsai, M.-J., (2009) International Electron Devices Meeting (IEDM), p. 5.5.
11. Chen, F.T., Lee, H.-Y., Chen, Y.-S., Hsu, Y.-Y., Zhang, L., Chen, P.-S., Chen, W.-S., Gu, P.-Y., Liu, W.-H., Wang, S.-M., Tsai, C.-H., Sheu, S.-S., Tsai, M.-J., and Huang, R. (2011) *Sci. China Inf. Sci.*, **54**, 1073.
12. Hsiung, C.-P., Liao, H.-W., Gan, J.-Y., Wu, T.-B., Hwang, J.-C., Chen, F., and Tsai, M.-J. (2010) *ACS Nano*, **4**, 5414.
13. Chen, F.T., Chen, Y.-S., Lee, H.-Y., Chen, W.-S., Gu, P.-Y., Wu, T.-Y., Tsai, C.-H., Liao, Y.-Y., Chen, P.-S., Sheu, S.-S., Chiu, P.-F., Lin, W.-P., Lin, C.-H., Tsai, M.-J.,

- and Ku, T.-K. (2012) China Semiconductor Technology International Conference (CSTIC).
14. Ho, C.H., Hsu, C.-L., Chen, C.-C., Liu, J.-T., Wu, C.-S., Huang, C.-C., Hu, C.M., and Yang, F.-L. (2010) IEEE International Electron Devices Meeting (IEDM), p. 19.1.
 15. Fackenthal, R., Kitagawa, M., Otsuka, W., Prall, K., Mills, D., Tsutsui, K., Javanifard, J., Tedrow, K., Tsushima, T., Shibahara, Y., and Hush, G. (2014) Solid-State Circuits Conference Digest of Technical Papers (ISSCC), p. 19.7.
 16. Yoo, I.-K., Lee, M.-J., Seo, S.-A., and Seo, D. US Patent 7,821,809, : Non-volatile memory device and method including resistor and transistor. Filed Nov. 7 2005, assigned to Samsung.
 17. Seok, J.Y., Song, S.J., Yoon, J.H., Yoon, K.J., Park, T.H., Kwon, D.E., Lim, H., Kim, G.H., Jeong, D.S., and Hwang, C.S. (2014) *Adv. Funct. Mater.*, **24**, 5316.
 18. Lee, W., Park, J., Shin, J., Woo, J., Kim, S., Choi, G., Jeung, S., Park, S., Lee, D., Cha, E., Lee, H.D., Kim, S.G., Chung, S., and Hwang, H. (2012) Symposium on VLSI Technology, p. 37.
 19. Gopalakrishnan, K., Shenoy, R.S., Rettner, C.T., Virwani, K., Bethune, D.S., Shelby, R.M., Burr, G.W., Kellock, A., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., Jackson, B., Friz, A.M., Topuria, T., Rice, P.M., and Kurdi, B.N. (2010) Symposium on VLSI Technology, p. 205.
 20. Chen, A. (2013) IEEE International Electron Devices Meeting (IEDM), p. 30.3.
 21. Longos, F., Vianello, E., Molas, G., Palma, G., Souchier, E., Carabasse, C., Bernard, M., DeSalvo, B., Bretegnier, D., and Liebault, J. (2013) International Memory Workshop, p. 96.
 22. Baek, I.G., Park, C.J., Ju, H., Seong, D.J., Ahn, H.S., Kim, J.H., Yang, M.K., Song, S.H., Kim, E.M., Park, S.O., Park, C.H., Song, C.W., Jeong, G.T., Choi, S., Kang, H.K., and Chung, C. (2011) IEEE International Electron Devices Meeting (IEDM), p. 737.
 23. Yu, S., Chen, H.-Y., Deng, Y., Gao, B., Jiang, Z., Kang, J., and Wong, H.-S.P. (2013) Symposium on VLSI Technology, p. 158.
 24. Chien, W.C., Lee, F.M., Lin, Y.Y., Lee, M.H., Chen, S.H., Hsieh, C.C., Lai, E.K., Hui, H.H., Huang, Y.K., Yu, C.C., Chen, C.F., Lung, H.L., Hsieh, K.Y., and Chih-Yuan, L. (2012) Multi-layer sidewall WOX resistive memory suitable for 3D RRAM. Symposium on VLSI Technology, p. 153.
 25. Park, J. and Park, Y. US Patent 20130153852: Variable resistance memory devices and methods of forming the same. Filed Nov. 30 2012, assigned to Samsung.
 26. Chen, F.T., Chen, Y.-S., Wu, T.-Y., and Kun, T.-K. (2014) *IEEE Electron Device Lett.*, **35**, pp. 223–225.
 27. Tanaka, H., Kido, M., Yahashi, K., Oomura, M., Katsumata, R., Kito, M., Fukuzumi, Y., Sato, M., Nagata, Y., Matsuoka, Y., Iwata, Y., Aochi, H., and Nitayama, A. (2007) Symposium on VLSI Technology, p. 14.
 28. Gan, O., Barkai, A., Frish, H., Allen, P., Buck, P., Connolly, B., and Pindo, M. (2012) BACUS News, June 2012. Programmed resist sidewall profiles using sub-resolution binary grayscale masks for Si-photonics applications.
 29. Nakao, S., Nakahara, S., and Singubara, S. (2012) *Proc. SPIE*, **8281**, 828116.
 30. Swanson, G.J. (1989) Binary Optics Technology: The Theory and Design of Multi-level Diffractive Optical Elements. Technical Report 854.
 31. Zhang, L., Cosemans, S., Wouters, D.J., Govoreanu, B., Groeseneken, G., and Jurczak, M. (2013) International Memory Workshop, p. 155.
 32. Hsu, C.-W., Wan, C.-C., Wang, I.-T., Chen, M.-C., Lo, C.-L., Lee, Y.-J., Jang, W.-Y., Lin, C.-H., and Hou, T.-H. (2013) 2013 IEEE International Electron Devices Meeting (IEDM), p. 10.4.
 33. Kim, J., Hong, A.J., Kim, S.M., Song, E.B., Park, J.H., Han, J., Choi, S., Jang, D., Moon, J.-T., and Wang, K.L. (2009) Symposium on VLSI Technology, p. 186.
 34. Chen, F.T. 2011 IEEE VLSI Test Symposium (VTS) Special Session.

21

Reliability Aspects

Dirk J. Wouters, Yang-Yin Chen, Andrea Fantini, and Nagarajan Raghavan

21.1

Introduction

For any memory technology, reliability is a major concern. Most widely known reliability characteristics are the data retention (ability to store data for extended periods of time in a specified temperature range) and the cyclability or endurance (ability to rewrite the data multiple times). However, in general, memory reliability includes many different aspects related to data storage (retention), to writing (program variability, endurance, and write disturb), and to reading processes (read disturb and random telegraph noise or RTN).

Different memory applications, as stand-alone nonvolatile memories (NOR or NAND Flash-type memories), dynamic and static RAMs, and embedded nonvolatile memories each have specific requirements for all these reliability properties. Although there is in general some trade-off possible between different reliability aspects (as, e.g., between endurance and retention), there is no single universal memory technology that can cover all these requirements, and as a result different technologies are used for different applications.

The drive for the development of new emerging memory technologies, such as resistive switching RAM (RRAM), is coming from scaling issues that most – if not all – of the current (charge-based) memory technology face beyond the 1Xnm technology node. Although we do see a steady improvement of the reliability characteristics of RRAM as the technology is becoming better controlled, it is important to assess the real limitations (and/or possible mitigations) to assess the potential of RRAM to substitute for one of these technologies in further scaled nodes.

Therefore, this chapter tries to give a comprehensive overview of the different aspects of the RRAM reliability focusing not only on “best” achieved specifications, but also on the models and understanding of the reliability physics involved.

In relation to the type of RRAM cell, this chapter focuses on the reliability aspects of (filamentary) bipolar switching transition metal-oxide (TMO) cells.

21.2

Endurance (Cyclability)

21.2.1

Endurance Summary of Bipolar Switching TMO RRAM

Most memory applications require the ability to rewrite stored data. This is measured as the maximum numbers of alternative state (0/1) program/read cycles that can be applied without failure, that is, cyclability or “endurance.” Endurance and data retention (see Section 21.3) are two of the most important reliability characteristics defining the application domain of a certain memory technology.

Filamentary-based bipolar switching TMO RRAM, in general, demonstrates good cyclability. Table 21.1 summarizes the achieved number of cycles for different TMO stacks, on single-cell basis [1–10]. Compared with floating gate memories, which typically fail after 10^4 – 10^5 cycles, much higher endurance (up to 10^{12} cycles) is achieved in RRAM. Over the years (2008–2012), the maximum endurance cycle number on RRAM has also increased (Figure 21.1a, [1–5, 7, 8, 10]). The high cyclability potential of RRAM enables applications in a broad range of the memory hierarchy, covering on the one hand the mass storage, low

Table 21.1 Summary of high-endurance RRAM devices.

	TMOs	SET	RESET	Cycles
SAIT [1, 2]	TaO based	4.5 V, 10 ns	7 V, 10 ns	10^{12}
ITRI [3, 4]	HfO ₂ /Ti	3.2 V, 40 ns	2.7 V, 40 ns	10^{10}
IMEC [5, 6]	HfO ₂ /Hf, Ti	1.8 V, 5 ns	1.8 V, 10 ns	10^{10}
HP [7]	TaO based	1.9 V, 1 μ s	2.2 V, 1 μ s	1.5×10^{10}
Panasonic [8, 9]	TaO based	1.5 V, 100 ns	2 V, 100 ns	10^9
SEMATECH [10]	HfO _x	1.5 V, 50 ns	1.5 V, 50 ns	> 10^9

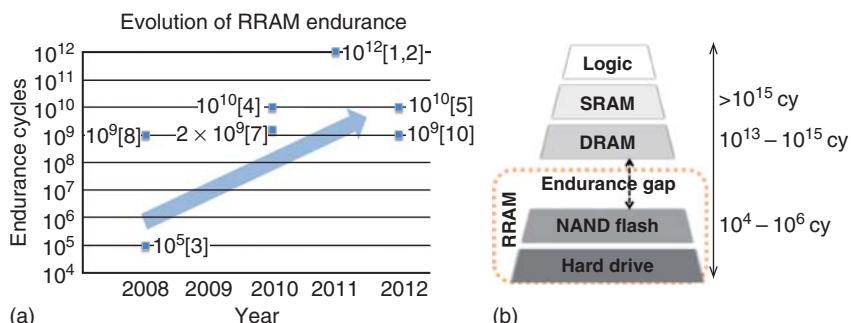


Figure 21.1 (a) High-endurance RRAM devices in publications, 2008–2012. (b) RRAM position in memory hierarchy.

cyclability part, while on the other hand bridging the endurance gap between Flash memory and DRAM (illustrated in Figure 21.1b).

Although TMO RRAM demonstrates potential high endurance, concerns remain especially due to its defect-based filamentary switching nature. Uncontrolled, stochastic changes during the defect switching may lead to variable endurance degradation, which needs to be better understood.

21.2.2

Balancing the Bipolar Switching for Better Endurance

Based on the understanding of bipolar switching in filamentary RRAM, SET/RESET switching is achieved by drift of defects (oxygen vacancy) in the oxide. In this view, it is important to drift an equal amount of oxygen vacancies forth and back during each SET/RESET operation in order to maintain the same levels of both the low-resistance state (LRS) and high-resistance state (HRS). Over-SET or over-RESET, which drifts an excessive amount of oxygen vacancies toward either LRS or HRS states, will result in the possible failure of the following SET/RESET operations.

In order to balance the defect drift during SET/RESET operations, a delicate tuning of SET/RESET switching conditions is necessary [5]. Through such tuning of SET/RESET conditions, as shown in Figure 21.2, the failure of either over-SET

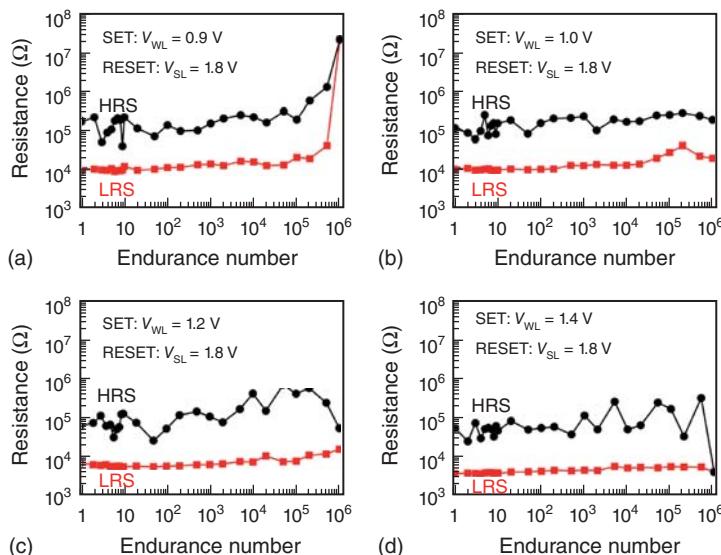


Figure 21.2 Pulse endurance behavior of the 40 nm Hf/HfO₂ 1T1R devices, with fixed RESET pulse at word line (WL) = 3 V, source line (SL) = 1.8 V, and 10 ns. The SET pulse amplitude was varied using different WL pulses: (a) 0.9 V, (b) 1.0 V, (c) 1.2 V, and (d) 1.4 V. The SET pulse width

is fixed as 100 ns, 1.8 V on the bit line (BL). With increasing SET WL voltage, endurance failure mode shifts from LRS failure (a) to HRS failure (d). (© 2012 IEEE. Reprinted, with permission, from Ref. [5].)

or over-RESET can be avoided and endurance can be substantially improved. Though this balance point of SET/RESET operations may be material stack specific, balancing the switching operations is a general requirement for the bipolar oxygen-vacancy-based RRAM devices.

21.2.3

Understanding of Endurance Degradation

As demonstrated in Refs. [11–13], endurance degradation in bipolar oxygen vacancy RRAM can show up as LRS and/or HRS state degradation. Different models are proposed to explain the degradation behavior, which can be mainly divided into two types: failure to RESET to high enough HRS state and failure to SET to low enough LRS state. Hereby, RESET failure (HRS degradation) is typically attributed to the exhaust of oxygen (e.g., due to nonideal drift of oxygen), resulting in incomplete recombination of oxygen and oxygen vacancies so that a larger, oxygen-vacancy-controlled conduction path remains. Alternatively, the filament may gradually become too big during SET (e.g., by extra oxygen generation during switching), eventually prohibiting sufficient RESET. For example, three different RESET failure types are discussed by B. Chen *et al.* [11], each with a different signature in their HRS and LRS behavior.

SET failure (LRS degradation), however, can be attributed to local material changes. For instance, the SET failure model in Ref. [12] is based on an increase of the SET voltage due to a local modification (“recrystallization”) of the oxide material in the filament region, caused by the high temperatures (~ 895 K for a $30\text{ k}\Omega$ LRS state) generated by the current during switching. As a consequence, the mobility of oxygen vacancy is reduced, and the SET operation becomes unsuccessful. In general, the switching power/energy and the temperature in the oxygen vacancy filament are important parameters impacting both LRS and HRS degradations. Due to filamentary conduction, the temperature and power distribution is strongly nonuniform, a local high temperature and high power may modify or even damage the original atomic configuration of the oxide host material.

Cycling-induced degradation can be further related to the thermodynamic (in)stability of the filament and the host matrix, and has to be considered in the material design of the resistive switching system [7]. For example, Ta oxide system has exhibited a small variance from switching with cycling and thus a high endurance. The reason for this may be that this oxide system has only two stable material phases, that is, Ta–O solid solution (filament) and Ta pentoxide (host matrix). These two material phases do not react with each other thermally to form another material phase even at high temperatures induced by Joule heating. Furthermore, there is a large oxygen solubility in the Ta–O solid solution phase, which allows for the filament to accommodate and release oxygen ions without a phase change during cycling. An almost identical system is Hf oxide system, which has exhibited similar cycling ability.

Simulations may help to model and better understand the endurance degradation process. For example, a Monte-Carlo simulation on atomic scale is reported

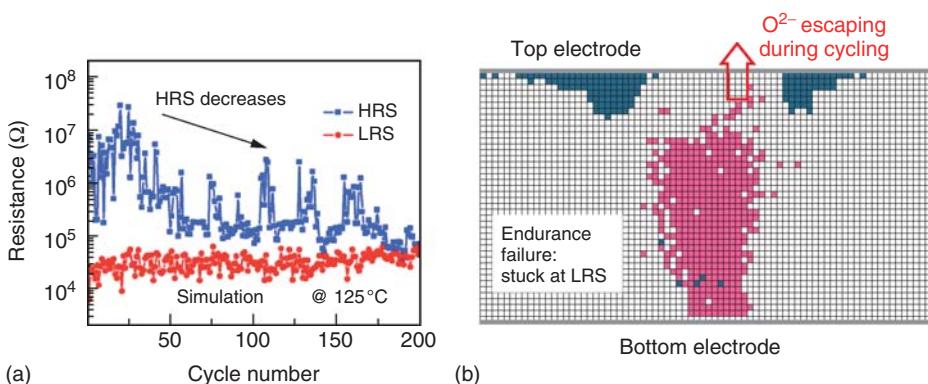


Figure 21.3 (a) Simulated endurance at 125°C. HRS decreases, and the final failure is at LRS. (b) V_{ox} (pink) and O^{2-} (blue) distribution at the endurance failure at the

end of cycling in (a). Insufficient O^{2-} at the interface makes the reset impossible for the RRAM cell. (© 2012 IEEE. Reprinted, with permission, from Ref. [13].)

Table 21.2 Summary of retention degradation behaviors in RRAM devices.

	TMOs	LRS	HRS
IMEC [6, 12]	HfO_2/Hf	Current decreases	Current decreases/increases
Panasonic [14, 15]	TaO_x	Current decreases	Current increases
Politecnico di Milano [16]	NiO	Current decreases	N/A
Stanford University [17]	HfO_2	Current decreases	N/A
Fudan University [18]	CuSiO	Current decreases	N/A

by S. Yu *et al.*, [13]. They could simulate HRS failure, by taking into account the oxygen and oxygen vacancy recombination during switching (see Figure 21.3). The escape of oxygen during switching results in an incomplete recombination of the oxygen vacancies during RESET and leads to the HRS degradation. This again points out the importance of balancing the oxygen drift during SET/RESET conditions for a better endurance performance of RRAM devices.

21.3 Retention

21.3.1 Retention Summary of Bipolar TMO RRAM

As RRAM is targeting nonvolatile memory applications, the retention property of RRAM is a critical reliability aspect. Typical nonvolatile memory requires 5–10 years data retention up to 85–125 °C. Various studies (Table 21.2) [6, 12, 14–18] have analyzed the retention behavior of the LRS and HRS states of oxygen vacancy filamentary RRAMs. In general, after a certain period, the current

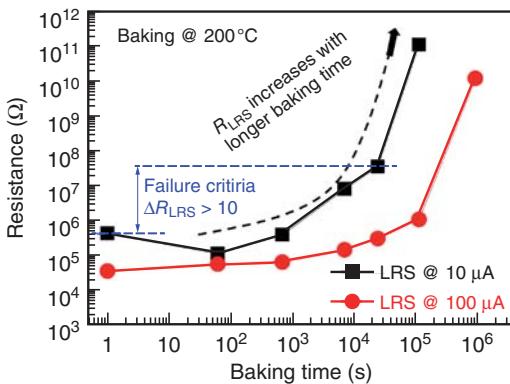


Figure 21.4 Typical LRS retention failure of a 40 nm HfO_2/Hf RRAM, for LRS programmed by either 100 μA CC or 10 μA CC. The LRS retention failure criterion was defined as a 10x LRS resistance increases. (© 2012 IEEE. Reprinted, with permission, from Ref. [12].)

(resistance) of LRS tends to decrease (increase), which indicates a degradation of the conducting filament (CF) due to a loss of oxygen vacancies. The current (resistance) of HRS shows mixed behaviors, and both increase and decrease have been observed. Figure 21.4 shows the typical LRS resistance increase for HfO_2 -based RRAM. Understanding and optimizing the retention property of RRAM is critical and challenging, due to the complex and even stochastic behavior of the oxygen vacancy defects. In general, LRS degradation is the more critical reliability issue, and will be the focus here.

Though the oxygen vacancy filamentary switching has stochastic nature, the (LRS) retention degradation on a large statistical basis follows the classical Arrhenius law dependence with temperature. Figure 21.5 demonstrates the Arrhenius behavior of a typical $\text{HfO}_2/\text{metal cap}$ (Hf) RRAM system [12]. The LRS

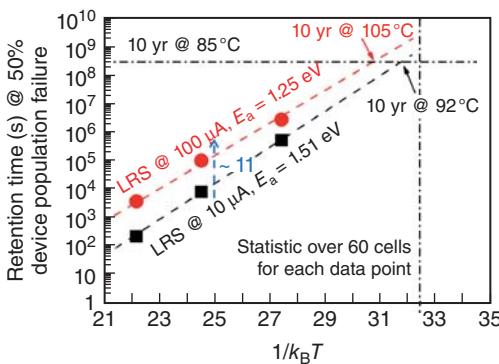


Figure 21.5 The LRS retention (programmed by 100 and 10 μA) measured at 150, 200, and 250 $^\circ\text{C}$. The retention follows the Arrhenius law, with $E_a = 1.2 - 1.5 \text{ eV}$. (© 2012 IEEE. Reprinted, with permission, from Ref. [12].)

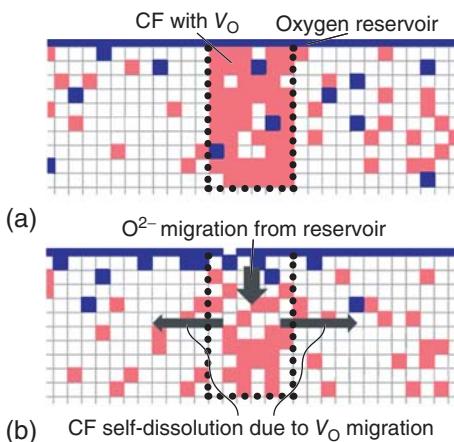


Figure 21.6 (a) An example of the initial filament configuration in LRS: pink sites are V_{ox} , blue sites are O^{2-} , the top boundary is the oxygen reservoir at the electrode/oxide interface. (b) An illustration of oxygen and

oxygen vacancy migration processes during baking. (Reproduced with permission from Ref. [17]. Copyright 2012, AIP Publishing LLC.)

retention degradation is clearly accelerated with increasing stress temperature, with an extracted energy barrier E_a of 1.2–1.5 eV. The diffusion of oxygen and oxygen vacancies inside the host oxide of the RRAM cells is responsible for the retention degradation, as is discussed later.

21.3.2

Understanding of Retention Degradation in Bipolar TMO RRAM

In the oxygen-vacancy-based RRAM, the possible (LRS) degradation mechanisms considered are as follows (Figure 21.6 [17]):

- 1) diffusion of oxygen from outside of the filament (i.e., from electrode or from surrounding oxide region) and recombination with an oxygen vacancy in the filament;
- 2) diffusion of an oxygen vacancy¹⁾ out of the filament.

Both mechanisms will result in a decrease of the number of oxygen vacancies in the filament, which leads to an increase of the LRS resistance (shown in Figure 21.4). Which mechanism – oxygen diffusion (process 1) or vacancy diffusion process 2) – dominates, may depend on both the material stack and LRS program conditions.

1) While in both processes, the actual moving species is oxygen, we can make a distinction between oxygen diffusion, which we relate to the diffusion of interstitial (excess) oxygen, and oxygen vacancy diffusion, which we relate to the substitutional diffusion of oxygen. Analogous to the hole conduction model in semiconductors as Si, the latter process is more easily described by the motion of a virtual oxygen vacancy particle.

A general trend on the LRS retention behavior is that the retention improves with higher LRS current (lower LRS resistance). For lower LRS resistances, the filament consists of more oxygen vacancies, making it more retention robust as the relative change of the filament current is smaller for the same amount of vacancies lost. That is, the relationship between number of oxygen vacancies (n_c) in the filament constriction,²⁾ and the LRS resistance can be calculated using the quantum point contact (QPC) filament conduction model [15]. The nonlinear relationship between number of oxygen vacancies and the resistance explains the fact that a disappearing oxygen vacancy has a much larger relative impact for a narrow filament. The current dependence of the LRS³⁾ retention needs to be considered when evaluating the retention property of a particular device.

Another observation is a degradation of the data retention after cycling (at the same programming current level) [19]. The physical mechanism proposed is a loss of oxygen vacancies in the filament after cycling. This postcycling retention is a coupled reliability issue of both cycling and retention, which depends on the cycling programming condition as well. By optimizing the cycling properties, the postcycling retention can be improved [6].

21.3.3

Trade-Off between Retention/Endurance

As both the endurance and retention reliability depend on the oxygen vacancy filament, optimization of one property may impact the other. As identified in [6], an endurance and retention trade-off can be observed in HfO₂-based RRAM, in this case by modifying the oxygen-scavenging layers in the memory element stack. Oxygen-scavenging layers are used in some types of RRAM devices with the effect of creating oxygen vacancies in the metal oxide, lowering the required forming conditions to create the switching filament. Using a strong oxygen-scavenging layer capability, better resistance window and longer endurance are achieved. The better cyclability results from the fact that the RESET failure can be compensated due to the larger amount of available oxygen. Furthermore, better retention can be achieved in the weak oxygen-scavenging layer property case. The smaller amount of scavenged oxygen gives better retention, due to the reduced availability of oxygen that may diffuse and recombine with the oxygen vacancies in the filament.

Endurance and retention trade-off may be achieved by modifying either material or electrical parameters in the oxygen vacancy filament RRAM. Essentially,

- 2) The constriction is the smallest, conduction controlling part of the filament. Most RRAM switching models assume that during SET/RESET filament changes mainly occur in this part of the filament only.
- 3) The definition of LRS and HRS is actually arbitrary, and depends on the actual resistance levels targeted in an application. The same resistance level can be HRS in a higher power application, while LRS in a low-power application, so more properly we should discuss retention as function of resistance (or current levels).

however, due to the two terminal structure of RRAM, cycling, and the retention are always coupled, and have to be optimized together.

21.4 Variability

21.4.1 Introduction

Although aggressive scalability and easy manufacturability of RRAM have been demonstrated [3, 20, 21], one of the major issues of RRAM is the stochastic variability in the device operation that still needs further understanding to enable its commercialization.

In the well-established Flash NAND technology, random device-to-device (D2D) fluctuations of the programmed state occur as a consequence of process-related device scaling. In this case, we properly discuss *variability* of device characteristics such as channel width, length (W, L), drain current IDS , and threshold voltage VT . However, if we then look into one of these devices and we monitor the evolution of the programmed state during the device lifetime, we observe a constant state value⁴⁾ between consecutive readouts (not accounting here for a long-term change of the state due to degradation mechanisms).

Resistive memories, due to their fundamentally different operating mechanism, additionally display significant intradevice, cycle-to-cycle (C2C) dispersion from the very beginning of device lifetime. In this sense, we should distinguish the typical *process variability* from the *stochastic variability* where each characteristic is subject to *independent and (ideally) uncorrelated* C2C fluctuations.

As resistive memories store information as a logical value associated to the device resistance state, and transitions between these resistance states are obtained by voltage stimuli of appropriate magnitude (and polarity), the stochasticity of the RRAM operation mechanism is reflected in both the variability of the required switching voltage, and in the variability of the resultant resistance state. Furthermore, the initial device forming step strongly influences its variability.

21.4.2 Experimental Aspects of Variability

21.4.2.1 Variability of Forming Operation

Differently from FLASH technology, resistive memory involves a one-time initialization step called *electroforming*. During this step, a conductive channel is created within the oxide matrix, which hosts the subsequent programming operation.

4) Flash also is affected by program variability (one to a few electrons C2C difference) and is also affected by RTN variations in the read current; however the overall variability is much smaller compared to RRAM.

This forming process has been studied within the context of classic oxide breakdown theory via ramped voltage stress (RVS) or constant voltage stress (CVS) [22, 23]. Distributions of forming voltage and its respective forming time have been shown to obey Weibull (or weakest-link) statistics for both polycrystalline and amorphous oxides. It has been shown [24] that the power involved in this operation typically affects subsequent switching operation. For instance, forming performed at much higher power than used in following switching cycles generally decreases the device resistance window and uniformity. For this reason, optimization of the forming operation is a subject of growing interest. Measures that minimize the randomness of filament formation, for example, by modifying device geometry, introducing filament precursors [25] or optimizing electrical operation protocols, could help to reduce the formation variance from device to device.

21.4.2.2 Intrinsic and Extrinsic Variability

Even assuming an ideal forming operation, program operations are not self-limiting as in the Flash scenario. For instance, applying voltage programming, the SET operation requires a current limiting device to prevent device breakdown while, conversely, in a current driven programming, RESET operation may need a voltage limiting device to prevent degradation. In real applications, the memory device, sometimes called *1R*, is thus usually coupled together with some limiting device, typically a transistor, a diode, or a limiting resistor, giving rise to the so-called “1T1R,” “1S1R,” and “1R1R” configurations. However, it is important to realize that also these current limiting devices are themselves affected by a spread of characteristics or are subject to parasitic components such as stray capacitances, the latter of these being responsible for the poor control of switching properties in the early stage of technology development.

Thus, care has to be taken to decouple the characteristics of the current limiting device from those of the memory element. In this sense, it is proper to speak of *intrinsic variability* of the memory device (related to the randomness of the switching process) and *extrinsic variability* associated with the combined effect of memory device and cell (and array) architecture.

From an experimental point of view, the study of the intrinsic memory device variability can be accomplished either by using nearly ideal limiting device (such as a large area long channel metal-oxide-semiconductor field-effect transistor (MOSFET)) or by carefully compensating for additional contributions.

System-level studies of the combined impact of the variability of both the memory and the limiting devices are, however, lacking, while in particular important for assessing the performance in scaled technologies at the array level. Architectural studies presented so far [26, 27] typically focus on the impact of access line resistance, selector nonlinearity, and biasing scheme on the overall power consumption assuming uniform memory device characteristics, and the power penalty introduced by nonuniformity in the limiting device itself has still to be carefully addressed.

21.4.3

Physical Aspects of Variability

21.4.3.1 Variability in Unipolar Devices

In the early stage of RRAM development, studying unipolar switching devices, variability was mainly viewed as the spread of SET and RESET switching voltage, while variability of the programmed resistance state itself was less of a concern. This can be understood considering that unipolar operation, involving a thermally assisted rupture of the filament, is characterized by relatively high operating current and high ON–OFF ratio. So, the resistance window did not pose large concerns; however, the RESET voltage required to disrupt the filament and the SET voltage required to reform it are subject to wide fluctuations. More importantly, since in this case both operations are done in the same voltage polarity, an overlap between the two voltage distributions may induce unwanted parasitic SET during RESET operation leading to strong switching instability. One approach to qualitatively model variability in this approach is linked to the “Random Circuit Breaker” model [28]. More recently, the same model was updated to take into account interfacial properties in order to model bipolar switching behavior as well [29].

21.4.3.2 Variability in Bipolar Devices

As the research focused more and more on bipolar switching devices, the need for an accurate estimation of resistance variability became prominent. This is both due to the fact that the resistance window is much smaller than in the unipolar case and that the opposite polarity of the operation voltage solves the problem of SET–RESET instability typical of unipolar devices.

Experimentally, independently from the particular oxide or stack used, all bipolar switching (oxide) RRAMs show similar switching characteristics. The resistance distribution for both ON and OFF states can in first approximation be described by means of a lognormal distribution, and to establish a useful figure of merit it is possible to consider the standard deviation of resistance distribution σ_R normalized by median resistance σ_R/R . Using this criterion, it becomes apparent that lowering the operating current increases not only the absolute spread (as it would be expected from a fluctuation proportional to filament “radius”) but also the relative spread. In a simplified view, this behavior is consistent with a picture where the same fluctuation of the filament radius affects a thinner (“weaker”) filament more strongly than a thicker (“stronger”) filament up to the limit where the filament size may be so small that its geometry affects the conduction much more than its radius. This is captured by the σ_R/R trend in Figure 21.7 clearly depicting a change in regime for extremely resistive (=“thin”) filaments. Trying to better define what defines the filament “radius” in bipolar devices, it has been recently evidenced that switching operation involves a change into a number of conductive defects (where V_O^{2+} oxygen vacancies are generally accepted as the dominant defect in TMO oxides) either by forming a percolation path [31], a conductive sub-band [32], or a quantum point-defined constriction [33]. Figure 21.8 illustrates the microscopic origin of resistive variability in a quantum

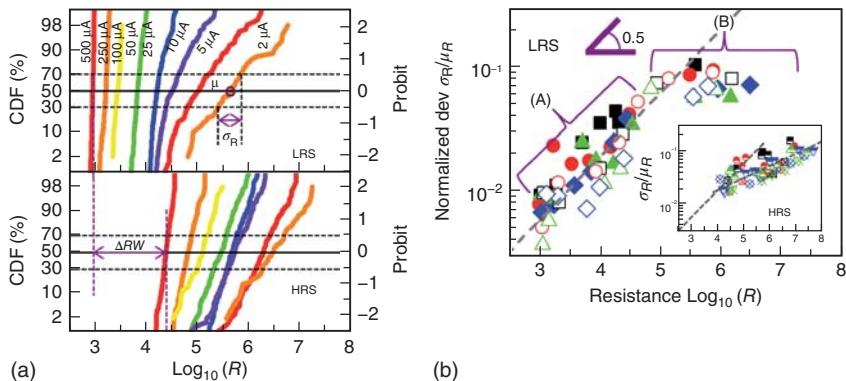


Figure 21.7 (a) Resistance distribution of LRS and HRS states obtained at different compliance currents for a TiN/HfO₂/Hf/TiN stack. The decrease of operating current induces an increase of median and dispersion of resistances for both states.

(b) Normalized standard deviation versus median resistance for LRS states and (inset) HRS states for different statistic, compliance currents, and stack types. Two clearly different regions can be identified. (© 2012 IEEE. Reprinted, with permission, from Ref. [30].)

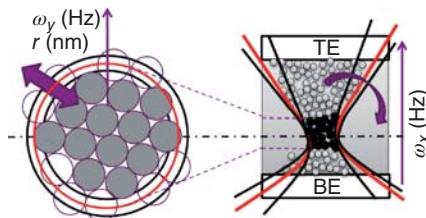


Figure 21.8 Microscopic origin of resistance variability as induced by fluctuation in the number and geometry of discrete defects defining a quantum point defined conductive path. (© 2012 IEEE. Reprinted, with permission, from Ref. [30].)

point conduction path model [33]. A quantitative explanation for LRS resistance variability is given in Ref. [34].

In bipolar switching device, the modeling and physical description of variability becomes thus linked to the role of the oxygen vacancies and consequently cannot be decoupled from the particular switching mechanism proposed. Two mainstream views are proposed: in one switching is described as the generation and recombination of oxygen defects [31, 32], and in the other the switching action is instead caused due to a (ionic) movement of the same defect without any change in their number [33, 35].

- When switching is explained in terms of generation/recombination of oxygen vacancies [31], the resistance in the LRS state is produced by the creation of a percolating network of defects (oxygen vacancies) while the HRS state is due to the creation of defect-free gap due to defect recombination (i.e.,

oxidation of the vacancies). In this sense, the spread in HRS can be justified by the random generation of defects in the gap area, while for the percolation path regeneration during SET action can be described in a similar way as standard time-dependent dielectric breakdown (TDDB) degradation in oxides, and should follow Weibull statistics.

- 2) When an ion-movement model is considered [36, 30, 34], ions are moving from preexisting reservoirs located at the top electrode (TE) and bottom electrode (BE) to enlarge or shrink (and maybe even disrupt) the filament. The source of variation in this case is linked to the number of discrete defects defining the radius of a single filament. Concerning LRS variability, this view matches well with the experimentally evidenced Poissonian behavior of resistance when modeling defects as independently emitted in time or space. In this sense, the SET action describes the “effort” necessary to “nucleate” a filament either by injecting ions into a gap or by gradually changing the shape of quantum-defined CF.

21.5

Random Telegraph Noise (RTN)

21.5.1

Introduction

Another important metric to be considered in assessing the robustness of RRAM is the RTN phenomenon, which is intrinsic to any dielectric with defects (traps). The presence of RTN can cause a large spread in the distribution of the HRS and LRS and induce “soft errors” in reading the wrong memory state (if the memory window is relatively small). Although RTN, which is relevant mainly at read voltage ($V_{\text{READ}} = 0.1$ V) conditions, does not cause irreversible damage unlike endurance test conditions, it affects the variability of the resistance distribution [37–39] and introduces large magnitude of noise in addition to the standard $1/f$ flicker noise and thermal white noise. Detection of RTN signals also serves as a spectroscopy tool enabling us to determine the trap properties (spatial location from dielectric – electrode interface and energy depth below conduction band of dielectric) [40]. As we will discuss in the following sections, the RTN signal can also help to characterize the shape and size of the filament in the HRS state, taking the QPC formulation for the defect cluster [24].

There are two major types of RTN signals observed. One is the steady-state fluctuations involving stochastic electron capture and emission events [37, 40–42] through the defects with their corresponding time constants (which depend on applied voltage, trap position, and trap energy). The other components are the non-steady-state fluctuations that arise due to structural disturbances in the CF due to removal or addition of oxygen vacancies [42]. It is important to note that resistive switching can occur due to two mechanisms: (i) oxide-based RRAM (OXRAM) where switching is caused by oxygen vacancy/ion generation – recombination and drift/diffusion and (ii) CBRAM (conducting

bridge RAM), where switching is caused by nucleation and rupture of metallic filament due to ionic migration/electromigration. Considering that the fundamental physics of switching may be completely different for these two cases, it is obvious that the kinetics of RTN will also be very different. Our analysis here is predominantly focused only on RTN in OXRAM, as there are very few studies that have been carried out for RTN in CBRAM [43]. For the rest of this chapter, we first investigate the steady-state RTN, followed by the non-steady-state component of RTN. The dependence of RTN on the compliance level, dielectric material, and microstructure are addressed as we proceed.

21.5.2

Charge Carrier Transport-Induced RTN

As mentioned in the previous section, the steady-state component of RTN is mainly attributed to electron capture and emission process through the oxygen vacancy defects (traps) in the dielectric, by an inelastic multiphonon trap-assisted tunneling (ITAT) process [44]. Another possibility is the Coulomb repulsion effect [45] where charged defects in the vicinity of the filament can restrict the conductivity of the filament (reduced effective filament cross section) due to Coulombic interactions. The RTN signal that is measured is an indicator of the number of “critical traps” that affect the stability of the resistance state. In the HRS, the RTN trends are more clearly observed. If the signal is sensed for a long period of time and N distinct current levels are observed (this can be detected using a hidden Markov model (time-lag plot)-based analysis [44]), this implies that there are $\log_2(N)$ traps present in the dielectric. Most signals observed are multilevel RTN, where each two-level deconvoluted RTN arises from the stochastic electron capture – emission events in one single oxygen vacancy defect.

With deeper reset, there are less number of active defects and RTN signals tend to be more discrete with larger spread in the current ($\Delta I/I$) (Figure 21.9), as clearly shown by the statistical RTN study by Veksler *et al.* [45] and Ielmini *et al.* [46]. Ideally, if the RRAM can be reset to very deep states with zero defects, then we can achieve very good noise immunity. However, this is in most cases not feasible. As for the LRS state, for typical high compliance levels of $100\text{ }\mu\text{A} - 1\text{ mA}$, there are far too many traps and the sum of many RTN signals with different time constant and current step distributions will average out to produce a $1/f$ random noise signal [47], with very low ($\Delta I/I$). Therefore, RTN in the LRS state is generally not a critical issue, unless we talk about ultra-low power-switching devices with forming/SET compliance levels as low as $1\text{ }\mu\text{A}$ (which currently show very low endurance and retention).

Even in the HRS state, the values of ($\Delta I/I$) are relatively small ($\sim 10 - 100\%$) [38, 39] for the charge carrier transport compared with the vacancy fluctuations in the filament (structural disturbances), for which the value of ($\Delta I/I$) can be even an order of magnitude at times [42]. Moreover, since structural disturbances are mostly irreversible (only occasionally the current levels jump back to their initial state), their role is more detrimental in the stability of the memory state. In the

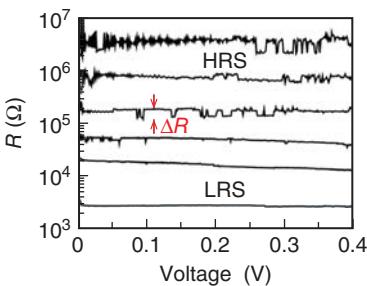


Figure 21.9 Pattern of the resistance ($R = V/I$) signal as a function of the resistance state. In the HRS with few defects, the RTN Lorentzian signal from every defect is more apparent, resulting in large values of $(\Delta I/I)$. In the LRS, due to a large number of defects with widely distributed time constants, the Lorentzian RTN signals add up

and the fluctuations average out to give a 1/f noise trend with very low $(\Delta I/I)$. Note that the fluctuations appear to be small in the HRS as well, which is an artifact – this is because the resistance data is plotted on a logarithmic scale. (Reproduced with permission from Ref. [46]. Copyright 2010, AIP Publishing LLC.)

next section, we focus on these vacancy-induced RTN effects. Before we discuss that, some of the recent noteworthy references for carrier-induced RTN are the works done by Lee *et al.* on TiO_x (20 nm) [40], Puglisi *et al.* on HfO_2 (5 nm) [44], and Ielmini *et al.* on NiO (20 nm) [46].

21.5.3 Oxygen Vacancy Transport-Induced RTN

21.5.3.1 Experimental Identification of Vacancy Perturbations

Given any measured RTN signal, the first task to be carried out is the identification of the different RTN jumps and finding out which of these jumps correspond to vacancy-induced fluctuations and which are due to carrier transport-based fluctuations. A physical model formulation using the QPC model is used in this context. The QPC formulation has been previously proposed by Miranda and Suñé [48] and Cester *et al.* [49] to describe the nonlinear conduction in the post-soft-breakdown regime for high- κ dielectrics with very good fit to the measured $I-V$ data. This formulation has been adapted successfully to describe the conduction in RRAM for both the HRS and LRS states by Degraeve *et al.* [33] and Suñé *et al.* recently [50]. As proposed by Degraeve and co-workers, the filament can be phenomenologically represented as a cluster of vacancies with a narrow “constriction” comprising N_C particles (vacancies). It is this constriction that controls the conductivity of the state. The size of this constriction is governed by N_C and the shape is determined by two parabolic energy band frequency parameters, ω_X and ω_Y , where these two quantities qualitatively indicate the length and width of the constriction, respectively. The lower the ω_X , the longer the constriction; the lower the ω_Y , the wider the constriction. Using the QPC model, the $I-V$ curve for any integer value of N_C can be simulated. To identify vacancy-induced RTN effects, the measured $I-V$ sweep

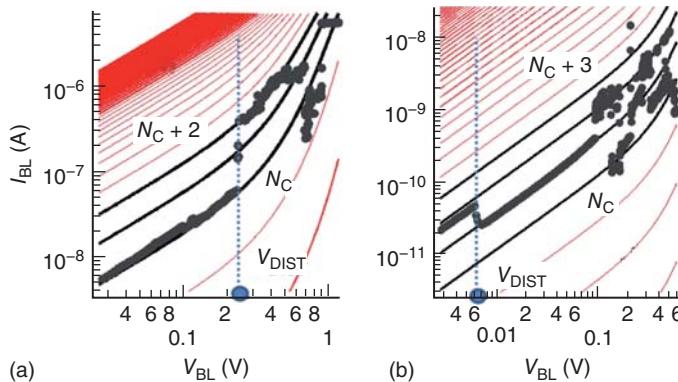


Figure 21.10 Methodology to extract the value of the disturb voltage (V_{DIST}) for every SET cycle. The I - V plots simulated for different number of constriction defects (N_C) is superimposed on to the electrical I - V measurement data, for the case of (a) shallow and (b) deep RESET. The first voltage level at which the current consistently jumps

to a neighboring I - V curve is classified as the V_{DIST} value. Note that RTN trends still exist even for $V_{\text{BL}} < V_{\text{DIST}}$ (not visible here in the logarithmic scale); however, they are small in magnitude and correspond to the electron capture – emission process. (© 2013 IEEE. Reprinted, with permission, from Ref. [51].)

data can be superimposed on to the I - V simulated curves and the data and its multiple jumps can be fit to the QPC simulation model by optimizing the $\{\omega_X, \omega_Y\}$ values. The first voltage level at which a jump in the I - V data occurs from one level of N_C to the adjacent one (either $N_C + 1$ or $N_C - 1$) is classified as the “disturb” voltage (V_{DIST}) [51], which is an indicator of the stability of the HRS state against vacancy-perturbation-induced RTN. Figure 21.10 illustrates this methodology of V_{DIST} identification. Along with the value of V_{DIST} , the model also provides us with the initial value of N_C . Note here that the underlying assumption is that the filament does not rupture during RESET; instead, it only shrinks in size. However, as shown later, this model can be extended to analyze the case of ruptured filaments as well.

21.5.3.2 Vacancy-Induced RTN for Shallow to Moderate Reset

The higher the forming/SET compliance (I_{comp}), the shallower the reset is expected to be, as there are more ion–vacancy recombination events needed for a given reset sweep as I_{comp} is increased. From a logical perspective, we would expect the value of V_{DIST} to be higher for lower I_{comp} (deeper reset). However, contrary to our expectation, the deeper the reset (lower I_{READ}), the lower is the measured V_{DIST} value [51]. This is hard to logically interpret; however, it turns out that this is a unique feature of the QPC model as illustrated by Figure 21.11 [52]. When the filament is not ruptured and the reset gets deeper (implying less number of defects in the constriction), the overall voltage applied across the dielectric redistributes itself to be localized at the filament-constriction (F–C) interfaces only. As a result, with decreasing N_C , the F–C interface width gets

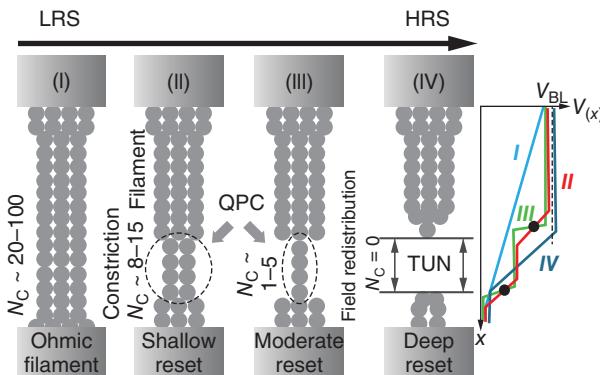


Figure 21.11 The four possible scenarios for the shape and size of the conductive filament (CF) ranging from (I) LRS Ohmic filament to (II) HRS shallow reset, (III) HRS moderate reset with very few defects in the “constriction” to (IV) HRS deep reset, when a tunneling (TUN) barrier is created. The probability to end up in the states (II, III, and IV) depends on the dielectric material

parameters and vacancy transport properties. The plot at the right end shows the potential drop profile for each of these four scenarios. The HRS states in (II and III) correspond to QPC mode of conduction with majority of the voltage dropping across the two F–C interfaces. (Reprinted from Ref. [52], Copyright 2013 The Japan Society of Applied Physics.)

sharper and the potential drop (electric field) is more concentrated and locally enhanced. The interesting feature of the QPC is that the field is almost zero anywhere outside the F–C interface and therefore, the immunity to RTN is dependent on the magnitude of the local field at the F–C interface. The deeper the reset (without filament rupture), the higher the local QPC field and, therefore, the lower the V_{DIST} value as observed in Ref. [51]. From a quantum physics perspective, the potential is concentrated at the F–C interface due to interference of the incident and reflected electron wavefunctions there [53]. Considering the dependence of V_{DIST} on the ramp rate and using the thermochemical model for defect generation (creating an oxygen vacancy requires bond breakage of Hf–O bonds), it can be estimated that the “time to disturb” (at $V_{READ} = 0.1$ V) can be as low as a few milliseconds to as high as a few megaseconds depending on the extent of reset within the QPC regime [52]. Therefore, although deep reset is desired for higher memory window, if the filament does not rupture, then the system will be highly prone to vacancy-induced RTN effects. When operating within the QPC regime, there is always a trade-off involved in the depth of reset and the RTN immunity.

21.5.3.3 Vacancy-Induced RTN for Very Deep Reset

If the filament is able to undergo rupture during the reset process thereby introducing a tunnel barrier in the dielectric for HRS, the disturb trends are completely reversed. As shown in Figure 21.12 [54], which shows the $V_{DIST} - I_{READ}$ trend for the same device where deep reset was observed for a few cases when I_{comp} is as low as $0.3 \mu\text{A}$, the value of V_{DIST} starts to increase again with deeper reset, against

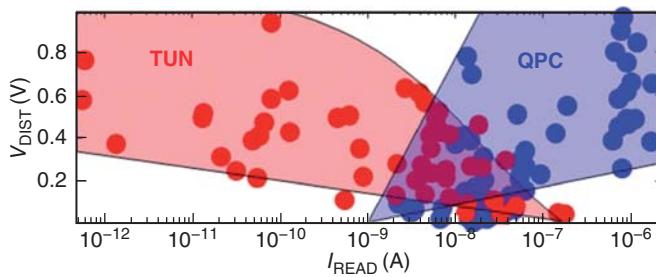


Figure 21.12 Dependence of V_{DIST} on the reset level (I_{READ}) for a very wide range of reset ranging from $I_{\text{READ}} = 10^{-12}$ to 10^{-6} A corresponding to different degrees of soft breakdown (SBD) and progressive

breakdown (PBD). The harder the breakdown, the larger is the filament and the lower the chance for rupture (filament remains in QPC mode). (© 2013 IEEE. Reprinted, with permission, from Ref. [54].)

the hypothesis of QPC. This is precisely because the QPC ceases to hold true in this regime and we have basically entered the tunnel (TUN) regime with a dielectric barrier across which the potential drops uniformly. Deeper reset corresponds to thicker tunnel barrier and therefore prolonged disturb time (enhanced RTN immunity). Whether a particular RESET state is in the QPC or TUN regime can be verified by analyzing the $\{\omega_X, \omega_Y, N_c\}$ values. When force-fitting a deep reset $I-V$ curve to the QPC model, we end up with an unrealistically high value of N_c that contradicts the QPC assumption [54]. We, therefore, have a robust methodology in place to distinguish between QPC and TUN configurations, according to which the immunity to vacancy-induced RTN is determined.

21.5.3.4 Bimodal Filament Configuration and Disturb Immunity

Considering that RESET is a purely stochastic process that involves recombination of many pairs of oxygen ions and vacancies, we should expect to see a bimodality in the filament configuration that can have a finite nonzero probability of ending up in the QPC or TUN modes. This hypothesis is well confirmed by our analysis of switching for many cycles at different compliance values [54]. Although lower compliance forming and SET enhance the probability of filament rupture (soft breakdown regime), there still exists a finite chance of staying in the QPC regime for the HRS. The vice versa holds true for the high compliance case (progressive breakdown regime – see Chapter 8 for more details on different modes of breakdown). Note that the LRS state is not analyzed here because it is more resilient to vacancy perturbations given the uniform potential drop across the whole dielectric for a large size filament (implying low electric field). With the filament configuration being bimodal, we can conclude that the V_{DIST} distribution should also be bimodal.

21.5.3.5 Role of Dielectric Microstructure on RTN Immunity

When the role of the dielectric microstructure is considered, where the dielectric can be amorphous or polycrystalline with grain boundaries (GBs), the presence

of GB causes a reduction in the V_{DIST} value probably because the GB serves as an easy diffusion path for vacancies to migrate along [55]. Therefore, although the presence of GB may help reduce the forming power, control the variability in the filament size and shape [54], and reduce the bimodality in the filament configuration, it suffers from lower V_{DIST} and has a shallower reset [54] (small memory window) both of which are undesirable.

21.5.4

Summary of RTN Analysis Studies

In summary, we have analyzed the two key mechanisms of RTN in RRAM and identified them to be electron transport-based (steady-state) and vacancy-perturbation-based (non-steady-state). While the former serves as a good defect spectroscopy tool, the latter is the more critical factor that can disturb the stability of the HRS state. Using the QPC formulation, the vacancy-induced RTN phenomena was studied in depth for various compliance levels and dielectric material/microstructure. To achieve good switching with robust disturb immunity, it is desirable to have a low LRS state and very deep HRS state (TUN regime). However, it is hard to find dielectric materials that satisfy both these criteria at the same time. From a material design perspective, the field acceleration factor γ (as defined in studies of electrical breakdown [56]) of the dielectric plays an important role. A high value of γ can simultaneously ensure low forming/SET voltage and a high V_{DIST} value. This factor depends on the relative permittivity and the permanent dipole moment of the high- κ material. Further studies on RTN are essential considering that future devices are being downscaled to areas as low as $10 \times 10 \text{ nm}^2$ [22], where the background noise is relatively low and the RTN effects are expected to be more dominant.

21.6

Disturb

21.6.1

Phenomena

To complete our overview of reliability aspects, this section discusses “disturb”. A (read) disturb error is an (undesired) change of the programmed state of a memory cell by reading that cell for extended times. More in general, a change of the content of the cell can occur during different memory operation conditions, that is, not only by multiple reads of the specified bits, but also because of multiple writes or (less likely) a read of other (neighboring) cells in the same memory array. In the latter case, a voltage stress on a nonselected bit may result from directly applied voltages (e.g., on half selected bit or word lines) and/or capacitive coupling phenomena.

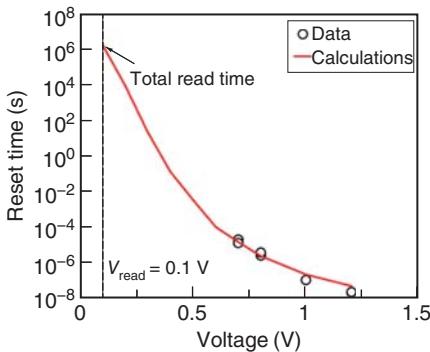


Figure 21.13 Measured and calculated pulse width required for reset (reset time) as function of pulse amplitude (voltage). (© 2012 IEEE. Reprinted, with permission, from Ref. [59].)

The occurrence of neighboring cell read-and-write disturbs is strongly dependent on the details of the memory array organization and design. It is clear, however, that the weaker is the isolation between different cells, the more susceptible the array will be to these disturb effects. In particular, they constitute severe design limitations for raw crossbar RRAM arrays, see for example [57].

A general study of disturb susceptibility can be done by applying different voltage stress conditions on a memory cell (with voltage amplitudes that are lower than those required to directly switch the cell).

21.6.2

Understanding and Modeling

The intrinsic susceptibility to switching due to low voltage stress can be understood from the program voltage versus program pulse width characteristics of an RRAM cell. While measuring this relationship for short pulse widths (spanning the range of normal program pulse widths) indicates a strong nonlinearity approaching a kind of threshold voltage for longer pulse times for both SET and RESET switching [58, 59], true saturation is not evidenced and switching at lower voltages is still expected, albeit for pulse widths increasing in a superlinear way.

If the effect of multiple pulses is the same as of that of one single pulse with a pulse width equal to the sum of the pulse widths, disturb effects can be predicted out of the extrapolation of the measured program voltage versus pulse width behavior (see Figure 21.13 [59]). This assumption is correct only if there are no important transients (e.g., of the internal temperature) in the cell during even a single pulse, so that “equilibrium” is reached “instantly” – which may require further extensive studies to possibly validate.

21.6.3

Anomalous Disturb Behavior

For a bipolar switching cell, one would intuitively only expect a possible disturb-caused switching from the cell HRS (RESET) state to the cell LRS (SET)

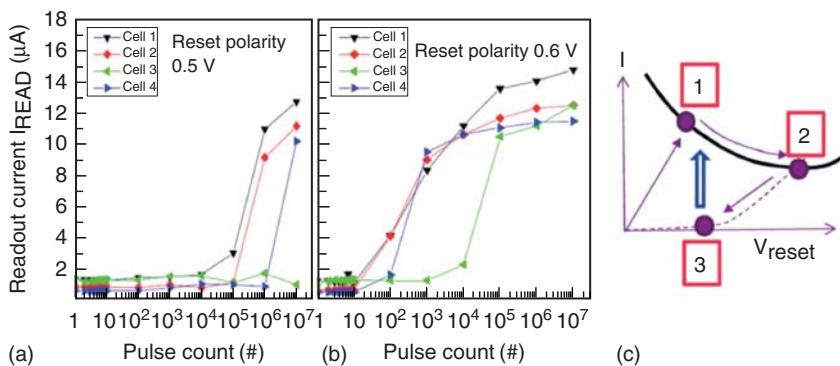


Figure 21.14 Disturb measurements (100 ns pulses) at (a) 0.5 and (b) 0.6 V in reset polarity. (c) Schematic explanation: (1) transition voltage is reached. (1 → 2) Reset along dynamic balance line. (2 → 3) Fast down

ramp to the OFF state. (3) System tends to return to the ON state (upwards), but because of the low voltage, a time delay is observed (a,b). (© 2012 IEEE. Reprinted, with permission, from Ref. [36].)

state for voltage stress having the same polarity as the polarity required for SET programming, and, similarly, an LRS to HRS disturb switching only for voltages with the RESET polarity. However, experimentally a decrease of the cell resistance has been observed when applying low voltage pulses with RESET polarity, see Figure 21.14 [36]. This effect has been understood by that during the initial (fast) RESET, the cell is not in equilibrium as the cell resistance tends to a “balance value” that is dependent on the applied voltage.

So, this effect should rather be interpreted as a resistance retention issue with the cell resistance moving to an equilibrium state, and the voltage pulses giving the system the required energy to evolve to that state.

21.7 Conclusions and Outlook

This chapter discussed the different reliability aspects of bipolar switching TMO RRAM cells. It is shown that insight into the filamentary switching processes on the atomic scale is key to understand the reliability physics. Based on that understanding, material, process, and operation conditions may be further tuned to improve the RRAM memory performance toward the targeted application specifications. While different reliability figures (as endurance) have indeed considerably improved, low current operation retention and significant cycle-to-cycle variability are identified as major concerns for reliable operation of scaled RRAM memories.

Furthermore, we should remain aware that, as the RRAM technology is not yet at the stage of making large density memory arrays in advanced scaled technology, we have today only limited experimental data – mostly on individual cells and/or

small arrays. This means that available statistics are still limited, and while we have an understanding on the scale of the “main” bit distributions, we are far from exploring the behavior of ppm and lower tail bits that may eventually decide on the application of the technology in a real high density memory. This should be further addressed in future reliability studies.

Acknowledgments

DW acknowledges support by the Deutsche Forschungsgemeinschaft through the SFB 917.

References

1. Kim, Y.-B., Lee, S.R., Lee, D., Lee, C.B., Chang, B., Hur J.H., Lee, M.-J., Park, G.-S., Kim, C.J., Chung, U.-I., Yoo I.-K., and Kim K. (2011) Bi-layered RRAM with unlimited endurance and extremely uniform switching. *IEEE 2011 VLSI Technology Symposium (VLSI) Technical Digest*, pp. 52–53.
2. Lee, M.-J., Lee, C.B., Lee, D., Lee, S.R., Chang, B., Hur, J.H., Kim, Y.-B., Kim, C.-J., Seo, D.H., Seo, S., Chung, U.-I., Yoo, I.-K., and Kim, K. (2011) A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta₂O₅–x/TaO₂–x bilayer structures. *Nat. Mater.*, **10**, 625–630.
3. Lee, H.Y., Chen, P.S., Wu, T.Y., Chen, Y.S., Wang, C.C., Tzeng, P.J., Lin, C.H., Chen, F., Lien, C.H., and Tsai, M.-J. (2008) Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO₂ based RRAM. *IEEE 2008 International Electron Device Symposium (IEDM) Technical Digest*, pp. 297–300.
4. Lee, H.Y., Chen, Y.S., Chen, P.S., Gu, P.Y., Hsu, Y.Y., Wang, S. M., Liu, W.H., Tsai, C.H., Sheu, S.S., Chiang, P.C., Lin, W.P., Lin, C.H., Chen, W.S., Chen, F.T., Lien, C.H., and Tsai, M.-J.. (2010) Evidence and solution of over-RESET problem for HfO_X based resistive memory with sub-nanosecond switching speed and high endurance. *IEEE 2010 International Electron Device Symposium (IEDM) Technical Digest*, pp. 460–463.
5. Chen, Y.Y., Govoreanu, B., Goux, L., Degraeve, R., Fantini, A., Kar, G.S., Wouters, D.J., Groeseneken, G., Kittl, J.A., Jurczak, M., and Altimime, L. (2012) Balancing SET/RESET pulse for $>10^{10}$ endurance in HfO₂/Hf 1T1R bipolar RRAM. *IEEE Trans. Electron Devices*, **59** (12), 3243–3249.
6. Chen, Y.Y., Goux, L., Clima, S., Govoreanu, B., Degraeve, R., Kar, G.S., Fantini, A., Groeseneken, G., Wouters, D.J., and Jurczak, M. (2013) Endurance/retention trade-off on HfO₂/metal cap 1T1R bipolar RRAM. *IEEE Trans. Electron Devices*, **60** (3), 1114–1121.
7. Yang, J.J., Zhang, M.-X., Strachan, J.P., Miao, F., Pickett, M.D., Kelley, R.D., Medeiros-Ribeiro, G., and Williams, R.S. (2010) High switching endurance in TaO_x memristive devices. *Appl. Phys. Lett.*, **97**, 232102.
8. Wei, Z., Kanzawa, Y., Arita, K., Katoh, Y., Kawai, K., Muraoka, S., Mitani, S., Fujii, S., Katayama, K., Iijima, M., Mikawa, T., Ninomiya, T., Miyanaga, R., Kawashima, Y., Tsuji, K., Himeno, A., Okada, T., Azuma, R., Shimakawa, K., Sugaya, H., Takagi, T., Yasuhara, R., Horiba, K., Kumigashira, H., and Oshima, M. (2008) Highly reliable TaO_x ReRAM and direct evidence of redox reaction mechanism. *IEEE 2008 International Electron Device Symposium (IEDM) Technical Digest*, pp. 293–296.
9. Kawahara, A., Azuma, R., Ikeda, Y., Kawai, K., Katoh, Y., Tanabe, K.,

- Nakamura, T., Sumimoto, Y., Yamada, N., Nakai, N., Sakamoto, S., Hayakawa, Y., Tsuji, K., Yoneda, S., Himeno, A., Origasa K-I, Shimakawa, K., Takagi, T., Mikawa, T., and Aono, K. (2012) An 8Mb multi-layered cross-point ReRAM macro with 443MB/s write throughput. *IEEE 2012 International Solid-State Circuits Conference (ISSCC) Technical Digest*, pp. 432–434.
10. Koveshnikov, S., Matthews, K., Min, K., Gilmer, D.C., Sung, M.G., Deora, S., Li, H.F., Gausepohl, S., Kirsch, P.D., and Jammy, R. (2012) Real-time study of switching kinetics in integrated 1T/HfO_x 1R RRAM: intrinsic tunability of set/reset voltage and trade-off with switching time. *IEEE 2013 International Electron Device Symposium (IEDM) Technical Digest*, pp. 486–488.
11. Chen, B., Lu, Y., Gao, B., Fu, Y.H., Zhang, F.F., Huang, P., Chen, Y.S., Liu, L.F., Liu, X.Y., Kang, J.F., Wang, Y.Y., Fang, Z., Yu, H.Y., Li, X., Wang, X.P., Singh, N., Lo, G.Q., and Kwong, D.L. (2011) Physical mechanisms of endurance degradation in TMO-RRAM. *IEEE 2011 International Electron Device Symposium (IEDM) Technical Digest*, pp.283–286.
12. Chen, Y.Y., Degraeve, R., Clima, S., Govoreanu, B., Goux, L., Fantini, A., Kar, G.S., Pourtois G., Groeseneken, G., Wouters, D., and Jurczak, M. (2012) Understanding of the endurance failure in scaled HfO₂-based 1T1R RRAM through vacancy mobility degradation. *IEEE 2012 International Electron Device Symposium (IEDM) Technical Digest*, pp. 482–485.
13. Yu, S., Guan, X., and Wong, H.-S.P. (2012) Understanding metal oxide RRAM current overshoot and reliability using Kinetic Monte Carlo simulation. *IEEE 2012 International Electron Device Symposium (IEDM) Technical Digest*, pp. 585 – 588.
14. Wei, Z., Takagi, T., Kanzawa, Y., Katoh, Y., Ninomiya, T., Kawai, K., Muraoka, S., Mitani, S., Katayama, K., Fujii, S., Miyanaga, R., Kawashima, Y., Mikawa, T., Shimakawa, K., and Aono, K. (2011) Demonstration of high-density ReRAM ensuring 10-year retention at 85°C based on a newly developed reliability model. *IEEE 2011 International Electron Device Symposium (IEDM) Technical Digest*, pp. 721-724.
15. Wei, Z., Takagi, T., Kanzawa, Y., Katoh, Y., Ninomiya, T., Kawai, K., Muraoka, S., Mitani, S., Katayama, K., Fujii, S., Miyanaga, R., Kawashima, Y., Mikawa, T., Shimakawa, K., and Aono, K. (2012) Retention model for high-density ReRAM. *Proceedings EEE 2012 International Memory Workshop (IMW)*, pp. 14–17.
16. Ielmini, D., Nardi, F., Cagli, C., and Lacaita, A.L. (2011) Size-dependent retention time in NiO-based resistive switching memories. *IEEE Electron Device Lett.*, **31**, 353–355.
17. Yu, S., Chen, Y.Y., Guan, X., Wong, H.-S.P., and Kittl, J.A. (2012) A Monte Carlo study of the low resistance state retention of HfO_x based resistive switching memory. *Appl. Phys. Lett.*, **100**, 043507.
18. Wang, Y.-L., Song, Y.-L., Yang, L.-M., Lin, Y.-Y., Huang, R., Zou, Q.-T., and Wu, J.-G. (2012) Algorithm-enhanced retention based on megabit array of Cu_xSi_yO RRAM. *IEEE Electron Device Lett.*, **33** (10), 1408–1410.
19. Ninomiya, T., Wei, Z., Muraoka, S., Yasuhara, R., Katayama, K., and Takagi, T. (2013) Conductive filament scaling of TaO_x bipolar ReRAM for improving data retention under low operation current. *IEEE Trans. Electron Devices*, **60** (4), 1384–1389.
20. Baek, I.G., Lee, M.S., Seo, S., Lee, M.J., Seo, D.H., suh, D.-S., Park, J.C., Park, S.O., Kim, H.S., Yoo, I.K., Chung, U.-I., and Moon, J.T. (2004) Highly scalable nonvolatile memory using simple binary oxide driven by asymmetric pulses. *IEEE 2004 International Electron Device Symposium (IEDM) Technical Digest*, pp. 587 – 590.
21. Govoreanu, B., Kar, G.S., Chen, Y.-Y., Parashiv, V., Kubicek, S., Fantini A., Radu, I.P., Goux, L., Clima, S., Degraeve, R., Jossart, N., Richard, O., Vandeweyer, T., Seo, K., Hendrickx, P., Pourtois, G., Bender, H., Altimime, L., Wouters, D.J., Kittl, J.A., and Jurczak, M. (2011) 10x10nm² Hf/HfO crossbar resistive

- RAM with excellent performance, reliability and low-energy operation. IEEE 2011 International Electron Device Symposium (IEDM) Technical Digest, p. 792.
22. Vandelli, L., Padovani, A., Larcher, L., Bersuker, G., Gilmer, D., and Pavan, P. (2011) Modeling of the forming operation in HfO₂-based resistive switching memories. Proceedings of IEEE 2011 International Memory Workshop (IMW), pp. 1–4.
 23. Raghavan, N., Fantini, A., Degraeve, R., Roussel, P.J., Goux, L., Govoreanu, B., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) Statistical insight into controlled forming and forming free stacks for HfO_x RRAM. *Microelectron. Eng.*, **109**, 177–181.
 24. Degraeve, R., Goux, L., Clima, S., Govoreanu, B., Chen, Y.Y., Kar, G.S., Roussel, P.J., Pourtois, G., Wouters, D.J., Altimime, L., Jurczak, M., Groeseneken, G., and Kittl, J.A. (2012) Modeling and tuning the filament properties in RRAM metal oxide stacks for optimized stable cycling. Proceedings 2012 International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA), pp. 1–2.
 25. Choi, B.J., Torrezan, A.C., Norris, K.J., Miao, F., Strachan, J.P., and Zhang, M.X. (2013) Electrical performance and scalability of Pt dispersed SiO₂ nanometallic resistance switch. *Nano Lett.*, **13** (7), 3213–3217.
 26. Liang, J., Yeh, S., Wong, S., and Wong, H.-S.P. (2012) Scaling challenges for the cross-point resistive memory array to sub-10nm node—an interconnect perspective. Proceedings IEEE 2012 International Memory Workshop (IMW), pp. 61–64.
 27. Zhang, L., Cosemans, S., Wouters D.J., Govoreanu, B., Groeseneken, G., and Jurczak, M. (2013) Analysis of vertical cross-point resistive memory (VRRAM) for 3D RRAM design. Proceedings IEEE 2013 International Memory Workshop (IMW), pp. 155–158.
 28. Chae, S.C., Lee, J.S., Kim, S., Lee, S.B., Chang, S.H., Liu, C., Kahng, B., Shin, H., Kim, D.-W., Jung, C.U., Seo, S., Lee, M.-J., and Noh, T.W. (2008) Random circuit breaker network model for unipolar resistance switching. *Adv. Mater.*, **20** (6), 1154–1159.
 29. Lee, S.B., Lee, J.S., Chang, S.H., Yoo, H.K., Kang, B.S., Kahng, B., Lee, M.-J., Kim, C.J., and Noh, T.W. (2011) Interface-modified random circuit breaker network model applicable to both bipolar and unipolar resistance switching. *Appl. Phys. Lett.*, **98**, 033502.
 30. Fantini, A., Goux, L., Degraeve, R., Wouters, D.J., Raghavan, N., Kar, G., Belmonte, A., Chen, Y.Y., Govoreanu, B., and Jurczak, M. (2012) Intrinsic switching variability in HfO₂ RRAM. Proceedings IEEE 2012 International Memory Workshop (IMW), pp. 45–48.
 31. Yu, S., Guan, X., and Wong, H.-S.P. (2012) On the switching parameter variation of metal oxide RRAM-Part II: model corroboration and device design strategy. *IEEE Trans. Electron Devices*, **59** (4), 1183–1188.
 32. Bersuker, G., Gilmer, D.C., Veksler, D., Kirsch, P., Vandelli, L., Padovani, A., Larcher, L., McKenna, K., Shluger, A., Iglesias, V., Porti, M., and Nafria, M. (2011) Metal oxide resistive memory switching mechanism based on conductive filament properties. *J. Appl. Phys.*, **110** (12), 2011.
 33. Degraeve R., Roussel P., Goux L., Wouters D.J., Kittl J , Altimime L., Jurczak M., and Groeseneken, G. (2012) Generic learning of TDDB applied to RRAM for improved understanding of conduction and switching mechanism through multiple filaments. IEEE 2010 International Electron Device Symposium (IEDM) Technical Digest, pp. 632–635.
 34. Balatti, S., Ambrogio, S., Gilmer, D.C., and Ielmini, D. (2013) Set variability and failure induced by complementary switching in bipolar RRAM. *IEEE Electron Device Lett.*, **34** (7), 861–863.
 35. Ielmini, D. (2011) Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Devices*, **58** (12), 4309–4317.
 36. Degraeve, R., Fantini, A., Clima, S., Govoreanu, B., Goux, L., Chen, Y.Y.,

- Wouters, D.J., Roussel, P., Kar, G.S., Pourtois, G., Cosemans, S., Kittl, J.A., Groeseneken, G., Jurczak, M., and Altimime, L. (2012) Dynamic 'hour glass' model for SET and RESET in HfO_x RRAM. IEEE 2012 VLSI Technology Symposium (VLSI) Technical Digest, pp. 75–76.
37. Chen, Y.S., Lee, H.Y., Chen, P.S., Gu, P.Y., Chen, C.W., Lin, W.P., Liu, W.H., Hsu, Y.Y., Sheu, S.S., Chiang, P.C., Chen, W.S., Chen, F.T., Lien, C.H., and Tsai, M.J. (2009) Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity. IEEE 2009 International Electron Device Symposium (IEDM) Technical Digest, pp. 105–108.
38. Terai, M., Sakotsubo, Y., Saito, Y., Kotsuji, S., and Hada, H. (2010) Memory-state dependence of random telegraph noise of Ta₂O₅/TiO₂ stack ReRAM. *IEEE Electron Device Lett.*, **31** (11), 1302–1304.
39. Lee, D., Lee, J., Jo, M., Park, J., Siddik, M., and Hwang, H. (2011) Noise-analysis-based model of filamentary switching ReRAM with ZrO_x/HfO_x stacks. *IEEE Electron Device Lett.*, **32** (7), 964–966.
40. Lee, J.K., Lee, J.W., Park, J., Chung, S.W., Roh, J.S., Hong, S.J., Cho, I.W., Kwan, H.I., and Lee, J.H. (2011) Extraction of trap location and energy from random telegraph noise in amorphous TiO_x resistance random access memories. *Appl. Phys. Lett.*, **98**, 143502.
41. Tseng, Y.H., Shen, W.C., Huang, C.E., Lin, C.J., and King, Y.C. (2010) Electron trapping effect on the switching behavior of contact RRAM devices through random telegraph noise analysis. IEEE 2010 International Electron Device Symposium (IEDM) Technical Digest, pp. 28.5.1–28.5.4.
42. Raghavan, N., Degraeve, R., Fantini, A., Goux, L., Strangio, S., Govoreanu, B., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability. IEEE 2013 International Reliability Physics Symposium (IRPS), Monterey, CA, pp. 5E.3.1–5E.3.7.
43. Soni, R., Meuffels, P., Petraru, A., Weides, M., Kügeler, C., Waser, R., and Kohlstedt, H. (2010) Probing Cu doped Ge_{0.3}Se_{0.7} based resistance switching memory devices with random telegraph noise. *J. Appl. Phys.*, **107**, 024517.
44. Puglisi, F.M., Pavan, P., Padovani, A., Larcher, L., and Bersuker, G. (2010) Random telegraph signal noise properties of HfO_x RRAM in high resistive state. Proceedings of the IEEE 2012 European Solid-State Device Research Conference (ESSDERC), pp. 274–277.
45. Veksler, D., Bersuker, G., Chakrabarti, B., Vogel, E., Deora, S., Matthews, K., Gilmer, D.C., Li, H.F., Gausepohl, S., and Kirsch, P.D. (2012) Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics. IEEE 2012 International Electron Device Symposium (IEDM) Technical Digest, pp. 219–222.
46. Ielmini, D., Nardi, F., and Cagli, C. (2010) Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories. *Appl. Phys. Lett.*, **96**, 053503.
47. Lee, J.K., Jeong, H.Y., Cho, I.T., Lee, J.Y., Choi, S.Y., Kwon, H.I., and Lee, J.H. (2010) Conduction and low-frequency noise analysis in Al / TiO_x / Al bipolar switching resistance random access memory devices. *IEEE Electron Device Lett.*, **31** (6), 603–605.
48. Miranda, E. and Suñé, J. (2001) Analytic modeling of leakage current through multiple breakdown paths in SiO₂ films. Proceedings IEEE 2001 International Reliability Physics Symposium (IRPS), pp. 367–379.
49. Cester, A., Bandiera, L., Suñé, J., Boschiéro, L., Ghidini, G., and Paccagnella, A. (2001) A novel approach to quantum point contact for post soft breakdown conduction. IEEE 2001 International Electron Device Symposium (IEDM) Technical Digest, pp. 305–308.
50. Lian, X., Long, S., Cagli, C., Buckley, J., Miranda, E., Liu, M., and Suñé, J. (2012) Quantum point contact model of filamentary conduction in resistive switching memories. Proceedings 13th

- 2012 International Conference on Ultimate Integration on Silicon (ULIS), pp. 101–104.
51. Raghavan, N., Degraeve, R., Fantini, A., Goux, L., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) Modeling the impact of reset depth on vacancy induced filament perturbations in HfO_2 RRAM. *IEEE Electron Device Lett.*, **34** (5), 614–616.
 52. Raghavan, N., Degraeve, R., Goux, L., Fantini, A., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) RTN insight to filamentary instability and disturb immunity in ultra-low power switching HfO_x and AlO_x RRAM. IEEE 2013 VLSI Technology Symposium (VLSI) Technical Digest, pp. T163–T164.
 53. Ulreich, S. and Zwerger, W. (1998) Where is the potential drop in a quantum point contact? *Superlattices Microstruct.*, **23** (3-4), 719–730.
 54. Raghavan, N., Degraeve, R., Fantini, A., Goux, L., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) Stochastic variability of vacancy filament configuration in ultra-thin dielectric RRAM and its impact on OFF-state reliability. IEEE 2013 International Electron Device Symposium (IEDM) Technical Digest, pp. 554–557.
 55. McKenna, K. and Shluger, A. (2009) The interaction of oxygen vacancies with grain boundaries in monoclinic HfO_2 . *Appl. Phys. Lett.*, **95**, 222111.
 56. McPherson, J.W. and Mogul, H.C. (1998) Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films. *J. Appl. Phys.*, **84** (3), 1513–1523.
 57. Chen, A. (2011) Accessibility of nano-crossbar arrays of resistive switching devices. Nanotechnology (IEEE-NANO), August 2011, pp. 1767–1771.
 58. Schindler, C., Staikov, G., and Waser, R. (2009) Electrode kinetics of $\text{Cu}-\text{SiO}_2$ -based resistive switching cells: overcoming the voltage-time dilemma of electrochemical metallization memories. *Appl. Phys. Lett.*, **94**, 072109.
 59. Larentis, S., Nardi, F., Balatti, S., Gilmer, D.C., and Ielmini, D. (2012) Resistive switching by voltage-driven ion migration in bipolar RRAM—Part II: modeling. *IEEE Trans. Electron Devices*, **59** (9), 2468–2475.

22

Select Device Concepts for Crossbar Arrays

Geoffrey W. Burr, Rohit S. Shenoy, and Hyunsang Hwang

22.1

Introduction

Crossbar arrays have been investigated for memory applications for more than 60 years [1]. Such a memory ought to be capable of extremely high densities, since both the wires and the spaces between them are of width F , and thus the area per connection is only $4F^2$. However, a crossbar memory application requires a memory device with a particularly diverse set of critical characteristics.

Such a crossbar memory device must be a two-terminal device that can be reliably, repeatedly, and readily switched between at least two resistance states, preferably with a large resistance contrast. This switching operation must not require excessive power, and the device should be capable of surviving through many millions, if not billions or more, of switching cycles.

Each read operation, at lower voltages and currents than the switching operation, should supply sufficient current for distinguishing the various states. Yet, even a very large number of successive read operations must not induce a switching event. The resistance states should be nonvolatile over the normal operation range of the system (typically up to 85 °C for nonmilitary applications) over a long lifetime (often 10 years is specified). Error rates should be both predictable and low enough throughout the device lifetime for correction by error-correction coding (ECC) with a reasonable amount of redundancy overhead.

In the past 10–15 years, memory devices capable of these kinds of specifications have moved closer to becoming a reality. Research begun in the late 1990s, originally motivated by the desire to have a “backup device” in case either NAND or NOR Flash had trouble scaling to smaller dimensions, has produced a number of viable two-terminal nonvolatile memory (NVM) devices. These include ferroelectric RAM (FeRAM) [2], phase-change memory (PCM) [3], spin-torque-transfer magnetic RAM (STT-MRAM) [4], and the subject of the book you are holding: Resistance RAM (ReRAM). Even though NAND Flash succeeded at scaling and, as of 2014, has not required “replacement,” there has been a general realization that many of these two-terminal NVM devices might offer better performance through much lower latencies and much higher program-erase endurance than

NAND Flash. This opportunity is collectively referred to as *Storage Class Memory* [5, 6].

In many of the other chapters in this book, you can find consideration of the physics and engineering of resistive switching materials, and discussions of NVM device-specific attributes such as switching power, switching speed, resistive window, cycling endurance, retention across a range of temperatures, reliability, and the scalability of the size of the device. Other chapters address applications that extend beyond the simple memory application of ReRAM to new forms of memory-intensive computation, some motivated by the neural architecture of the brain.

The focus of this chapter is on the leap from a single, “state-holding” memory device to a large array of memory devices. Ideally, we would have a vast matrix of densely packed devices, within which we could write and read any small subset of the memory devices at will, while all other devices would remain completely unperturbed, dissipating zero additional power beyond that required for writing or reading. Although this ideal situation is, as one might expect, not really attainable, one can come close by introducing a strongly nonlinear I - V (current–voltage) characteristic – a *select device* – together with the ReRAM or NVM element at every crossbar node.

This nonlinearity can be introduced by a second, discrete two-terminal device – such as a diode, switch, or other similar element – in series with each state-holding (ReRAM) element. Here, the memory element and the select device can each be optimized separately, at the cost of additional processing steps. Alternatively, the necessary nonlinearity can be simply added to the list of necessary criteria that the prospective state-holding element must deliver. Once successfully developed, there is only one device to integrate; however, this list of criteria is quite daunting already, without adding a strong nonlinearity of many orders of magnitude – at exactly the “right” voltage – to the list.

In this chapter, we begin with a brief overview of crossbar biasing schemes, and discuss the need to minimize leakage through all the various nonselected cells, while delivering the right voltages and currents to the selected cell. We then summarize the criteria that a select device must fulfill in order to enable crossbar memory. We review current research on various discrete select device options, ranging from conventional silicon-based semiconductor devices, to oxide semiconductors, threshold switch devices, oxide tunnel barriers, and devices based on mixed-ionic-electronic-conduction (MIEC). Various approaches for self-selected ReRAM are discussed, and then the chapter concludes.

22.2

Crossbar Array Considerations

The task for a crossbar array is fairly simple: we need to apply a set of voltages at the edge of the array such that the desired operations (reads and writes) take place at the desired *selected cells*, yet all unselected cells remain unperturbed and

overall power dissipation remains manageable [7]. Over the past few years, a large number of crossbar analyses have been performed [8–21], which all show that it is nearly impossible to implement a crossbar array of any practical size without a strong nonlinearity at each node – we will not repeat such analyses here.

The overall memory chip, storing anywhere from megabytes to terabytes of memory, will be composed of numerous crossbar subarrays, which might each be of the size 1000×1000 wires. We want each of these subarrays to be as large as possible, in order to amortize the peripheral circuitry we must place at the edge and to maximize the *area efficiency*: the fraction of silicon used for memory bits.

22.2.1

Problems Associated with Large Subarrays

This strong incentive to make subarrays as large as possible must be traded off against problems associated with large subarrays. The longer the wire, the larger the capacitance and the larger the worst-case resistance drop. These problems are exacerbated at advanced technology nodes, as wires get both thinner and are spaced more closely together [22]. The worst-case access is a write to the “far” corner of the array, and the need to boost the voltage at the far edge of the array runs the risk of perturbing the state of the unselected devices that sit right at the near edge.

The large capacitance of long bit lines is a well-known issue in memory design. An added complication for ReRAM is the critical importance of *compliance* currents. When large bit line capacitances are present, a transistor or other circuit at the edge of the array can no longer instantaneously limit the current passing through a cell within the array. Failure to terminate ReRAM filament formation at a particular targeted current can lead to excessively thick filaments, which then require significant switching power to sever or remove during the subsequent “RESET” operation.

22.2.2

Considerations During NVM-Write

When a device is selected for writing or reading, we typically energize one of the two wires leading to this device to some higher voltage, and the other one to a lower voltage. The biasing scheme specifies the intermediate voltages that are applied to all the other wires within the array, so as to minimize the overall leakage and to prevent any undesired disturbance to previously stored data. As shown in Figure 22.1, these other devices in the array fall into three classes: the *half-selected* devices along the same row as a selected device, the *half-selected* devices along the same column as a selected device, and the *unselected* devices that share neither a row nor a column with any selected device.

We first address the scenario of writing the worst-case device, since this is the largest applied voltage differential that will ever be used. One biasing scheme that

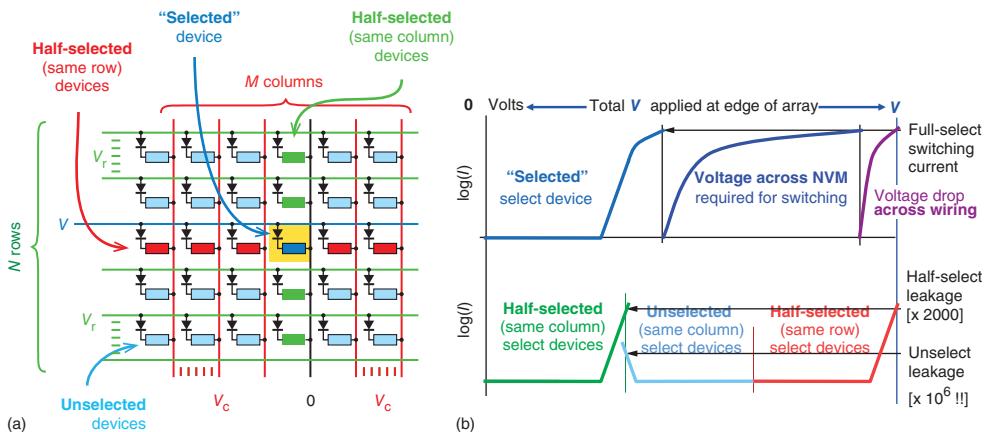


Figure 22.1 (a) In a crossbar array of ReRAM devices, voltage applied at the edge of the array, to select one or more ReRAM/select-device pairs, also affects three other sets of ReRAM + select devices: those in the same row, in the same column (both *half-selected*), and all others (*unselected*). (b) The total voltage drop across any selected ReRAM device must be sure to switch the

device, despite the additional voltage drop in its own select device and in the wiring. However, that same applied voltage must not lead to excessive leakage in either the half-selected or unselected devices. In a 1000×1000 array, for example, there will be ~ 2000 half-selected devices and nearly 1 million unselected devices.

works well for this scenario is referred to as the “V/3” scheme [8–21]. In its simplest form, the applied voltage is split into three parts, one for each of the three classes of nonselected devices. Both sets of half-selected devices experience the same voltage polarity as the selected device, while the unselected device experiences the opposite voltage polarity. As shown in Figure 22.1, the sum of these three voltage drops is roughly equal to the effective voltage drop across the selected device. Note that the presence of significant line resistance causes the wire and local device voltages to vary continuously along the selected rows and columns.

Although the name “V/3” would suggest division into three equal voltages, this is rarely the optimal choice, because there are many more unselected than half-selected devices. In the case of a bipolar select device with symmetric characteristics, it makes sense to reduce the voltage across the many unselected devices until total aggregate leakage is minimized. This scenario is illustrated in Figure 22.1. In a scenario where the select device and the ReRAM device are unipolar (using only one polarity for all switching operations), it might be preferable to significantly increase the opposite polarity voltage across the unselected devices.

During write operations, it is critical to deliver the necessary power and current to the selected cell while avoiding two undesirable outcomes: dissipating too much aggregate power and chanceing even the remote possibility that any of the nonselected cells might accidentally switch. The probability of such a *write disturb* during any one write event must be extremely low. Any excess power dissipated

during write will reduce the effective write bandwidth. As the size of the subarray gets larger, the number of potentially leaky devices increases. The energy required to drive the wires to the desired voltages and activate the subarray increases as $(1/2)CV^2$, thus favoring low wire capacitance and low-voltage operation.

22.2.3

Considerations During NVM-Read

During read operations, both the currents and the associated device voltages are typically significantly lower. Since read currents are typically 10 \times smaller than write currents, line resistance can frequently be ignored during read operations. Many authors have studied the use of a “V/2” scheme, where the voltage applied to the wires leading to *unselected* devices is either brought to zero or, alternatively, these lines are allowed to float [12]. There are different tradeoffs associated with these choices. Driving these lines to the same voltage has the advantage that the leakage within the array is sure to be low; however, the act of driving these wires consumes some amount of power in the peripheral circuitry.

During write, we were concerned with perturbing stored data and with wasted power; during read, we are additionally concerned with being able to accurately sense the state of individual devices. Without strong nonlinearity at each node, there are many potential “sneak paths” from the higher potential selected wire to the lower potential selected wire. In fact, every path that leads through a series combination of three devices in the low-resistance state – one each from the pool of half-selected/same-row, unselected, and then half-selected/same-column devices – can contribute significant signal at the readout node. However, this is not representative of the state of the selected device [8–21]. Given the large number of possible connections, the *sensing margin* – the detectable difference at the edge of the array between the high- and low-resistance states (HRS and LRS) – collapses in the absence of a strong nonlinearity, even for a fairly modest-sized array. The presence of this nonlinearity cuts off these sneak paths, either by suppressing the current through all three legs or in some cases, simply by blocking the current through the reverse-biased *unselected* devices, making it possible to have reasonable sense margin for large arrays. Typically, writing with good power efficiency is a much more difficult task for an access device than simply cutting off the sneak paths during readout [23, 24].

22.3

Target Specifications for Select Devices

An ideal select device has a number of critical characteristics for application in high-density resistive crossbar memory arrays.

In each selected cell, the select device should be capable of supplying high currents as needed to program and erase the memory element that is in series with it. Delivering such high currents through the low cross-sectional area

corresponding to devices of $\sim 1F^2$ implies that the select device should support a **high ON-state current density** that is on the order of several Mega-ampere per square centimeter.

Since the unselected cells vastly outnumber the selected cells in typical crossbar memory arrays, the select device's **OFF-state leakage current needs to be as low as possible** so as to prevent the aggregate leakage through all the unselected cells from dominating the overall system power budget. Select device OFF-state leakage currents that are several orders of magnitude lower than the ON-state current allow the use of large arrays that are desirable to increase overall area efficiency. Satisfying both of these points (high ON-state current and low OFF-state leakage) simultaneously implies that the select device needs to have a highly nonlinear *I-V* curve.

Many ReRAM device candidates show better cycling reliability and lower power requirements if operated in bipolar mode (program and erase operations use opposing voltage polarities). The associated select device must also be capable of such **bidirectional operation**, offering low OFF-state leakage and high ON-state current density under both voltage polarities.

If, like most emerging resistive memory elements, the select device can be located above the wafer surface without consuming active Si wafer area, this frees up Si area for placement of some peripheral circuits underneath the crossbar memory array. Such **process compatibility with 3D multilayer stacking** leads to more efficient utilization of the overall chip area for maximizing memory density. Furthermore, if the select device can be fabricated in a back-end-of-the-line (BEOL)-compatible process, then this enables layering of multiple crossbar arrays on top of each other for lower effective cost per bit. BEOL-compatibility means that the maximum thermal budget during the select device fabrication process should not exceed $\sim 400^\circ\text{C}$, yet the select devices themselves must able to withstand up to $\sim 400^\circ\text{C}$.

Various resistive memory elements have different program/erase voltages and currents. It is important that the select device achieves **voltage compatibility with the memory element** with which it is paired. Select device turn-on voltages should have (or be tunable to) the appropriate values to ensure that the half-selected and unselected cells still have low leakage currents, while delivering the necessary voltage drop across, and current through, the selected memory element for read and write conditions. The need to consider this compatibility *after* specifying the memory element is why you will find no table summarizing the performance characteristics of access devices accompanying this chapter. Without specification of the particular memory element, such a table has little to no value [23, 24].

Since it is important that the select device properties should not limit the overall memory chip performance and reliability, **all other properties of the select device should be better than or equal to that of the memory element**. These properties – such as switching speed, cycling endurance, array yield, and variability – should be as good as or better than the corresponding properties of the memory element.

The set of requirements listed here have made it very challenging to find select devices that are fully suitable for high-density, 3D multilayer resistive crossbar memory. In the next section, we survey a number of device candidates that have been proposed as select devices and highlight some of their strengths and weaknesses.

22.4

Types of Select Devices

22.4.1

Si Based

In this section, we discuss three main types of Si-based select devices: vertical transistors, diodes, and bidirectional switches.

Transistors can drive large currents, exhibit very low OFF-state leakage, and can even be operated in bidirectional mode (with some complications due to the asymmetry introduced by the placement of the NVM). It is also relatively easy to adjust their turn-on threshold voltage. Moreover, the ability to use MOSFET gate voltage to constrain the saturation drain current offers an excellent way to enforce local current compliance during ReRAM forming and SET operations. Finally, Si transistors can be mass-produced with very high yield and reliability. As a result of all these advantages, Si transistors are widely used as select devices in single-layer memory arrays.

The main drawbacks of transistor-based select devices for resistive memories are relatively large cell size, complicated process, and incompatibility with 3D multilayer BEOL stacking concepts due to high processing thermal budget and usage of Si wafer surface area. While the latter two issues will likely be very difficult to solve, the use of vertical transistors (wherein current flows perpendicular to the Si wafer surface) offers a path to reducing the cell size. Wang *et al.* proposed a compact ($4F^2$ cell-size) FET select device based on vertical Si nanopillars with gate-all-around [25] and experimentally demonstrated successful operation of HfO_2 -based ReRAM stacked above these transistors.

Wang *et al.* showed switching of HfO_2 -based ReRAM stacked above vertical bipolar junction transistors (BJTs) built in a standard commercial foundry CMOS logic process [26], achieving higher current-driving capability than planar MOSFETs built in the same CMOS process. Vertical BJT select devices, capable of achieving very high drive currents ($300 \mu\text{A}$) and low base-emitter reverse bias leakage ($<0.1 \text{ pA}$), have been employed to drive PCM cells in a 45 nm generation 1 Gb chip [27].

Oh *et al.* reported a 512 Mbit PCM chip [28], which had vertical Si PN junction diode select devices made using selective epitaxial growth (SEG) of Si over highly doped N-type Si substrate, and ion implantation to create superficial P-doped regions. They achieved a $5.8F^2$ cell size at the 90 nm technology node. The reported I - V curves suggest very high ON-state current density ($>25 \text{ MA cm}^{-2}$)

and ON/OFF current ratio $\sim 10^8$. However, the area efficiency was somewhat degraded due to contacts made to the heavily N+ doped substrate once every eight cells to reduce series resistance by strapping with metal. Also, one needs to carefully consider the effect of vertical and lateral parasitic BJT currents in such structures, especially as the inter-cell spacing is decreased [28, 29]. There have been numerous studies of SEG optimization [30, 31] as well as alternative approaches based on solid-phase epitaxy [32] to build vertical Si PN diodes.

Unlike epitaxially grown Si diodes that need single-crystal Si as a growth template, polysilicon diodes can be built on non-Si films, allowing the possibility of having at least one crossbar memory array layer above the Si substrate for better area efficiency. As expected from their poorer crystalline quality, the electrical properties of polysilicon diodes are usually worse than their single-crystal Si counterparts. Sasago *et al.* showed polysilicon P-I-N diodes built above metal wiring [33] and reduced contact resistances to get high ON-state current density $\sim 8 \text{ MA cm}^{-2}$. The worst-case forward/reverse rectification ratio was $> \sim 8 \times 10^4$. The authors were able to drive PCM cells with $160 \mu\text{A}$ RESET current through the polysilicon diodes. While the maximum process temperature was not specified in this paper, it is unlikely that these devices could be made in a typical sub- 400°C Cu-BEOL process while maintaining good enough crystalline quality and dopant activation for such high current density. In order to use polysilicon diodes for 3D-stacked multilayer crossbar memory, one would need to either resort to high-temperature-compatible metallization schemes [34], or employ processes such as laser annealing [35, 36] for crystallization and activation of ion-implanted dopants in an effectively low thermal budget process.

Given their rectifying nature, PN diodes cannot be used as select devices for bipolar memory elements. However, two-terminal NPN diodes do show bidirectional turn-on. Such structures have been proposed as select devices in simulation [37, 38] and experimental studies [39–41]. The basic idea, as shown in Figure 22.2, is to set up an NPN doping profile that results in the formation of a potential barrier for electron transport from one side of the device to the other at equilibrium (zero applied bias). Positive voltage applied to one of the electrodes will initially drop across the reverse-biased PN junction closer to that electrode. At higher voltages, the nonzero electric field that extends across all of the middle P-region will result in lowering of the potential barrier for electron injection from the other (grounded) electrode. This mechanism is similar to the well-known drain-induced barrier lowering (DIBL) or punch-through effect that causes enhanced drain-to-source OFF-state leakage current in short-channel MOSFETs. By symmetry, the same behavior would be observed if the opposite voltage polarity is applied across the NPN diode.

Srinivasan *et al.* [40] obtained $>1 \text{ MA cm}^{-2}$ ON-state current density and maximum ON/OFF ratio $>10^6$ (with OFF-current measured close to 0 V) on epitaxially grown Si NPN punchthrough diodes (Figure 22.2). Threshold voltage tunability by adjusting the doping profile to get asymmetric turn-on voltages was demonstrated in Ref. [41]. Since the epitaxial growth process requires high temperatures ($\sim 700^\circ\text{C}$, according to [40]) as well as a crystalline Si template,

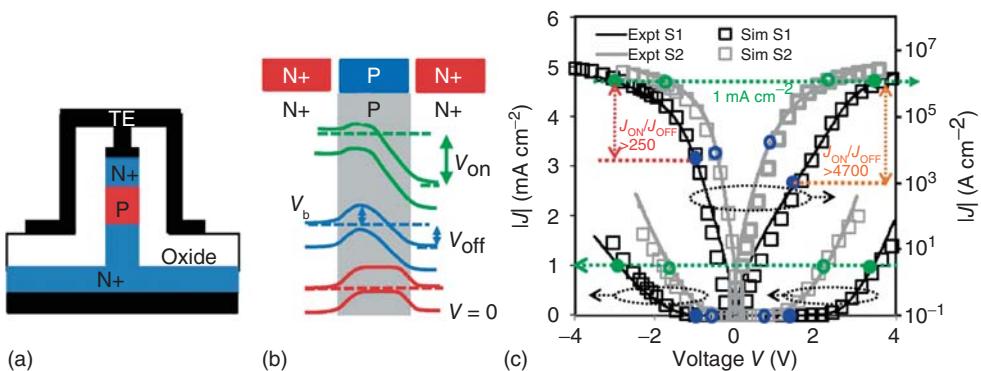


Figure 22.2 (a) Schematic cross-section through two-terminal bidirectional NPN Si punchthrough diode selector. (b) Band diagrams through NPN structure at equilibrium (zero applied bias to positive electrode), during exponential turn-on (V_{off} applied bias), and in the ON-state (applied voltage V_{on}). Note that the barrier V_b for electron injection reduces with increasing applied

voltage. (c) Experimental and simulated current density versus voltage plots for two structures, S1 and S2, with different doping profiles. (© IEEE. All rights reserved. Reprinted, with permission, from [40]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

such a device is unlikely to be either BEOL-compatible or easily stackable above the Si substrate. Polysilicon NPN structures made by Lee *et al.* [39] eliminate the need for a single-crystal Si seed layer but still require 800 °C annealing that rules out standard BEOL stacking. These devices showed bidirectional exponential turn-on with high maximum ON/OFF ratio but lower $\sim 0.1 \text{ MA cm}^{-2}$ ON-state current density.

Park *et al.* studied an MSM diode [42] with a P-type polysilicon layer contacted on both sides by metallic (Ta) electrodes, which had assumably low barriers to the Si conduction band edge. Simulations on such structures, proposed as select devices for STT-MRAM, showed $>1 \text{ MA cm}^{-2}$ ON-state current density and $>10^6$ maximum ON/OFF ratio. The authors suggest that this structure offers potentially better control of the middle P-region thickness compared with the doped NPN devices. However, the I - V characteristics are dependent on the metal–semiconductor barrier height.

22.4.2

Oxide Diodes

22.4.2.1 Oxide PN Junction

Most of the Si-based select devices discussed in the previous section are not compatible with 3D-stacked multilayer crossbar schemes due to the high processing temperatures needed during their fabrication. This has motivated research in the direction of non-Si-based select devices that can be made at lower

BEOL-compatible temperatures, such as diodes made from semiconducting oxide heterojunctions [43–47].

Baek *et al.* fabricated PN diodes using P-type NiO and N-type TiO₂ at sub-300 °C processing temperatures [43] and showed them working with unipolar oxide ReRAM elements. The ON-state current density of $\sim 1000 \text{ A cm}^{-2}$ in these diodes was quite low.

Lee *et al.* built two stacked layers of small crossbar arrays with unipolar NiO-based ReRAM accessed through P-CuO_x/N-InZnO_x diodes [45]. Their oxide diodes, deposited at room temperature, had higher ON-state current density ($> 10^4 \text{ A cm}^{-2}$) than the P-NiO/N-TiO₂ diodes from [43], attributed to the lower bandgaps of the oxides used here. Kang *et al.* published more detailed studies of P-CuO_x/N-InZnO_x diodes [46] including optical absorption measurements to estimate the bandgaps of the CuO_x and InZnO. Hall measurements revealed the conductivity of CuO_x to be P-type. While large area (electrode size $> 10 \mu\text{m}$) devices showed high rectification ratio (10^6) with ON-state current density of $3.5 \times 10^4 \text{ A cm}^{-2}$ and a forward bias ideality factor of 2 (this is 1 in ideal diodes; larger values indicate worse subthreshold slope), smaller 50 nm \times 50 nm crossbar-patterned diodes had much lower rectification ratio, possibly due to higher edge-leakage. Ahn *et al.* showed successful operation of one-time-programmable (OTP) Al₂O₃ memory elements accessed through P-CuO/N-InZnO_x thin film diodes and integrated in 8 \times 8 crossbar arrays [47].

22.4.2.2 Metal-Oxide Schottky Barrier

Rectifying *I-V* characteristics in metal–oxide–metal structures have been reported in the literature. Unlike the oxide PN junction devices discussed in the previous section, the rectification here is due to transport across a Schottky-barrier at the metal–oxide interface.

Tallarida *et al.* obtained high rectification ratio ($\sim 10^7$ – 10^8) and moderate ON-state current density ($\sim 10^4 \text{ A cm}^{-2}$) in their devices with ZnO sandwiched between Ag and Ti/Au electrodes [48]. In these diodes of size down to $36 \mu\text{m}^2$, ZnO was deposited by ALD at relatively low process temperatures ($\sim 100^\circ\text{C}$) and intrinsically N-doped polycrystalline films were obtained. ZnO was assumed to form a Schottky junction with the Ag bottom electrode and an ohmic contact to the Ti/Au top electrode (TE). The authors integrated these select devices with stacked unipolar NiO ReRAM elements in 10 Kbyte crossbar arrays and reported ReRAM SET switching through the diode.

Huang *et al.* measured rectification ratio $> 10^5$ and ON-state current density $\sim 2000 \text{ A cm}^{-2}$ in room-temperature-fabricated Ti/TiO₂/Pt MIM structures [49]. Based on the polarity of the *I-V* curves – low current and exponential turn-on with positive and negative voltage, respectively, applied to the Ti – the authors suggested, and confirmed via temperature-dependent measurements, that there were asymmetric barriers for electron transport from the metal electrodes into the TiO₂. The Schottky barrier height for electron injection from Pt into TiO₂ was found to be higher than that for injection from Ti into TiO₂. The role of inhomogeneous conduction through oxygen-deficient filaments in the TiO₂ was

also highlighted. Reliability studies showed that the diodes survived ± 3 V stress applied for 1000 s without degradation and 1000 \times repeated cycling between ± 3 V with some reduction in forward and reverse currents. However, application of higher voltages (~ 5 V) with current compliance caused breakdown and turned these diodes into bipolar ReRAM elements.

Park *et al.* [50] obtained very high rectification ratio $\sim 10^9$ in Pt/TiO₂/Ti Schottky-type diodes. The device operation was explained by assuming that the bottom Ti electrode interface was an ohmic contact while the TE Pt had a ~ 1 V barrier for electron transport into the TiO₂ deposited by ALD. While the measured ON-state current density was very low (~ 10 A cm⁻²) in these rather large devices ($\sim 60\,000$ μm^2), the authors used conducting AFM measurements to infer that the ohmic contact area at the Ti/TiO₂ interface was effectively confined to very small regions. Based on the estimated local ON-state current density of $\sim 3 \times 10^5$ A cm⁻² from these measurements, the authors suggested that these devices would show better ON-state current density when scaled down in size. Measurements on smaller crossbar-type Pt/TiO₂/Ti/Pt devices [51] indeed showed higher ON-state current density $\sim 3 \times 10^5$ A cm⁻² in $2\,\mu\text{m} \times 2\,\mu\text{m}$ devices although with lower rectification ratio ($\sim 2.4 \times 10^6$) compared with the earlier devices in Ref. [50].

There have been reports of Ni/TiO₂/Ni structures [52, 53] which show bipolar nonlinear *I-V* characteristics. The devices reported by Huang *et al.* [53] had ON/OFF ratio of $\sim 10^6$ between 0 and ± 2 V. Furthermore, the nonlinearity was maintained even under high temperature (125 °C) operation and after 1000 \times DC cycling. The device operation was ascribed to electron emission over the Ni/TiO₂ Schottky barrier whose height was estimated to be ~ 0.58 V at zero applied bias. Ni/TiO₂/Ni devices of varying sizes down to 0.36 μm^2 were reported by the same group in Ref. [52]. Figure 22.3 depicts some measured *I-V* curves from this work. The authors suggested the presence of nonuniform current conduction to explain the observed weak scaling of ON-state current with device area. They found that the calculated ON-state current density increased upon reducing the device area and obtained $\sim 10^5$ A cm⁻² in their smallest ($0.36\,\mu\text{m}^2$) devices. These Ni/TiO₂/Ni structures were used as bidirectional select devices to access Ni/HfO₂/Pt bipolar ReRAM. Leveraging the room temperature deposition process for the select devices as well as ReRAM elements, the authors built small 8 \times 8 crossbar 1S-1R arrays on flexible plastic substrates.

22.4.3 Threshold Switch

A threshold switch makes for an ideal select device, since from an initial highly resistive state the device turns ON to a highly conducting state as soon as a threshold voltage or current is applied. Various types of threshold switches such as ovonic threshold switching devices, Metal-insulator transition, (MIT)-based devices, and a threshold vacuum switch (TVS) have been proposed for application as a select device for crossbar memory.

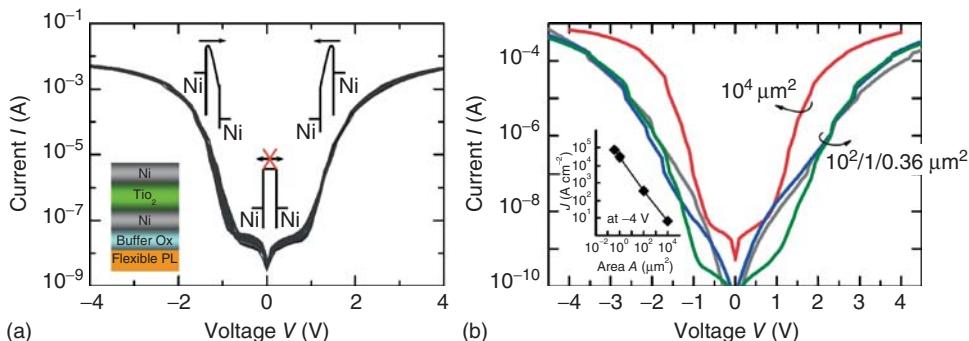


Figure 22.3 (a) Bidirectional exponential current–voltage curves measured on experimental Ni/TiO₂/Ni devices. (b) Experimental I–V curves measured on Ni/TiO₂/Ni bipolar selectors of varying area. The inset shows that the ON-state current density increases

as the device area is reduced. (© IEEE. All rights reserved. Reprinted, with permission, from [52]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

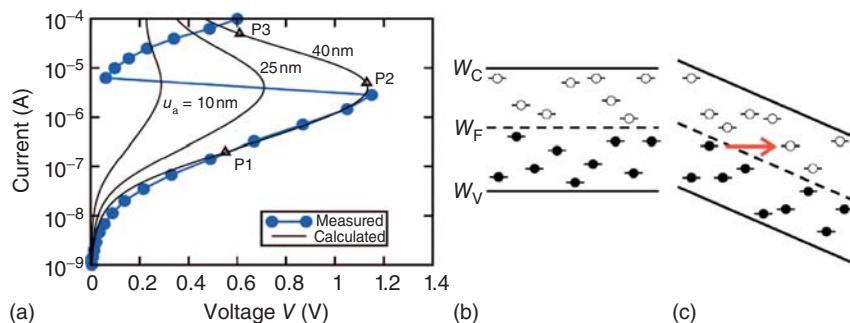


Figure 22.4 (a) Measured I–V curve for a phase-change memory cell with amorphous Ge₂Sb₂Te₅ chalcogenide and dependence of threshold voltage on thickness of the amorphous chalcogenide. (Reprinted with

permission from [55]. © The American Physical Society.) (b) Electron tunneling into higher energy traps can occur under high field. (Reprinted with permission from [56]. © 2007, AIP Publishing LLC.)

22.4.3.1 Ovonic Threshold Switching

The ovonic threshold switching (OTS) phenomenon was first reported by Stanford Ovshinsky in thin films of amorphous chalcogenide alloys [54]. Unlike high field electrical breakdown phenomena, OTS is repeatable and nondestructive. The current–voltage (I – V) curve of a simple two-terminal threshold switching device measured with current sweep mode is generally ‘S’ shaped (Figure 22.4a), switching rapidly from an initial high resistance ‘OFF’ state into a highly conductive dynamic ‘ON’ state once the applied voltage exceeds the *threshold voltage*, V_{th} . The device remains in the ON state as long as a minimum holding voltage, V_h (or current I_h), is present across the device, otherwise switching back into the OFF state.

This threshold switching occurs after a finite delay time (t_d) [57], which could potentially affect the upper limit for the speed of the integrated memory device [58]. However, it has been found that the switching delay time (t_d) decreases with increasing applied voltage and even \sim ns switching speed can be obtained [57–60]. The threshold voltage tends to decrease with decreasing thickness [61, 62], as well as with increasing temperature [62], showing that the threshold event is triggered by a threshold electric field [61, 62]. However, holding voltage (V_h) is independent of the chalcogenide thickness but is affected by the electrode materials [62].

To explain the threshold switching phenomenon, various theoretical models including thermally induced instabilities [63], impact ionization and Shockley–Hall–Read recombination [57, 64], polaron instabilities [65], and field-induced crystallization [59, 66] have been proposed. Recently, Ielmini proposed a model based on trap-limited current conduction, similar to Poole–Frenkel (PF) transport at high electric fields [55]. Applied electric field decreases the barrier height for electron hopping conduction, which in turn dramatically increase the conductivity [55]. As the applied voltage increases, the carriers gain energy and tunnel into higher energy traps (Figure 22.4b). When these traps are filled, mobility increases and current rises until switching occurs. This model has successfully explained experimentally observed trends for threshold voltage as a function of both temperature and thickness [55].

For use as a select device, it is important that any undesired structural rearrangement, such as crystallization, is suppressed during threshold switching and high-current operation of an OTS device [62]. While similar chalcogenide materials are used in phase change memory (PCM) devices, where both threshold switching and fast crystallization is desirable, it is possible to find materials that exhibit threshold switching without fast crystallization [57, 58, 60, 62]. Various materials have been reported for the OTS as a memory select device [58, 67, 68].

Kau *et al.* demonstrated the successful integration of OTS select device on top of a “mushroom-cell” PCM device in a stackable cross-point structure [67]. Figure 22.5a shows a TEM of the PCM device and OTS select device, while Figure 22.5b shows the measured I - V characteristics. Unfortunately, the use of a linear current scale makes it impossible to gauge the actual ON–OFF contrast of this device.

The higher threshold voltage of the OFF state of the stacked device pair is attributed to the serial connection of two amorphous alloys, OTS and amorphous PCM [67]. Figure 22.5c shows the cut-away SEM image of a 64 Mb “PCMS” (PCM + OTS) test chip fabricated using 90 nm technology node [67], which was used to explore cell sizes ranging from 40 to 230 nm. Single PCMS devices showed fast reset speed of 9 ns and pulse endurance of 10^6 cycles; statistics over a 2 Mb block showed more than 1 V of dynamic range between the highest threshold voltage of a PCMS device in the SET state and the lowest threshold voltage of a device in the RESET state.

Lee *et al.* have reported threshold switching in AsTeGeSi-based material by utilizing electronic charge injection [68]. Improved characteristics, including

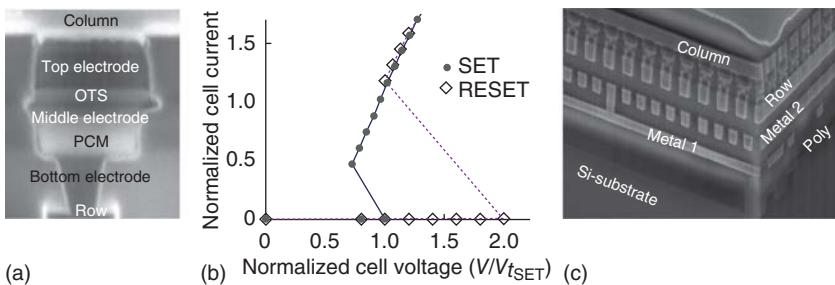


Figure 22.5 (a) SEM cross section of a PCMS (OTS + PCM) cell; (b) I - V characteristics of such a cell in SET (ON state) and RESET (OFF state) state; and (c) one layer of an integrated PCMS array. (© IEEE. All rights reserved. Reprinted, with permission, from [67]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

a reduction in threshold voltage and current, were observed after N_2 plasma nitridization, which led to higher tellurium (Te) composition. Fabricated AsTe-GeSiN threshold switches have shown pulse switching endurance of more than 10^8 cycles at a pulse width of 1 μs . Such an OTS device (1S) was integrated with TaO_x -based resistive memory (1R) to realize a 1S1R memory cell. The 30 nm (10 nm of Ta_2O_5 on 20 nm TaO_x) ReRAM cell was located under the 40 nm OTS cell, separated by the bottom TiN electrode of the OTS deposited onto the top Pt electrode of the ReRAM cell. An ON-OFF selectivity of 100 was observed for 30 μm devices with a leakage floor of 1 μA for $V_{SET}/2$, but scaling experiments suggested that a selectivity of 1000 could potentially be achieved for 30 nm devices.

22.4.3.2 Metal–Insulator Transition (MIT)

Metal–insulator transition (MIT) behavior in transition metal oxides has been a widely researched topic in the field of condensed matter materials [69]. It has been observed that various stimuli (magnetic, optical, and pressure) can activate MIT behavior [69]. Recently, electrically or thermoelectrically triggered MIT devices have become attractive candidates as bidirectional select devices for ReRAM application because of their simple metal/metal–insulator stack, high ON/OFF ratio, and fast transition speed.

Vanadium dioxide VO_2 possesses a first-order MIT. In micron-scale VO_2 devices, MIT behavior cannot be observed – in such large devices, even though the XRD pattern indicates only a single monoclinic phase, there are invariably numerous leakage current paths through nonstoichiometric oxide defects [70, 71]. In nanoscale devices, however, as applied voltage is increased from 0 to 0.5 V, the current suddenly increases at 0.35 V as the electrical property of VO_2 changes from an insulating to a metallic state [70, 71]. When the voltage is brought back down below 0.2 V, the current suddenly decreases as the VO_2 returns to its initial insulating state. Nanoscale VO_2 devices show very fast response (<20 ns) to an

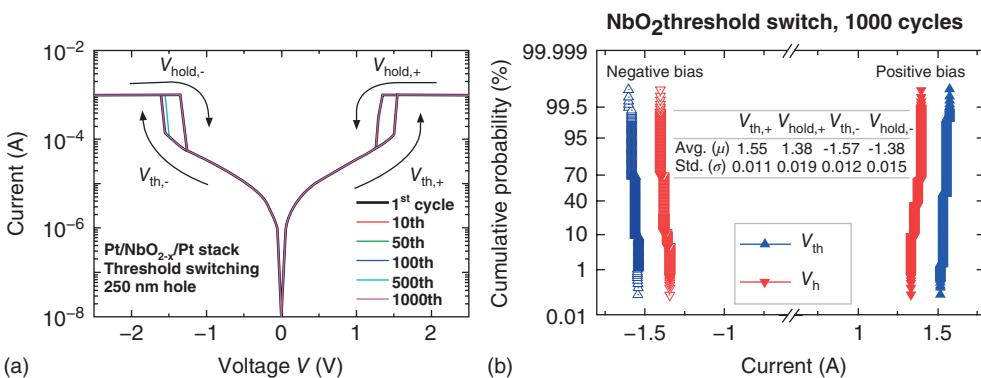


Figure 22.6 (a) $I-V$ characteristics of MIT behavior of Pt/NbO₂/Pt device. (b) Distribution of V_{th} and V_{hold} during single-device cycling. (© IEEE. All rights reserved. Reprinted, with permission, from [74].)

Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

input voltage signal, and individual devices show stable and uniform threshold and hold voltages over 100 successive switching cycles [70, 71].

While VO₂ exhibits excellent MIT characteristics at room temperature, its transition temperature is too low ($\sim 67^\circ\text{C}$) for practical device application at operating temperatures up to 85°C . NbO₂, however, offers a bulk transition temperature of 1081 K [72]. Recently, ultrathin NbO₂ film, intentionally fabricated by reactive sputtering method and patterned to nanoscale device area, has been investigated for select device applications [73, 74]. Similar to VO₂ devices, individual NbO₂ MIT-based select devices exhibit well-behaved and repeatable hysteretic $I-V$ characteristics over 1000 switching cycles (Figure 22.6) [74]. Thermal simulations have shown that Joule heating through the device causes the internal maximum temperature to exceed 1100 K, triggering the MIT [74]. Unlike the VO₂ device, NbO₂ select device shows excellent thermal stability up to 160°C to guarantee reliable memory application [74]. However, the threshold voltages are significantly larger and the hysteretic window smaller than VO₂, and the leakage current corresponding to the half-select condition is still fairly high.

22.4.3.3 Threshold Vacuum Switch

Recently, a TVS was proposed as a bidirectional select device element for 3D stackable crossbar arrays [75]. The TVS stack consists of W/TiN bottom electrode, TiN/vacuum TE with vacuum gap of ~ 1 nm (established by a SiN sacrificial layer), and WO_x as a switching layer, prepared by electrochemical oxidation.

Volatile switching behavior with high current density ($> 10^8 \text{ A cm}^{-2}$) and high ON–OFF selectivity of $> 10^5$ between ON and OFF states was observed in individual devices upon application of modest (1.5 V) dc or pulsed bias. The TVS ON state in the low voltage region corresponds to direct tunneling, with the high voltage region limited by the ohmic conduction WO_x layer. The TVS OFF state

is described as the superposition of direct tunneling and nonlinear conduction. Constant voltage stress tests with ± 2 V, 100 ns pulses applied with a sampling interval of 5 ms showed endurance over of $>10^8$ cycles. However, it is not clear how easily an array of devices with identical ultra-narrow vacuum gaps could be fabricated.

22.4.4

Oxide Tunnel Barrier

One attractive approach for achieving strongly nonlinear I - V characteristics is to use a thin oxide or nitride layer as a tunneling barrier. Such field-induced tunneling behavior is the basis of programming in conventional NAND FLASH memory [76], and allows currents to increase exponentially by several orders of magnitude at both polarities via electron tunneling behavior. Tunnel barriers can be formed by various high-k materials (HfO_2 , SiO_2 , ZrO_2 , and TiO_2) [77, 78] and the field sensitivity enhanced by tunnel-barrier engineering of multilayer oxides. According to how the dielectrics are stacked, such engineered tunnel barriers are classified as crested and variable oxide thickness (VARIOT)-type barriers [79, 80].

22.4.4.1 Single Layer Oxide-(Nitride-)Based Select Device (TiO_2 and SiN_x)

By choosing electrode materials with suitable work functions relative to the oxide/nitride-based semiconductor layers (TiO_x or SiN_x), Schottky barriers can be formed at both interfaces [81]. At low bias, electron injection is initially limited by one of the two barriers. When higher bias is applied, band bending leads to increased current. Symmetrical device structures lead to the symmetric I - V curves needed for bipolar ReRAM applications.

In devices composed of an ultrathin- TiO_2 layer deposited by atomic layer deposition (ALD) on a TiN bottom layer, when a positive voltage is applied to the Pt TE, electrons transfer from TiN layer to Pt layer [82]. In this case, current flow is relatively high (15 kA cm^{-2} at 2 V) because of the small ($\sim 0.4 \text{ eV}$) Schottky barrier at the TiN/ TiO_2 interface [82]. The current flow under negative bias is smaller (1 kA cm^{-2} at -2 V) because the barrier at the Pt/ TiO_2 interface is larger. Both the Schottky barrier height and the film thickness increase the current density exponentially. By changing the injection time of oxidizer in the TiO_2 -ALD process, the ratio of lattice to nonlattice oxygen can be controlled, leading to a change in the number of oxygen vacancies and a modulation in the work function of TiO_2 [82]. This allows the current density level of select devices to be adjusted to match with the particular ReRAM device of interest.

22.4.4.2 Multi-Layer Oxide-Based Select Device ($\text{TaO}_x/\text{TiO}_2/\text{TaO}_x$)

A crested-barrier device concept was proposed using combination of the TaO_x , and TiO_2 ultra-thin films [79, 83, 84]. In such a crested barrier device, the tunneling barrier height is maximum in the middle and gradually decreases toward the conducting electrodes [78]. In contrast, a VARIOT structure contains

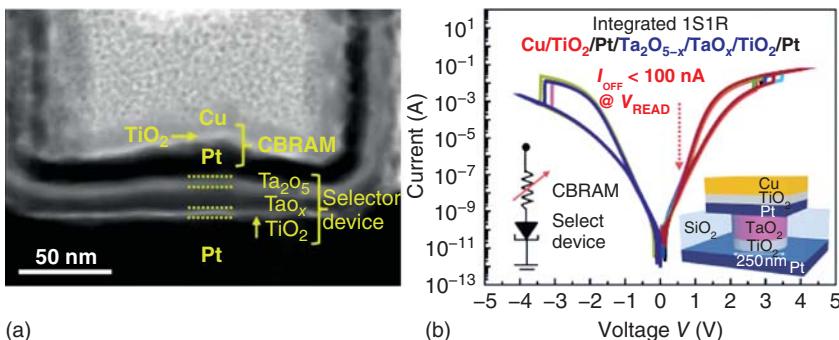


Figure 22.7 (a) Cross-sectional TEM image of a selector using engineered tunnel barriers (Pt/Ta₂O₅/TaO_x/TiO₂/Pt stack) integrated with CB-RAM in the 250-nm via-hole structure and (b) the *I*-*V* characteristics of CB-RAM and integrated 1S-1R device. (Reprinted, with permission, from [80]. © 2013, The Japan Society of Applied Physics.)

a central region with a lower bandgap than the outer regions. With a crested barrier structure, highly promising device performance can be obtained [80, 85]. The crested barrier was formed by Ta diffusion from the TaO_x layer into the TiO₂ layer, where it substitutes due to the similar ionic radii of Ti⁴⁺ and Ta⁵⁺. The resulting defect sites bend the energy bands of the TiO₂ film to form the crested barrier. This interpretation is supported by secondary ion mass spectroscopy of the depth profile of the structure [80, 85].

High current densities ($>10^7 \text{ A cm}^{-2}$) and a nearly symmetric *I*-*V* curve have been observed, together with a large ($\sim 10^4$) ON-OFF contrast. In the low bias OFF state, electron conduction is restricted by the higher total thickness. When the bands bend at higher bias, current increases exponentially under either polarity, via tunneling through the extremely thin oxide layer. No disturbance was observed under V_{READ} and $V_{\text{READ}}/2$ conditions out to 10^{10} pulses of 100 ns width [80].

A scalable Cu-based Conductive Bridging (CBRAM) was directly integrated on top of such a select device, using a via-hole structure (Figure 22.7a) [80]. Figure 22.7b shows the *I*-*V* characteristics of the 1S-1R combined device pair. The presence of the select device suppresses the current at the original $V_{\text{READ}}/2$ by a factor of 10^3 .

22.4.5

Mixed-Ionic-Electronic-Conduction (MIEC)

BEOL-friendly select devices based on Cu-containing MIEC materials [86–90] have become an intriguing choice as 3D-ready select devices for NVM. MIEC-based select devices offer large ON/OFF ratios ($>10^7$), high voltage margin V_m (over which current $< 10 \text{ nA}$), and ultra-low leakage ($< 10 \text{ pA}$), while also offering the high current densities needed for PCM and the fully bipolar operation needed for high-performance ReRAM [86, 87]. These devices contain a large amount of

mobile copper ions, which can move readily within the MIEC-based material. At low bias, Schottky barriers at the MIEC-electrode interfaces are believed to strongly suppress current flow, leading to the ultra-low leakage. As bias increases in either direction, copper ions and vacancies shift accordingly within the device, modulating the interfaces and leading to an exponential increase in electronic current. Although the current eventually saturates, extremely high current densities have been observed (up to 50 MA cm^{-2} [86]). Although higher drive currents have been observed to sharply reduce cycling endurance, $\sim 10^8$ cycles has been demonstrated at $\sim 150 \mu\text{A}$ [87].

Integration of MIEC-based select devices has been demonstrated on 8" wafers, with Cu-containing MIEC material sputtered into vias followed by an optimized Chemical-Mechanical Planarization (CMP) process [87] and a confined, non-ionizable TE. conductive-atomic force microscopy (AFM) testing of short-loop devices with minimal wiring has validated single-target deposition and revealed a wide processing temperature window (up to 500°C) for these MIEC-based select devices [88]. Although these select devices do not need high processing temperatures, MIEC-based select devices would not be adversely affected if a co-integrated ReRAM or other NVM did require such processing conditions.

Bidirectional array diagnostic monitor (ADM) arrays of up to 512×1024 integrated MIEC select devices have been tested using integrated 1-bit sense-amplifiers and a fast electrical tester (Magnum 2 EV) [88]. Cumulative distribution functions (CDFs) of the bit line voltage V_{BL} needed to produce various device currents I_d show tightly distributed array I - V characteristics (Figure 22.8a). All 524,288 MIEC devices – 100% – had $V_m > 1.1 \text{ V}$, and 99.955% of select devices fell within $\pm 150 \text{ mV}$ of the median voltage margin $V_m = 1.36 \text{ V}$ at 10 nA (Figure 22.8c) [88].

Thickness scaling experiments showed that as MIEC-based select devices become thinner, voltage margin remains mostly unchanged until the minimum gap between electrodes, d_{min} , reaches $\sim 11 - 12 \text{ nm}$ [89]. Short-loop MIEC select devices fabricated with ultra-scaled vias show both high-yield and high-voltage margin [89]. Despite their small size, these MIEC select devices can still rapidly drive the large currents needed for NVM switching (Figure 22.9a). Voltage margin V_m (at 10 nA) improves markedly as devices are scaled in lateral size. Aggressively scaled MIEC select devices retain all requisite characteristics, including ultra-low leakage ($< 10 \text{ pA}$) and the large voltage margins ($V_m > 1.50 \text{ V}$) needed for large arrays, even with both top and bottom critical dimensions (CDs) $< 30 \text{ nm}$ (Figure 22.9b,c). Unlike thickness scaling, where leakage increases sharply below $d_{min} \sim 11 \text{ nm}$, no lower limit to CD scaling has yet been identified [89].

Integration of MIEC select devices immediately above small (CD $\sim 35 \text{ nm}$) PCM devices allowed testing of integrated PCM + MIEC device pairs. Figure 22.10 demonstrates endurance in excess of 100 000 cycles, despite the repeated application of RESET pulses $> 200 \mu\text{A}$ and $5 \mu\text{s}$ long SET pulses at $\sim 90 \mu\text{A}$ [88]. To demonstrate NVM write speed capabilities, an MIEC-based select device was used to rapidly RESET a co-integrated PCM device. After each 15 ns RESET pulse

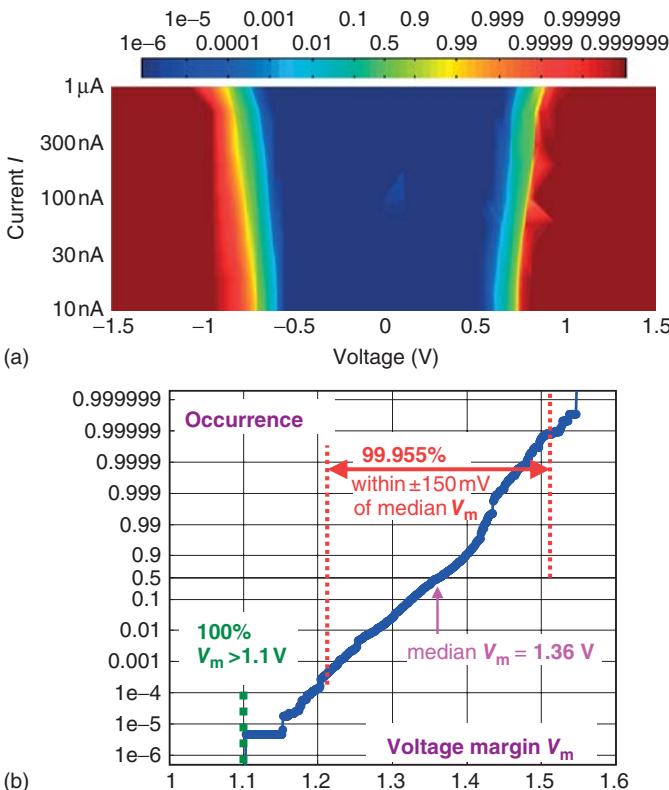


Figure 22.8 Cumulative distribution functions (CDFs) across bit line voltage V_{BL} at various device currents I_d (a) show array level I - V results with tight distributions across a 512×1024 array of integrated MIEC ADs [88]. (b) Within this same 512×1024 array, there were no leaky devices; 100% of the array showed $V_m > 1.1\text{ V}$, while 99.955%

of MIEC ADs had voltage margins V_m at 10 nA within $\pm 150\text{ mV}$ of the median of 1.36 V . (© IEEE. All rights reserved. Reprinted, with permission, from [88]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

at varying amplitude, a long SET pulse recrystallized the doped- $\text{Ge}_2\text{Sb}_2\text{Te}_5$ PCM material. After each pulse, bipolar DC I - V curves showed the expected change in the PCM resistance above the same low-leakage characteristics of the MIEC select device [88]. The application of shaped pulses or a transient “overvoltage” read readily allows MIEC select devices to support $\sim 5\text{--}10\ \mu\text{A}$ NVM reads in $< 50\text{ ns}$ [89, 90].

It has been shown that the ultra-low leakage offered by MIEC-based select devices representing the *unselected* state can be held for hours, while higher *half-selected* leakage can be sustained for at least a few seconds, more than long enough for millions of successive $1\ \mu\text{s}$ reads to the same bit line [90]. In addition,

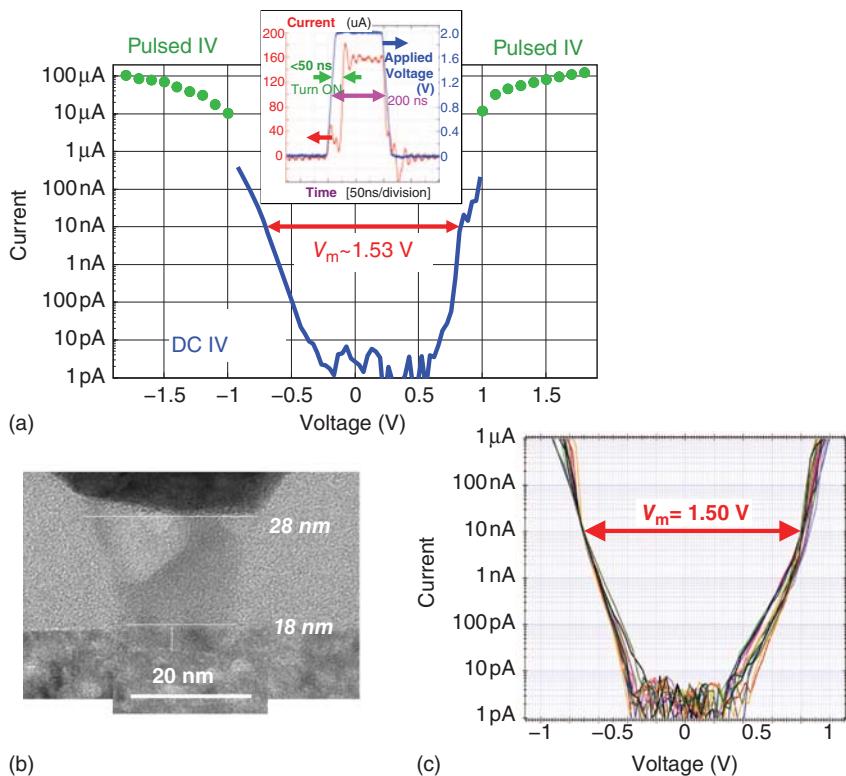


Figure 22.9 (a) In addition to high yield, scaled short-loop MIEC ADs [89] exhibit the same $>10^7$ ON–OFF contrast, <50 ns turn-ON times (test-setup-limited), and ultra-low leakage shown in larger devices [87, 88]. Aggressively scaled short-loop MIEC ADs, with (b) both TEC and BEC <30 nm, also

show (c) large voltage margins and ultra-low leakage. (© IEEE. All rights reserved. Reprinted, with permission, from [89]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

MIEC-based select devices can return rapidly from either the *selected-for-read* or *selected-for-write* states [90]. Read operations can be sub-50 ns, fast enough for use with MRAM [90]. Inherently fast thin MIEC select devices offer similar speeds at modest overvoltages, which is important for minimizing the possibility of read disturb [90].

Thus, MIEC-based select devices are well-suited for both the scaled CDs and thicknesses of advanced technology nodes, and for the fast read and write speeds of emerging NVM devices. Future improvements would include a better quantitative understanding of the interaction between the mobile Cu-ion dopants and the resulting electronic current [91], improvements in endurance at high current (currently $\sim 10^8$ cycles at $150\text{ }\mu\text{A}$ [87]), verification of high endurance in ultra-scaled

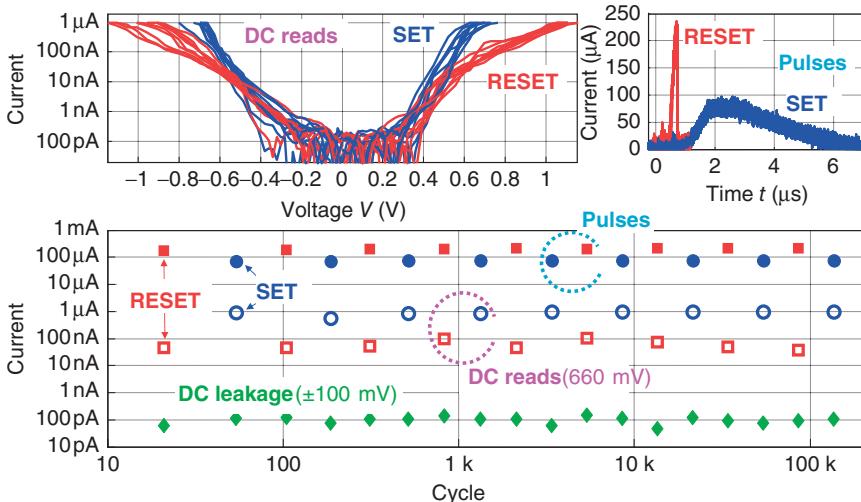


Figure 22.10 Endurance of an integrated PCM + MIEC device-pair to >100k cycles, with RESET currents >200 μ A and 5 μ s-long SET pulses (\sim 90 μ A). No select device degradation or PCM failure had occurred at the time testing was terminated. (© IEEE. All

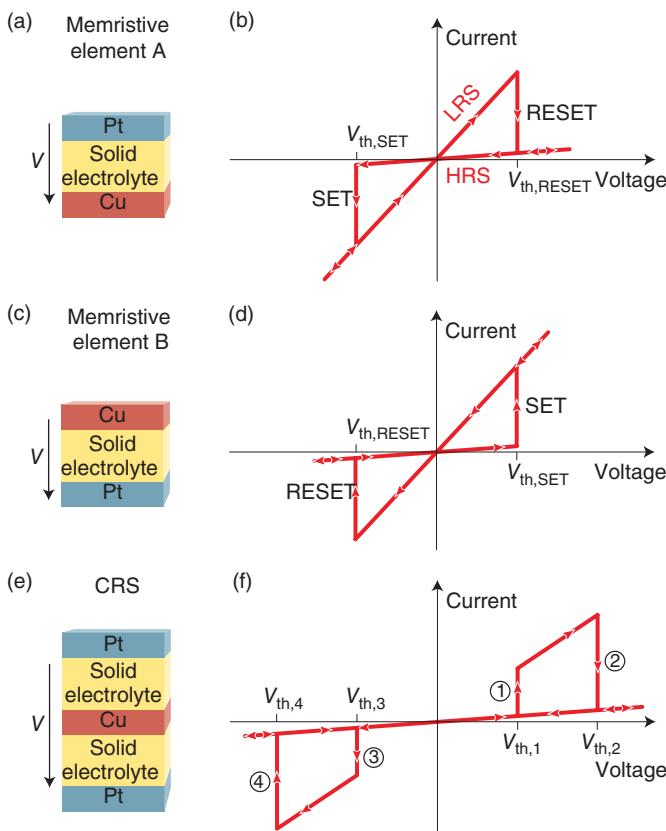
rights reserved. Reprinted, with permission, from [88]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

devices at both high and low operating temperatures, and increases in the voltage margin so as to enable large arrays for NVM devices requiring >1.5 V switching voltages.

22.5

Self-Selected Resistive Memory

In a self-selected resistive memory device, sneak path leakage currents are suppressed by using a ReRAM device whose I - V characteristics are strongly nonlinear in all resistance states. For example, in a complementary resistive switch (CRS), two bipolar ReRAM cells (preferably with symmetric set/reset voltages) are stacked back to back with one common electrode. In a hybrid ReRAM cell, built-in select device functionality is integrated in a single cell structure using one of the stand-alone select device mechanisms already discussed. In a nonlinear ReRAM cell, additional nonlinearity is introduced into the LRS by inserting barrier layers at the electrode/swing material interface to reduce the leakage current. The desired outcome in each of these approaches is select-device functionality with a simplified fabrication process and smaller cell size and/or thickness.



Possible states of the CRS

CRS state	Memristive element A	Memristive element B	Resistance CRS
0	HRS	LRS	\approx HRS
1	LRS	HRS	\approx HRS
ON	LRS	LRS	LRS + LRS; only during read operation
OFF	HRS	HRS	$>>$ HRS; only in initial state

State 'ON' occurs only during the read operation. State 'OFF' is the initial state without further importance for the operation and is present only after fabrication. The states '0' and '1' are the memory states for which the overall resistance of the CRS is the same:
 $R_{CRS} = R_{HRS} + R_{LRS} \approx R_{HRS}$ (with $R_{LRS} \ll R_{HRS}$).

Figure 22.11 By combining (a) bipolar resistive element A with a Pt/solid electrolyte/Cu stack with its (b) I - V characteristic, together with (c) a reversed copy of the same cell, bipolar element B with a Cu/solid electrolyte/Pt stack and its (d) I - V characteristic, the (e) stacked combination of elements A and B device acts as a complementary

resistive switch (CRS) with a (f) select device-like I - V characteristic exhibiting (g) two stored data states amongst the four possible combinations of individual device states, both offering high resistance under half-select conditions. (Reprinted by permission from Macmillan Publishers Ltd: Nature Materials [92], © (2010).)

22.5.1

Complementary Resistive Switch

The CRS proposed by Linn *et al.* [92] consists of two antiferentially connected bipolar resistive switches, as illustrated in Figure 22.11. In this first manifestation, each resistive switching element was a conductive bridging memory (CBRAM) with an oxidizable (Cu) electrode and a solid electrolyte (such as GeSe). A positive voltage on the TE can form a filament in the bottom element and dissolve the filament in the top element.

The four possible combinations in the two elements of the high resistance state (HRS, or no filament) and the LRS (or copper filament) are summarized in Figure 22.11g [92]. After initialization of the original HRS/HRS state, HRS/LRS can be defined as the state '0,' and LRS/HRS as the state '1.' The final state, LRS/LRS, is the 'ON' state of the CRS cell obtained during the reading operation. Sneak current is suppressed by the presence of at least one HRS state in both stored data states.

Unfortunately, the read operation of the CRS cell is destructive for the memory state of '1' (LRS/HRS), because a positive read voltage greater than V_{th1} (Figure 22.11f) triggers the filament in the bottom cell and places the CRS into the LRS/LRS or 'ON' state. Since a measurable current is detected, the '1' memory state can be detected. However, after the read operation, a write-back process is required to return the device stack from the 'ON' state back to memory state '1'. When the cell stores '0' (HRS/LRS), a positive readout voltage has no effect on the bottom cell, leading to low read current and no need for write-back. One challenge for empirical CRS device-pairs is attaining symmetric set and reset voltages in the two devices, lest the 'ON' state become unstable [11, 92, 93]. By using an external series resistor, threshold voltages can be shifted to help obtain a stable 'ON' state [11, 92], allowing the CRS concept to be applied to any bipolar resistive switching material, irrespective of the symmetry in set/reset voltages [11].

Rosezin *et al.* successfully demonstrated the vertical integration of CRS cells based on Cu/SiO₂/Pt bipolar switches for passive crossbar array applications with an OFF/ON resistance ratio of >1500 and switching speed of <120 µs [94]. Yu *et al.* have identified suitable pulse widths and pulse amplitudes for CRS devices based on Ag/GeSe elements [11]. In order to get similar read/write windows over

a wide timescale, it is best that the set/reset switching dynamics of the two individual elements be the same, because the speed of the aggregate CRS device will be limited to the speed of the slowest switching process. A CRS device formed by antiserially connecting two ReRAM elements with Pt/ZrO_x/HfO_x/Pt structure in a cross-point structure was reported by Lee *et al.* [95]. In this demonstration, each ReRAM exhibits typical linear on-current characteristics, yet acts as a select device for reset-read-selection cycling of the other device.

Lee *et al.* reported a CRS device fabricated by connecting two Pt/Ta₂O_{5-x}/TaO_{2-x}/Pt cells antiserially via external contacts [96]. Interestingly, the individual resistive memory cell in the mentioned stack has asymmetric *I-V* behavior with the Schottky barrier at the Pt/Ta₂O_{5-x} interface helping to suppress leakage current for voltages between -0.7 V and +0.7 V by either being reverse biased, or being below the threshold voltage of forward bias.

In contrast to these CRS devices, which require complex stacking of metal–oxide/metal layers, Nardi *et al.* recently demonstrated that the CRS properties can also be exhibited in single-layer nonpolar-ReRAM devices [97], by selectively severing the filament of oxygen vacancies either at the top or bottom interface. Then, instead of two separate devices each with its own filament, a single-layer device either contains a filament severed at the top or at the bottom. This behavior was empirically observed (Figure 22.12) in an oxide-ReRAM device in simple TiN/HfO_x/TiN structure with 5 nm thick HfO_x active layer [97]. In these devices, bipolar switching, CRS behavior, and unipolar switching could each be obtained by careful sequencing of switching operations and current compliance [97]. However, the current levels at which CRS behavior is obtained are fairly high, nor is it clear how the voltage and current conditions for CRS behavior might vary across a large array of such devices.

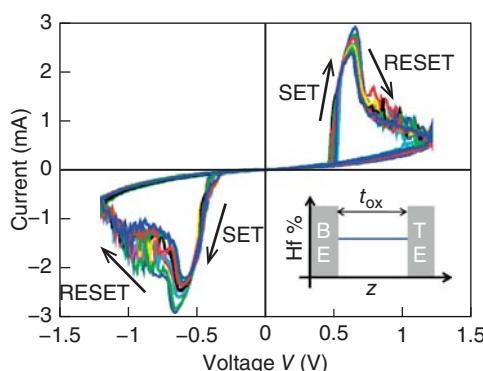


Figure 22.12 Measured *I-V* curves for symmetric ReRAM, demonstrating complementary switching (CS) in single stack nonpolar ReRAM. (© IEEE. All rights reserved. Reprinted, with permission, from [97].)

Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

Yang *et al.* have reported CRS behavior in a tantalum oxide-based single stack ReRAM device [98, 99], attributed to the redistribution of oxygen vacancies in the tantalum oxide layers. As with the HfO_x work, systematic adjustment of the component materials and stoichiometries allows different switching behaviors – including unipolar, bipolar, and complementary switching – to be obtained in bilayer ($\text{Pd}/\text{Ta}_2\text{O}_{5-x}/\text{TaO}_y/\text{Pd}$) and trilayer stacks ($\text{Pd}/\text{Ta}_2\text{O}_{5-x}/\text{TaO}_y/\text{TaO}_z/\text{Pd}$) [99]. In another study, CRS properties with low operating voltages (~ 1 V) were reported in $\text{Pt}/\text{TiO}_{2-x}/\text{TiN}_x\text{O}_y/\text{TiN}$ by Tang *et al.* [100]. One-step plasma oxidation of TiN film was used to partly oxidize the TiN bottom electrode, creating an oxygen reservoir layer of TiN_xO_y . The CRS behavior is attributed to redistribution of oxygen vacancies between the $\text{Pt}/\text{TiO}_{2-x}$ top interface and the $\text{TiO}_{2-x}/\text{TiN}_x\text{O}_y$ bottom interface.

22.5.2

Hybrid ReRAM-Select Devices

A hybrid memory is one in which a single device exhibits both memory as well as a select device properties. For instance, a combined MIT and memory device was achieved by controlling oxygen concentration in a vanadium oxide film [71]. While stoichiometric VO_2 film exhibited typical MIT behavior with selection properties, and a nonstoichiometric V_2O_{5-x} film showed typical resistive memory switching behavior, vanadium oxide film with various intermediate oxygen content showed hybrid characteristics. Co-existence of multiple phases of vanadium oxide at the electrode interface region was observed by X-ray photoelectron spectroscopy (XPS) [71].

Similar to vanadium oxide but offering a significantly higher transition temperature, niobium oxide can exhibit both memory and select device properties for intermediate oxygen concentrations (Figure 22.13) [74]. In TEM and EDX analysis, a 10 nm thick NbO_x layer is clearly observed. To confirm the high temperature stability of the hybrid memory device, DC measurement at 125°C was implemented, showing uniform memory and select device characteristics even at high temperature conditions [74].

A scalable transistorless cross-point technology was demonstrated by Meyer *et al.* by combining a novel oxide memory element and a co-integrated nonlinear tunnel diode [101]. The basic memory cell stack includes Pt as top and bottom electrodes, a thick (25 nm) crystalline perovskite conductive metal oxide (CMO) as the nonfilamentary switching layer, and a crystalline, high-quality dielectric tunnel oxide with the thickness of about 2–3 nm. Good crystallinity of both CMO and tunnel oxide are crucial in order to get desired memory and select device characteristics. Four layers of memory cells stacked in the BEOL above the silicon have been demonstrated in a $0.13\ \mu\text{m}$ 64 Mb test chip [102]. BEOL integration is possible because the sputter deposition of both CMO and tunnel oxide materials can be performed at temperatures from 380 to 450°C [101].

The I - V curves of the aggregate memory device can be tuned by varying the tunnel oxide thickness, and fit well to a trap-assisted tunneling mechanism

[101]. When positive voltage is applied to the TE, the oxygen ions migrate toward it and excess negative charge accumulates in the tunnel oxide, which repels the tunneling electrons. As a result, barrier height increases, tunneling current decreases and the measured device resistance increases (lower curves in Figure 22.14a). Conversely, negative voltage on TE forces the oxygen ions to move out of the tunnel oxide into the conductive metal oxide. Consequently, the tunneling current increases due to reduction in the effective barrier height, which leads to the LRS. The excellent scaling of initial, programmed, and erased states with area (Figure 22.14b) indicates the nonfilamentary switching mechanism.

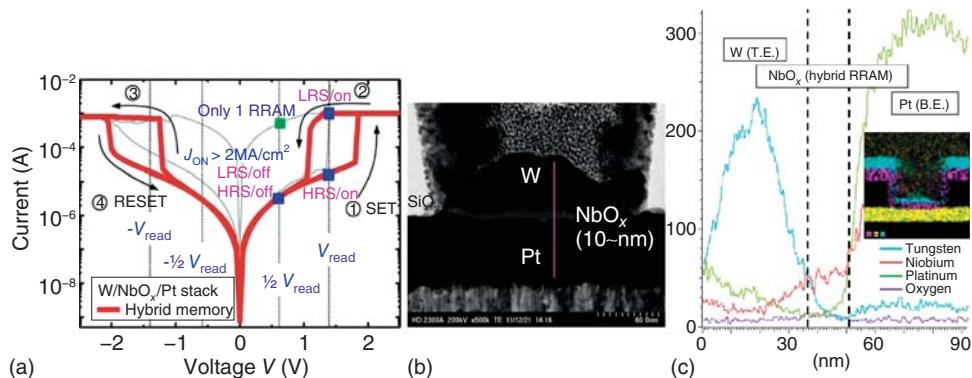


Figure 22.13 (a) I - V characteristics, (b) cross-sectional TEM image and (c) EDX line profile of hybrid ($\text{W}/\text{NbO}_x/\text{Pt}$) memory device. (© IEEE. All rights reserved. Reprinted, with

permission, from [74]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

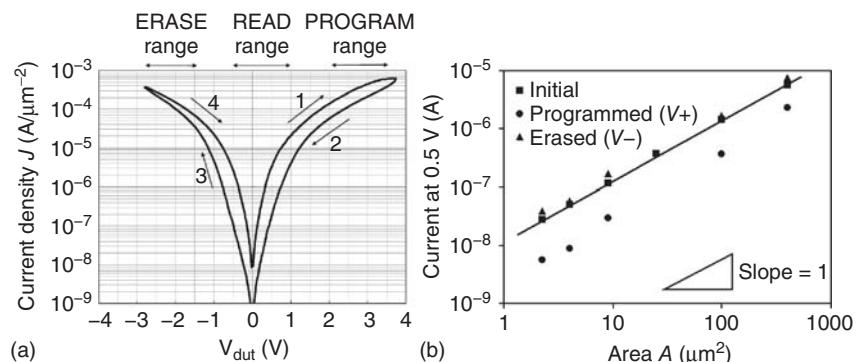


Figure 22.14 (a) High-field DC programming [102] and (b) area scaling of initial, programmed, and erased states [101]. (© IEEE. All rights reserved. Reprinted, with permission, from [101, 102]. Personal use

of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

No forming is required to obtain this resistive switching behavior. As both the resistance states are retained after program and erase operations, the memory device is nonvolatile with typical resistance ratio of ~ 10 . A cycling endurance of $> 10^6$ cycles with typical operation voltage of ± 3 V and $10\ \mu\text{s}$ pulse was obtained [101]. Since half-select leakage currents are fairly large, an asymmetric array favoring long wordlines and short bit lines is used [102, 103]. The readout scheme is timed carefully to sense the device state before the steadily increasing leakage from half-selected lines overwhelms the signal [102]. Since read currents are fairly low, readout is slow; however, incorporation of local gain stages allows the read to occur in $5\ \mu\text{s}$ [104] instead of $50\ \mu\text{s}$ [103].

22.5.3

Nonlinear ReRAM

Another alternative for suppressing sneak-path leakage is to introduce a strong nonlinearity in the switching I - V curve of the memory element, by inserting a thin layer that acts as a tunnel-barrier. Various types of nonlinear characteristics in ReRAM device have been reported [105–112], including a 2 Mb ReRAM crossbar array fabricated with 54 nm technology without select devices [111].

Switching that occurs at or near a $\text{TiO}_x/\text{Ta}_2\text{O}_5$ interface can be attributed to either charge trapping/detrapping (Type I switching, requiring high voltages) or oxygen vacancy migration (Type II switching, requiring high switching currents). It has been reported that the Type II shows better switching characteristics with higher switching speed ($< 10\ \text{ns}$), smaller operation voltage ($< 4\ \text{V}$), and better retention characteristics ($> 20\ \text{h}$ at 150°C) [111].

To achieve larger array size, low reset current ($< 3\ \mu\text{A}$) and high nonlinearity ($\gg 30$) are needed, depending on driver capability [111]. Further improvement in I - V curves can be obtained either by using different spacer materials or by changing the Ti/O ratio. Park *et al.* have reported a nonlinear ReRAM cell with ultralow operating current of $< 1\ \mu\text{A}$ by engineering the switching oxide and by inserting thin barrier layers between oxide and electrodes. Such cells have been shown to switch with $1\ \mu\text{A}$ current compliance and exhibit nonlinearities (the ratio between currents at the writing, and at half the writing voltage) up to $17 \times$ [112].

These two barrier layers have different roles in switching. In the bottom barrier, direct tunneling is dominant in the low-voltage regime but Fowler–Nordheim (FN) tunneling takes place in the high-voltage regime (Figure 22.15a), leading to high nonlinearity in the LRS [112]. The top barrier provides self-current compliance, improving pulse endurance by suppressing current overshoot but also offering a wider operation margin for low current switching [112]. The overall conduction mechanism in LRS is PF emission with tunneling through barrier layers as shown in Figure 22.15b. Such devices can be switched between LRS and HRS states for $> 10^7$ cycles with $1\ \mu\text{s}$ pulses, while maintaining a resistance ratio of $\sim 10^2$ and showing very little degradation in switching characteristics. As with conventional ReRAM, multiple different LRS resistance levels can be obtained for MLC operation by controlling the current compliance.

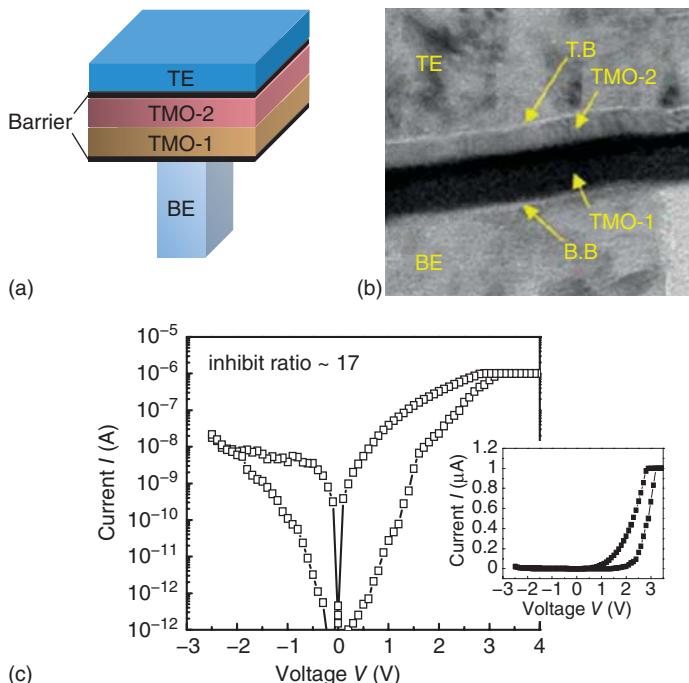


Figure 22.15 (a) Schematic of vertical ReRAM structure. Thin barrier layer is inserted between TMO and electrode on both sides. (b) X-TEM image ReRAM cell. (c) I - V curve of low current cell switching set current of $\sim 1 \mu\text{A}$ with nonlinear I - V

characteristic. (© IEEE. All rights reserved. Reprinted, with permission, from [112]. Personal use of this material is permitted. However, permission to reuse this material for any other purpose must be obtained from the IEEE.)

The dependence of nonlinearity on the current compliance and thickness of the switching material was investigated by Lentz *et al.* [113]. The device had a TiN/TiO₂/Ti/Pt stack with two different TiO₂ thicknesses, 5 and 25 nm. For the thicker sample, the highest nonlinearity (of 7.4 \times) was achieved at the lowest Forming/SET currents, because as SET current increases, the thicker filament exhibits a more ohmic behavior. For the thinner (5 nm) sample, a higher nonlinearity (of 30 \times) was measured [113].

Yang *et al.* successfully demonstrated a nonlinearity of ~ 100 in TaO_x-based devices by engineering the interface between TaO_x and TiO_{2-x} in a Pt/TaO_x/TiO_{2-x}/Pt stack [114]. They found that the Pt/TaO_x interface was responsible for the resistive switching, while the oxide heterojunction of TaO_x/TiO_{2-x} with Schottky-like metal/semiconductor contact produced the nonlinearity. The operation of this device is attributed to electrothermal effects, and a new oxide phase is formed in the conducting channel where local temperatures are higher because of Joule heating, leading to a negative differential resistance (NDR) effect.

22.6 Conclusion

In this chapter, we have discussed the need for select device functionality, to allow many “state-holding” memory devices to be organized into large arrays while supporting low-power programming and reliable readout. Such select devices should offer a large ON–OFF ratio between ultra-low leakage for half- and unselected devices, yet high ON current densities for rapid switching of selected memory devices. Various discrete select devices, and ways to allow the memory device itself to provide this functionality, were discussed.

After decades of research and development, Si-based select devices are very well understood. However, three-terminal devices are nontrivial to implement in a $4F^2$ footprint, and devices that require single-crystal silicon preclude 3D multilayer stacking. Polysilicon devices can be stacked, but high ON-state current density can typically be achieved only with high ($>400\text{ }^\circ\text{C}$) fabrication temperatures.

The main advantage of oxide PN junction select devices is their low fabrication temperature, providing compatibility with 3D multilayer stacking approaches. Select devices based on metal-oxide Schottky barriers offer both relative ease of integration and low ($<400\text{ }^\circ\text{C}$) processing temperatures. However, the ON-state current density of both oxide PN junctions and metal-oxide Schottky barriers needs improvement by several orders of magnitude for operation with most resistive memory elements. Both Si rectifiers and oxide PN junction diodes support only unipolar operation, and will not work with bipolar memory elements. However, it is possible to get bidirectional operation in metal–oxide–metal Schottky barrier devices by appropriate choice of metal electrodes.

Three types of threshold switches were discussed as potential select devices for ReRAM: OTS devices, MIT-based devices, and the TVS. OTS devices have been shown with good switching performance and in large arrays, but require complex materials and improvements in switching endurance and leakage current will be needed. It is also not yet clear what sort of ReRAM switching voltages and state resistances that OTS or other threshold select devices might be able to accommodate.

In contrast, MIT and TVS switches have been demonstrated only at the few-device level, both individually and in series with ReRAM devices. While MIT devices using materials such as NbO_2 with acceptably high threshold temperatures have been demonstrated, it will be important to further reduce the half-select and unselect leakage currents. Also, higher threshold temperatures implies that added electrothermal power is being used to trigger the MIT behavior, which could potentially increase the overall power required for NVM switching. The TVS device is interesting, but it will be critical to demonstrate that vacuum gaps with sufficiently tight thickness distributions, leading to tightly distributed electrical switching characteristics, can be demonstrated at high yield over large arrays of TVS select devices.

Oxide tunnel barriers offer steeply nonlinear curves, and have attained success in combination with various memories including CB-RAM and nonfilamentary

conductive metal oxides (CMOs). However, it is extremely important that the switching voltages of the ReRAM remain low so that the half-select leakage – as evaluated at half the applied voltage across both memory and select device, as well as across any extra bias dissipated in the wiring – can still be low.

BEOL-friendly select devices based on copper-containing MIEC materials offer large currents ($>100\text{ }\mu\text{A}$), bipolar operation, and ultra-low leakage ($<10\text{ pA}$). Co-integration with PCM, integration in large (512 kbit) arrays with 100% yield and tight distributions, fast transient operation, long-term persistence of the required ultra-low-leakage, and scalability to aggressive technology nodes have all been demonstrated. However, larger voltage margin than the $V_m \sim 1.6\text{ V}$ achieved to date will be required, if MIEC select devices are to be used with ReRAM devices with switching voltages larger than $\sim 1.3\text{ V}$ [23, 24].

Methods for adding select device functionality to ReRAM memory include the CRS, hybrid devices in which select device functionality is incorporated together with the memory functionality, and nonlinear ReRAM, in which barrier layers introduce nonlinearity to help reduce leakage current. Although a CRS device can significantly reduce sneak-path currents, destructive readout with subsequent write-back is required. In addition, each CRS device must either be fabricated from a stack of two well-behaved and symmetric memory elements, or a single ReRAM layer with very precise control over operating voltages. A particularly important step will be the demonstration of reliable CRS operation at the low switching currents ($20\text{--}50\text{ }\mu\text{A}$) required for implementation in the narrow-pitch wiring of advanced technology nodes [19].

Another aspect of self-selected memories is the difficulty in independently tuning select-device and memory functionality. This is more of an issue in hybrid devices using MITs or using tunnel barriers, which participate in the motion of oxygen ions, since the select function is nearly inseparable from the memory function. While nonlinearity can be tuned by increasing the thickness of an added tunnel barrier, the maximum nonlinearities achieved so far fall well short of the large ($10^6\text{--}10^7$) values needed for large (1000×1000) subarrays.

Although ReRAM devices are interesting in and of themselves, they will only have technological impact when they can be packed densely in vast arrays offering many Gigabytes if not Terabytes of solid-state storage. If this goal is to be reached, select device technologies will likely be an important, if not critical, contributor.

References

1. Wright, E.P.G. (1954) Electric connecting device. US:2,667,542.
2. Kim, K. and Song, Y.J. (2004) Current and future high density FRAM technology. *Integr. Ferroelectr.*, **61**, 3–15.
3. Raoux, S., Burr, G.W., Breitwisch, M.J., Rettner, C.T., Chen, Y.C., Shelby, R.M., Salinga, M., Krebs, D., Chen, S.H., Lung, H.L., and Lam, C.H. (2008) Phase-change random access memory: a scalable technology. *IBM J. Res. Dev.*, **52** (4/5), 465–480.
4. Hosomi, M., Yamagishi, H., Yamamoto, T., Bessho, K., Higo, Y., Yamane, K., Yamada, H., Shoji, M., Hachino, H., Fukumoto, C., Nagao, H.,

- and Kano, H. (2005) A novel non-volatile memory with spin torque transfer magnetization switching: Spin-RAM. *IEDM Technical Digest*, pp. 459–462.
5. Freitas, R. and Wilcke, W. (2008) Storage-class memory: the next storage system technology. *IBM J. Res. Dev.*, **52** (4/5), 439–448.
 6. Burr, G.W., Kurdi, B.N., Scott, J.C., Lam, C.H., Gopalakrishnan, K., and Shenoy, R.S. (2008) Overview of candidate device technologies for storage-class memory. *IBM J. Res. Dev.*, **52** (4/5), 449–464.
 7. Burr, G.W., Shenoy, R.S., Virwani, K., Narayanan, P., Padilla, A., Kurdi, B., and Hwang, H. (2014) Access devices for 3D crosspoint memory. *J. Vac. Sci. Technol., B*, **32** (4), 040 802.
 8. Flocke, A. and Noll, T.G. (2007) Fundamental analysis of resistive nano-crossbars for the use in hybrid nano/CMOS-memory. *ESSCIRC (European Solid State Circuits Conference)*, pp. 328–331.
 9. Flocke, A., Noll, T.G., Kugeler, C., Nauenheim, C., and Waser, R. (2008) A fundamental analysis of nano-crossbars with non-linear switching materials and its impact on TiO₂ as a resistive layer. *Nanotechnology, 2008. NANO'08. 8th IEEE Conference on*, IEEE, pp. 319–322.
 10. Liang, J. and Wong, H.S.P. (2010) Cross-point memory array without cell selectors – device characteristics and data storage pattern dependencies. *IEEE Trans. Electron Devices*, **57** (10), 2531–2538.
 11. Yu, S., Liang, J., Wu, Y., and Wong, H.S.P. (2010) Read/write schemes analysis for novel complementary resistive switches in passive crossbar memory arrays. *Nanotechnology*, **21**, 465 202.
 12. Chen, A. (2011) Accessibility of nano-crossbar arrays of resistive switching devices. *Nanotechnology (IEEE-NANO), 2011 11th IEEE Conference on*, IEEE, pp. 1767–1771.
 13. Chen, A. and Lin, M.R. (2011) Variability of resistive switching memories and its impact on crossbar array performance. *Reliability Physics Symposium (IRPS), 2011 IEEE International*, IEEE, pp. 843–846.
 14. Xu, C., Dong, X., Jouppi, N.P., and Xie, Y. (2011) Design implications of memristor-based RRAM cross-point structures. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1–6.
 15. Chen, A., Krivokapic, Z., and Lin, M.R. (2012) A comprehensive model for crossbar memory arrays. *Device Research Conference (DRC), 2012 70th Annual*, pp. 219–220.
 16. Liang, J., Yeh, S., Wong, S.S., and Wong, H.S.P. (2012) Scaling challenges for the cross-point resistive memory array to Sub-10nm node – an interconnect perspective. *IMW 2012*, pp. 61–64.
 17. Niu, D., Xu, C., Muralimanohar, N., Jouppi, N.P., and Xie, Y. (2012) Design trade-offs for high density cross-point resistive memory. *Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design*, ACM, pp. 209–214.
 18. Zhang, L., Cosemans, S., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2012) Analysis of the effect of cell parameters on the maximum RRAM array size considering both read and write. *ESSDERC*, pp. 282–285.
 19. Liang, J.L., Yeh, S., Wong, S.S., and Wong, H.S.P. (2013) Effect of word-line/bitline scaling on the performance, energy consumption, and reliability of cross-point memory. *ACM J. Emerging Technol. Comput. Syst.*, **9** (1), 9.
 20. Lo, C.L., Hou, T.H., Chen, M.C., and Huang, J.J. (2013) Dependence of read margin on pull-up schemes in high-density one selector–one resistor crossbar array. *IEEE Trans. Electron Devices*, **60** (1), 420–426.
 21. Chen, A. (2013) A comprehensive crossbar array model with solutions for line resistance and nonlinear device characteristics. *IEEE Trans. Electron Devices*, **60** (4), 1318–1326.
 22. International Technology Roadmap for Semiconductors (2013) Welcome to the ITRS! www.itrs.net (accessed 12 August 2015).

23. Narayanan, P., Burr, G.W., Shenoy, R.S., Stephens, S., Virwani, K., Padilla, A., Kurdi, B., and Gopalakrishnan, K. (2014) Exploring the design space for resistive nonvolatile memory crossbar arrays with mixed-ionic-electronic-conduction (MIEC)-based access devices. *Device Research Conference*, pp. V.–A5.
24. Narayanan, P., Burr, G.W., Shenoy, R., Virwani, K., and Kurdi, B. (2014) Circuit-level benchmarking of access devices for resistive nonvolatile memory arrays. *IEDM Technical Digest*, p. 29.7.
25. Wang, X., Fang, Z., Li, X., Chen, B., Gao, B., Kang, J., Chen, Z., Kamath, A., Shen, N., Singh, N., Lo, G., and Kwong, D. (2012) Highly compact 1T–1R architecture ($4F^2$ footprint) involving fully CMOS compatible vertical GAA nano-pillar transistors and oxide-based RRAM cells exhibiting excellent NVM properties and ultra-low power operation. *IEDM Technical Digest*, p. 20.6.
26. Wang, C.H., Tsai, Y.H., Lin, K.C., Chang, M.F., King, Y.C., Lin, C.J., Sheu, S.S., Chen, Y.S., Lee, H.Y., Chen, F.T., and Tsai, M.J. (2010) Three-dimensional $4F^2$ ReRAM cell with CMOS logic compatible process. *IEDM Technical Digest*, p. 29.6.
27. Servalli, G. (2009) A 45nm generation phase change memory technology. *IEDM Technical Digest*, p. 5.7.1.
28. Oh, J.H., Park, J.H., Lim, Y.S., Lim, H.S., Oh, Y.T., Kim, J.S., Shin, J.M., Park, J.H., Song, Y.J., Ryoo, K.C., Lim, D.W., Park, S.S., Kim, J.I., Kim, J.H., Yu, J., Yeung, F., Jeong, C.W., Kong, J.H., Kang, D.H., Koh, G.H., Jeong, G.T., Jeong, H.S., and Kim, K. (2006) Full integration of highly manufacturable 512Mb PRAM based on 90nm technology. *IEDM Technical Digest*, p. 2.6.
29. Zhang, C., Song, Z.T., Wu, G.P., Liu, B., Wan, X.D., Wang, L., Wang, L.H., Yang, Z.Y., Chen, B., and Feng, S.L. (2011) Design and fabrication of dual-trench epitaxial diode array for high-density phase-change memory. *IEEE Electron Device Lett.*, **32** (8), 1014–1016.
30. Lee, K.S., Yoo, D., Han, J.J., Hyung, Y.W., Kim, S.S., Kang, C.J., Jeong, H.S., Moon, J.T., Park, H., Jeong, H., Kim, K.R., and Choi, B. (2010) Selective epitaxial growth of silicon for vertical diode application. *Jpn. J. Appl. Phys., Part 1*, **49** (08), 08JF03.
31. An, H., Lee, K.S., Kang, Y., Jeong, S., Yoo, W., Han, J.J., Kim, B., Lim, H., Nam, S., Jeong, G.T., Kang, H.K., Chung, C., and Choi, B. (2012) Current–voltage characteristics of vertical diodes for next generation memories. *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, pp. 149–152.
32. Lee, K.S., Han, J.J., Lim, H., Nam, S., Chung, C., Jeong, H.S., Park, H., Jeong, H., and Choi, B. (2012) Cost-effective silicon vertical diode switch for next-generation memory devices. *IEEE Electron Device Lett.*, **33** (2), 242–244.
33. Sasago, Y., Kinoshita, M., Morikawa, T., Kurotsuchi, K., Hanzawa, S., Mine, T., Shima, A., Fujisaki, Y., Kume, H., Moriya, H., Takaura, N., and Torii, K. (2009) Cross-point phase change memory with $4F^2$ cell size driven by low-contact-resistivity poly-Si diode. *Symposium on VLSI Technology*, pp. T2B–1.
34. Herner, S.B., Bandyopadhyay, A., Dunton, S.V., Eckert, V., Gu, J., Hsia, K.J., Hu, S., Jahn, C., Kidwell, D., Konevecki, M., Mahajani, M., Park, K., Petti, C., Radigan, S.R., Raghuram, U., Vienna, J., and Vyvoda, M.A. (2004) Vertical p-i-n polysilicon diode with antifuse for stackable field-programmable ROM. *IEEE Electron Device Lett.*, **25** (5), 271–273.
35. Huet, K., Boniface, C., Negru, R., and Venturini, J. (2012) Ultra low thermal budget anneals for 3D memories: access device formation. *AIP Conference Proceedings*, vol. 1496, p. 135.
36. Ping, E.X., Erokhin, Y., Gossmann, H.J., and Khaja, F.A. (2012) Two-terminal diode steering element for 3D X-bar memory. *Junction Technology (IWJT), 2012 12th International Workshop on*, IEEE, pp. 24–28.

37. Song, Y.H., Park, S.Y., Lee, J.M., Yang, H.J., and Kil, G.H. (2011) Bidirectional two-terminal switching device for crossbar array architecture. *IEEE Electron Device Lett.*, **32** (8), 1023–1025.
38. Kil, G.H., Yang, H.J., Lee, G.H., Lee, S.H., and Song, Y.H. (2012) Bidirectional two-terminal switching device for non-volatile random access memory. *Jpn. J. Appl. Phys.*, **51** (4), 04DJ02.
39. Lee, M.H., Kao, C.Y., Yang, C.L., Chen, Y.S., Lee, H.Y., Chen, F., and Tsai, M.J. (2011) Reliability of ambipolar switching poly-Si diodes for cross-point memory applications. *Device Research Conference (DRC), 2011 69th Annual*, pp. 89–90.
40. Srinivasan, V.S.S., Chopra, S., Karkare, P., Bafna, P., Lashkare, S., Kumbhare, P., Kim, Y., Srinivasan, S., Kuppurao, S., Lodha, S., and Ganguly, U. (2012) Punchthrough-diode-based bipolar RRAM selector by Si epitaxy. *IEEE Electron Device Lett.*, **33** (10), 1396–1398.
41. Lashkare, S., Karkare, P., Bafna, P., Raju, M.V.S., Srinivasan, V.S.S., Lodha, S., Ganguly, U., Schulze, J., and Chopra, S. (2013) A bipolar RRAM selector with designable polarity dependent on-voltage asymmetry. *IMW 2013*, pp. 178–181.
42. Park, Y.S., Kil, G.H., and Song, Y.H. (2012) Bidirectional two-terminal switching device using Schottky barrier for Spin-Transfer-Torque magnetic random access memory. *Jpn. J. Appl. Phys.*, **51** (10), 106501.
43. Baek, I.G., Kim, D.C., Lee, M.J., Kim, H.J., Yim, E.K., Lee, M.S., Lee, J.E., Ahn, S.E., Seo, S., Lee, J.H., Park, J.C., Cha, Y.K., Park, S.O., Kim, S., Yoo, I.K., Chung, U., Moon, J.T., and Ryu, B.I. (2005) Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application. *IEDM Technical Digest*, pp. 750–753.
44. Lee, M.J., Seo, S., Kim, D.C., Ahn, S.E., Seo, D.H., Yoo, I.K., Baek, I.G., Kim, D.S., Byun, I.S., Kim, S.H., Hwang, I.R., Kim, J.S., Jeon, S.H., and Park, B.H. (2007) A low-temperature-grown oxide diode as a new switch element for high-density, nonvolatile memories. *Adv. Mater.*, **19** (1), 73–76.
45. Lee, M.J., Park, Y., Kang, B., Ahn, S., Lee, C., Kim, K., Xianyu, W., Stefanovich, G., Lee, J., Chung, S., Kim, Y., Lee, C., Park, J., Baek, I., and Yoo, I. (2007) Samsung 2-stack 1D–1R cross-point structure with oxide diodes as switch elements for high density resistance RAM applications. *IEDM Technical Digest*, p. 30.2.
46. Kang, B.S., Ahn, S.E., Lee, M.J., Stefanovich, G., Kim, K.H., Xianyu, W.X., Lee, C.B., Park, Y., Baek, I.G., and Park, B.H. (2008) High-current-density $\text{CuO}_x/\text{InZnO}_x$ thin-film diodes for cross-point memory applications. *Adv. Mater.*, **20**, 3066–3069.
47. Ahn, S.E., Kang, B.S., Kim, K.H., Lee, M.J., Lee, C.B., Stefanovich, G., Kim, C.J., and Park, Y. (2009) Stackable all-oxide based nonvolatile memory with Al_2O_3 antifuse and $p-\text{Cu}_x/n-\text{In-ZnO}_x$ diode. *IEEE Electron Device Lett.*, **30** (5), 550–552.
48. Tallarida, G., Huby, N., Kutrzeba-Kotowska, B., Spiga, S., Arcari, M., Csaba, G., Lugli, P., Redaelli, A., and Bez, R. (2009) Low temperature rectifying junctions for crossbar non-volatile memory devices. *International Memory Workshop*, 2009. IMW'09.
49. Huang, J.J., Kuo, C.W., Chang, W.C., and Hou, T.H. (2010) Transition of stable rectification to resistive-switching in $\text{Ti}/\text{TiO}_2/\text{Pt}$ oxide diode. *Appl. Phys. Lett.*, **96** (26), 262901.
50. Park, W.Y., Kim, G.H., Seok, J.Y., Kim, K.M., Song, S.J., Lee, M.H., and Hwang, C.S. (2010) A $\text{Pt}/\text{TiO}_2/\text{Ti}$ Schottky-type selection diode for alleviating the sneak current in resistance switching memory arrays. *Nanotechnology*, **21** (19), 195201.
51. Kim, G.H., Lee, J.H., Han, J.H., Song, S.J., Seok, J.Y., Yoon, J.H., Yoon, K.J., Lee, M.H., Park, T.J., and Hwang, C.S. (2012) Schottky diode with excellent performance for large integration density of crossbar resistive memory. *Appl. Phys. Lett.*, **100** (21), 213508.
52. Huang, J.J., Tseng, Y.M., Luo, W.C., Hsu, C.W., and Hou, T.H. (2011) One

- selector—one resistor (1S1R) crossbar array for high-density flexible memory applications. *IEDM Technical Digest*, p. 31.7.
53. Huang, J.J., Tseng, Y.M., Hsu, C.W., and Hou, T.H. (2011) Bipolar nonlinear Ni/TiO₂/Ni selector for 1S1R crossbar array applications. *IEEE Electron Device Lett.*, **32** (10), 1427–1429.
54. Ovshinsky, S.R. (1968) Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.*, **21** (20), 1450–1453.
55. Ielmini, D. (2008) Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses. *Phys. Rev. B*, **78** (3), 035 308.
56. Ielmini, D. and Zhang, Y.G. (2007) Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices. *J. Appl. Phys.*, **102** (5), 054 517.
57. Adler, D., Shur, M.S., Silver, M., and Ovshinsky, S.R. (1980) Threshold switching in chalcogenide-glass thin-films. *J. Appl. Phys.*, **51** (6), 3289–3309.
58. Anbarasu, M., Wimmer, M., Bruns, G., Salina, M., and Wuttig, M. (2012) Nanosecond threshold switching of GeTe₆ cells and their potential as selector devices. *Appl. Phys. Lett.*, **100** (14), 143 505.
59. Karpov, I.V., Mitra, M., Kau, D., Spadini, G., Kryukov, Y.A., and Karpov, V.G. (2008) Evidence of field induced nucleation in phase change memory. *Appl. Phys. Lett.*, **92** (17), 173 501A.
60. Wuttig, M. and Yamada, N. (2007) Phase-change materials for rewritable data storage. *Nat. Mater.*, **6** (11), 824–832.
61. Krebs, D., Raoux, S., Rettner, C.T., Burr, G.W., Salina, M., and Wuttig, M. (2009) Threshold field of phase change memory materials measured using phase change bridge devices. *Appl. Phys. Lett.*, **95** (8), 082 101.
62. Czubatyj, W. and Hudgens, S.J. (2012) Invited paper: thin-film ovonic threshold switch: its operation and application in modern integrated circuits. *Electron. Mater. Lett.*, **8** (2), 157–167.
63. Warren, A.C. (1973) Reversible thermal breakdown as a switching mechanism in chalcogenide glasses. *IEEE Trans. Electron Devices*, **20** (2), 123–131.
64. Pirovano, A., Lacaita, A.L., Benvenuti, A., Pellizzer, F., and Bez, R. (2004) Electronic switching in phase-change memories. *IEEE Trans. Electron Devices*, **51** (3), 452–459.
65. Emin, D. (2006) Current-driven threshold switching of a small polaron semiconductor to a metastable conductor. *Phys. Rev. B*, **74** (3), 035 206.
66. Karpov, I.V., Mitra, M., Kau, D., Spadini, G., Kryukov, Y.A., and Karpov, V.G. (2007) Fundamental drift of parameters in chalcogenide phase change memory. *J. Appl. Phys.*, **102** (12), 124 503.
67. Kau, D.C., Tang, S., Karpov, I.V., Dodge, R., Klehn, B., Kalb, J.A., Strand, J., Diaz, A., Leung, N., Wu, J., Lee, S., Langtry, T., Chang, K.W., Papagianni, C., Lee, J., Hirst, J., Erra, S., Flores, E., Righos, N., Castro, H., and Spadini, G. (2009) A stackable cross point phase change memory. *IEDM Technical Digest*, p. 27.1.
68. Lee, M.J., Lee, D., Kim, H., Choi, H.S., Park, J.B., Kim, H., Cha, Y.K., Chung, U.I., Yoo, I.K., and Kim, K. (2012) Highly-scalable threshold switching select device based on chalcogenide glasses for 3D nanoscale memory arrays. *IEDM Technical Digest*, p. 2.6.
69. Imada, M., Fujimori, A., and Tokura, Y. (1998) Metal–insulator transitions. *Rev. Mod. Phys.*, **70** (4), 1039–1263.
70. Son, M., Lee, J., Park, J., Shin, J., Choi, G., Jung, S., Lee, W., Kim, S., Park, S., and Hwang, H. (2011) Excellent selector characteristics of nanoscale VO₂ for high-density bipolar ReRAM applications. *IEEE Electron Device Lett.*, **32** (11), 1579–1581.
71. Son, M., Liu, X., Sadaf, S.M., Lee, D., Park, S., Lee, W., Kim, S., Park, J., Shin, J., Jung, S., Ham, M.H., and Hwang, H. (2012) Self-selective characteristics of nanoscale VO_x devices for high-density ReRAM applications. *IEEE Electron Device Lett.*, **33** (5), 718–720.
72. Chudnovskii, F.A., Odynets, L.L., Pergament, A.L., and Stefanovich, G.

- (1996) Electroforming and switching in oxides of transition metals: the role of metal–insulator transition in the switching mechanism. *J. Solid State Chem.*, **122** (1), 95–99.
73. Liu, X., Sadaf, S.M., Son, M., Shin, J., Park, J., Lee, J., Park, S., and Hwang, H. (2011) Diode-less bilayer oxide ($\text{WO}_x\text{--NbO}_x$) device for cross-point resistive memory applications. *Nanotechnology*, **22** (47), 475 702.
74. Kim, S., Liu, X., Park, J., Jung, S., Lee, W., Woo, J., Shin, J., Choi, G., Cho, C., Park, S., Lee, D.J., Cha, E., Lee, B.H., Lee, H., Kim, S., Jung, S., and Hwang, H. (2012) Ultrathin (<10nm) $\text{Nb}_2\text{O}_5/\text{NbO}_2$ hybrid memory with both memory and selector characteristics for high density 3D vertically stackable RRAM applications. *Symposium on VLSI Technology*, p. T18.3.
75. Ho, C., Huang, H.H., Lee, M.T., Hsu, C.L., Lai, T.Y., Chiu, W.C., Lee, M.Y., Chou, T.H., Yang, I., Chen, M.C., Wu, C.S., Chiang, K.H., Yao, Y.D., Hu, C., and Yang, F.L. (2012) Threshold vacuum switch (TVS) on 3D-stackable and 4F² cross-point bipolar and unipolar resistive random access memory. *IEDM Technical Digest*, p. 2.8.
76. Fazio, A. (2004) Flash memory scaling. *MRS Bull.*, **29** (11), 814–817.
77. Cimpoiasu, E., Tolpygo, S.K., Liu, X., Simonian, N., Kukens, J.E., Klie, R.F., Zhu, Y., and Likharev, K.K. (2004) Aluminum oxide layers as possible components for layered tunnel barrier. *J. Appl. Phys.*, **96** (2), 1088–1093.
78. Likharev, K.K. (1998) Layered tunnel barriers for nonvolatile memory devices. *Appl. Phys. Lett.*, **73** (15), 2137–2139.
79. Lee, W., Park, J., Shin, J., Woo, J., Kim, S., Choi, G., Jung, S., Park, S., Lee, D., Cha, E., Lee, H., Kim, S., Chung, S., and Hwang, H. (2012) Varistor-type bidirectional switch ($J_{MAX} > 10^7 \text{ A/cm}^2$, selectivity~ 10^4) for 3D bipolar resistive memory arrays. *Symposium on VLSI Technology*, p. T5.2.
80. Woo, J., Lee, W., Park, S., Kim, S., Lee, D., Choi, G., Cha, E., Lee, J.H., Jung, W.Y., Park, C.G., and Hwang, H. (2013) Multi-layer tunnel barrier ($\text{Ta}_2\text{O}_5/\text{TaO}_x/\text{TiO}_2$) engineering for bipolar RRAM selector applications. *Symposium on VLSI Technology*, p. T12.4.
81. Kawahara, A., Azuma, R., Ikeda, Y., Kawai, K., Katoh, Y., Hayakawa, Y., Tsuji, K., Yoneda, S., Himeno, A., Shimakawa, K., Takagi, T., Mikawa, T., and Aono, K. (2013) An 8 Mb multi-layered cross-point ReRAM macro with 443 Mb/s write throughput. *IEEE J. Solid-State Circuits*, **48** (1), 178–185.
82. Shin, J., Kim, I., Biju, K.P., Jo, M., Park, J., Lee, J., Jung, S., Lee, W., Kim, S., Park, S., and Hwang, H. (2011) TiO_2 -based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application. *J. Appl. Phys.*, **109** (3), 033 712.
83. Kim, S., Lee, W., and Hwang, H. (2012) Selector devices for cross-point ReRAM. *Cellular Nanoscale Networks and Their Applications (CNNA)*, 2012 13th International Workshop on, pp. 1–2.
84. Lee, W., Park, J., Kim, S., Woo, J., Shin, J., Choi, G., Park, S., Lee, D., Cha, E., Lee, B.H., and Hwang, H. (2012) High current density and non-linearity combination of selection device based on $\text{TaO}_x/\text{TiO}_2/\text{TaO}_x$ structure for one selector – one resistor arrays. *ACS Nano*, **6** (9), 8166–8172.
85. Woo, J., Lee, D., Cha, E., Lee, S., Park, S., and Hwang, H. (2013) Multilayer-oxide-based bidirectional cell selector device for cross-point resistive memory applications. *Appl. Phys. Lett.*, **103**, 202 113.
86. Gopalakrishnan, K., Shenoy, R.S., Rettner, C.T., Virwani, K., Bethune, D.S., Shelby, R.M., Burr, G.W., Kellock, A., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., Jackson, B., Friz, A.M., Topuria, T., Rice, P.M., and Kurdi, B.N. (2010) Highly-scalable novel access device based on Mixed Ionic Electronic Conduction (MIEC) materials for high density phase change memory (PCM) arrays. *Symposium on VLSI Technology*, p. 19.4.

87. Shenoy, R.S., Gopalakrishnan, K., Jackson, B., Virwani, K., Burr, G.W., Rettner, C.T., Padilla, A., Bethune, D.S., Shelby, R.M., Kellock, A.J., Breitwisch, M., Joseph, E.A., Dasaka, R., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., Friz, A.M., Topuria, T., Rice, P.M., and Kurdi, B.N. (2011) Endurance and scaling trends of novel access-devices for multi-layer crosspoint-memory based on Mixed Ionic Electronic Conduction (MIEC) materials. *Symposium on VLSI Technology*, pp. T5B-1.
88. Burr, G.W., Virwani, K., Shenoy, R.S., Padilla, A., BrightSky, M., Joseph, E.A., Lofaro, M., Kellock, A.J., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., Rettner, C.T., Jackson, B., Bethune, D.S., Shelby, R.M., Topuria, T., Arellano, N., Rice, P.M., Kurdi, B.N., and Gopalakrishnan, K. (2012) Large-scale (512kbit) integration of multilayer-ready access-devices based on Mixed-Ionic-Electronic-Conduction (MIEC) at 100% yield. *Symposium on VLSI Technology*, p. T5.4.
89. Virwani, K., Burr, G.W., Shenoy, R.S., Rettner, C.T., Padilla, A., Topuria, T., Rice, P.M., Ho, G., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., BrightSky, M., Joseph, E.A., Kellock, A.J., Arellano, N., Kurdi, B.N., and Gopalakrishnan, K. (2012) Sub-30nm scaling and high-speed operation of fully-confined access-devices for 3-D crosspoint memory based on Mixed-Ionic-Electronic-Conduction (MIEC) materials. *IEDM Technical Digest*, p. 2.7.
90. Burr, G.W., Virwani, K., Shenoy, R.S., Fraczak, G.H., Rettner, C.T., Padilla, A., King, R.S., Nguyen, K., Bowers, A.N., Jurich, M., BrightSky, M., Joseph, E.A., Kellock, A.J., Arellano, N., Kurdi, B.N., and Gopalakrishnan, K. (2013) Recovery dynamics and fast (sub-50ns) read operation with access devices for 3D crosspoint memory based on Mixed-Ionic-Electronic-Conduction (MIEC). *Symposium on VLSI Technology*, p. T6.4.
91. Padilla, A., Burr, G.W., Shenoy, R.S., Raman, K.V., Bethune, D.S., Shelby, R.M., Rettner, C.T., Mohammad, J., Virwani, K., Narayanan, P., Deb, A.K., Pandey, R.K., Bajaj, M., Murali, K.V.R.M., Kurdi, B.N., and Gopalakrishnan, K. (2015) On the origin of steep I-V nonlinearity in Mixed-Ionic-Electronic-Conduction-based access devices. *IEEE Trans. Electron Devices*, **62** (3), 963–971.
92. Linn, E., Rosezin, R., Kuegeler, C., and Waser, R. (2010) Complementary resistive switches for passive nanocrossbar memories. *Nat. Mater.*, **9** (5), 403–406.
93. Wouters, D.J., Zhang, L.Q., Fantini, A., Degraeve, R., Goux, L., Chen, Y.Y., Govoreanu, B., Kar, G.S., Groeseneken, G.V., and Jurczak, M. (2012) Analysis of complementary RRAM switching. *IEEE Electron Device Lett.*, **33** (8), 1186–1188.
94. Rosezin, R., Linn, E., Nielsen, L., Kugeler, C., Bruchhaus, R., and Waser, R. (2011) Integrated complementary resistive switches for passive high-density nanocrossbar arrays. *IEEE Electron Device Lett.*, **32** (2), 191–193.
95. Lee, J., Shin, J., Lee, D., Lee, W., Jung, S., Jo, M., Park, J., Biji, K., Kim, S., Park, S., and Hwang, H. (2010) Diode-less nano-scale ZrO_x/HfO_x RRAM device with excellent switching uniformity and reliability for high-density cross-point memory applications. *IEDM Technical Digest*, p. 19.5.
96. Lee, M.J., Lee, C.B., Lee, D., Lee, S.R., Chang, M., Hur, J.H., Kim, Y.B., Kim, C.J., Seo, D.H., Seo, S., Chung, U.I., Yoo, I.K., and Kim, K. (2011) A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures. *Nat. Mater.*, **10** (8), 625–630.
97. Nardi, F., Balatti, S., Larentis, S., and Ielmini, D. (2011) Complementary switching in metal oxides: toward diode-less crossbar RRAMs. *IEDM Technical Digest*, p. 31.1.
98. Yang, Y., Sheridan, P., and Lu, W. (2012) Complementary resistive switching in tantalum oxide-based resistive memory devices. *Appl. Phys. Lett.*, **100** (20), 203112.

99. Yang, Y., Choi, S.H., and Lu, W. (2013) Oxide heterostructure resistive memory. *Nano Lett.*, **13**, 2908–2915.
100. Tang, G., Zeng, F., Chen, C., Liu, H., Gao, S., Song, C., Lin, Y., Chen, G., and Pan, F. (2013) Programmable complementary resistive switching behaviors of a plasma-oxidized titanium oxide nanolayer. *Nanoscale*, **5**, 422–428.
101. Meyer, R., Schloss, L., Brewer, J., Lambertson, R., Kinney, W., Sanchez, J., and Rinerson, D. (2008) Oxide dual-layer memory element for scalable non-volatile cross-point memory technology. *NVMTS 2008*.
102. Chevallier, C.J., Siau, C.H., Lim, S.F., Namala, S.R., Matsuoka, M., Bateman, B.L., and Rinerson, D. (2010) A $0.13\mu\text{m}$ 64Mb multi-layered conductive metal-oxide memory. *ISSCC 2010*, p. 14.3.
103. Chevallier, C. (2011) Conductive metal oxide arrays for terabit memories. *2011 Leti workshop on Innovative Memory Technologies*, p. 15.
104. Chevallier, C. (2012) RRAM review. *ITRS workshop on Storage Class Memory*.
105. Jo, S.H. and Lu, W. (2008) CMOS compatible nanoscale nonvolatile resistance switching memory. *Nano Lett.*, **8** (2), 392–397.
106. Kim, K.H., Jo, S.H., Gaba, S., and Lu, W. (2010) Nanoscale resistive memory with intrinsic diode characteristics and long endurance. *Appl. Phys. Lett.*, **96**, 053 106.
107. Zuo, Q., Long, S., Yang, S., Liu, Q., Shao, L., Wang, Q., Zhang, S., Li, Y., Wang, Y., and Liu, M. (2010) ZrO_2 -based memory cell with a self-rectifying effect for crossbar WORM memory application. *IEEE Electron Device Lett.*, **31** (4), 344–346.
108. Yang, M., Hu, P., Lu, J.Q., Lv, Q.B., and Li, S.W. (2011) Coupled interfaces for misreading avoidance and write current reduction in passive crossbar memory. *Appl. Phys. Lett.*, **98** (21), 213 501.
109. Tran, X., Gao, B., Kang, J., Wu, X., Wu, L., Fang, Z., Wang, Z., Pey, K., Wang, X., Yeo, Y., Du, A., Nguyen, B., Li, M., and Yu, H. (2011) Self-rectifying and forming-free unipolar HfO_x based high performance RRAM built by fab-available materials. *IEDM Technical Digest*, p. 31.2.
110. Huang, Y.L., Huang, R., Pan, Y., Zhang, L.J., Cai, Y.M., Yang, G.Y., and Wang, Y.Y. (2012) A new dynamic selector based on the bipolar RRAM for the crossbar array application. *IEEE Trans. Electron Devices*, **59** (8), 2277–2280.
111. Lee, H., Kim, S., Cho, K., Hwang, H., Choi, H., Lee, J., Lee, S., Lee, H., Suh, J., Chung, S.O., Kim, Y., Kim, K., Nam, W., Cheong, J., Kim, J., Chae, S., Hwang, E.R., Park, S., Sohn, Y., Lee, C., Shin, H., Lee, K., Hong, K., Jeong, H., Rho, K., Kim, Y., Chung, S., Nickel, J., Yang, J., Cho, H., Perner, F., Williams, R., Lee, J., Park, S., and Hong, S.J. (2012) Integration of 4F^2 selector-less crossbar array 2Mb ReRAM based on transition metal oxides for high density memory applications. *Symposium on VLSI Technology*, p. T18.1.
112. Park, S.G., Yang, M., Ju, H., Seong, D.J., Lee, J., Kim, E., Jung, S., Zhang, L., Shin, Y., Baek, I.G., Choi, J., Kang, H.K., and Chung, C. (2012) A non-linear ReRAM cell with sub- $1\mu\text{A}$ ultralow operating current for high density vertical resistive memory (VRAM). *IEDM Technical Digest*, p. 20.8.
113. Lentz, F., Roesgen, B., Rana, V., Wouters, D.J., and Waser, R. (2013) Current compliance-dependent nonlinearity in TiO_2 ReRAM. *IEEE Electron Device Lett.*, **34** (8), 996–998.
114. Yang, J.J., Zhang, M.X., Pickett, M.D., Miao, F., Strachan, J.P., Li, W.D., Yi, W., Ohlberg, D.A.A., Choi, B.J., Wu, W., Nickel, J.H., Medeiros-Ribeiro, G., and Williams, R.S. (2012) Engineering non-linearity into memristors for passive crossbar applications. *Appl. Phys. Lett.*, **100** (11), 113 501.

23

Bottom-Up Approaches for Resistive Switching Memories

Sabina Spiga, Takeshi Yanagida, and Tomoji Kawai

23.1

Introduction

Resistive switching effects in metal–insulator–metal (MIM) structures are currently widely investigated for their potential applications in future nonvolatile storage, reconfigurable logic, and non-Boolean neuromorphic computing [1–6]. In particular, resistive switching memory (Resistive random access memory (ReRAM)) is considered today one of the most promising solution for ultra high-density mass storage to encompass the current limitation of NAND FLASH technology beyond the 16-nm technology node. Additional incentives for developing ReRAM technologies include their low cost-per-bit, low-voltage (and low-power) operation, high endurance, and suitability for integration in crossbar arrays stacked in multiple levels in 3D [1–6]. A ReRAM cell consists of a two terminal device, that is, a capacitor structure where the active material (such as an oxide or solid electrolyte) is sandwiched between two metal electrodes, and whose resistance can be reversibly switched between two or more different resistance states. The filamentary mechanism is the prevailing model for a broad class of resistive switching devices [6–8] and relies on the formation and rupture of conductive filaments (CFs), with dimension that can be decreased down to tens or even few nanometers.

Recently, bottom-up fabrication approaches for ultrascaled ReRAM have been proposed not only for exploring the fundamental nanoscale resistive switching phenomena but also for creating novel nanoscale devices, which cannot be easily achieved by using current state-of-the-art lithography techniques in a top-down approach [9–15]. Ultrasmall ReRAM nanodevices can be fabricated by exploiting the functionalities of nanodots, nanojunctions, and nanowires (NWs), synthesized by bottom-up methods, as building blocks for new devices. Indeed, by bottom-up approaches, it is possible to grow and assemble nanostructures with distinct chemical composition, structure, size, and morphology in order to build specific functional devices [16–20].

Since the NW structure can be relatively easily connected to metal electrodes when compared with nanodots, various applications toward nanoscale

electronics, including ReRAM, have been mainly demonstrated by using NWs [9–17, 20–23], even if nanodots or nanopillars have been also considered by some authors [24–26]. Nevertheless, it is still a key issue to improve the ability to synthesize and manipulate nano-objects on large areas [27, 28] and with controlled electrical and physical properties, in order to enable the fabrication of nanoscale devices and to address their future industrial applications. For this reason, bottom-up approaches can be combined with template-assisted fabrication and top-down fabrication methods to achieve ordered large arrays of ReRAM devices with the crossbar or 3D configurations. For instance, many bottom-up ReRAMs have been fabricated by taking advantage of the porous anodic aluminum oxide (AAO) membrane as template, or more recently by using emerging approaches such as nanosphere lithography (NSL) or self-assembly of block copolymers (BCPs) [15, 24, 29–33].

This chapter reviews the proposed bottom-up approaches for the fabrication of ReRAM devices, their resistive switching properties, and possible array configurations. We present the resistive switching properties and mechanism of ReRAM cells based on single all-oxide NWs and nanodots, segmented MIM NWs, and core/shell NWs. Finally, emerging bottom-up approaches and applications for nano-ReRAM are discussed.

23.2

Bottom-Up ReRAM Fabrication Methods

NWs or nanodots/nanopillars as building block of ReRAM were fabricated by various bottom-up approaches [5, 24–26, 34–38]. The most widely used techniques in the framework of resistive switching devices are:

- 1) Vapor–liquid–solid (VLS) growth of one-dimensional NW [39–41] via metal catalysts, eventually followed by an heterostructure formation technique in the radial or axial direction to form segmented or core/shell NWs, respectively.
- 2) Template-assisted fabrication of array of NWs by electrodeposition of metals into the pores of an AAO template followed by thermal oxidation or sulfurization to form single all-oxide NWs or segmented axial heterostructured NWs (i.e., metal oxides, $\text{Ag}_2\text{S}/\text{Ag}$, $\text{Ni}-\text{Ag}_2\text{Se}-\text{Ni}$) [12, 15, 42–44].

In addition, NWs fabricated by VLS or template methods can be released by the matrix and rearranged to form ReRAM devices also in a crossbar configuration [13, 45, 46].

23.2.1

Vapor–Liquid–Solid Method

The **VLS** method using a metal catalyst [39–41] allows to define the size and spatial position of NWs by defining those of the metal catalysts. In addition, the

heterostructures along axial or radial directions can be sequentially designed by using VLS process. These unique and fascinating features of VLS methods are hardly attainable by other NW fabrication methods. Metal oxides as typical ReRAM materials can be formed as NWs via VLS route [10, 11, 34, 47, 48], following two main approaches. The first one relies on the direct synthesis of oxide NWs by VLS to provide single crystalline NW-based ReRAM, allowing the study of the nanoscale resistive switching nature in single crystalline NW without grain boundaries. However, there is a limitation of materials that can be directly fabricated. The typical oxide NWs for ReRAM, which can be fabricated by VLS, comprises materials such as ZnO, SnO₂, Ga₂O₃, and In₂O₃ [34, 36, 45, 48, 49]. Other oxide materials, such as NiO, which is of great interest for unipolar ReRAM, are hardly formed as VLS NWs within the framework of existing direct VLS methods. To overcome this issue, Oka *et al.* fabricated heterostructured NWs consisting of MgO(core)/NiO(shell) [34] by using *in situ* heterostructure formation method without exposure to atmospheric atmosphere [34, 50]. First, MgO NWs were formed via VLS route with Au metal catalysts. Then, the single crystalline NiO shell layer was grown by altering the crystal growth mode from VLS to vapor–solid (VS) growth [34, 50]. Since both MgO and NiO are rock-salt crystal structures and their lattice constants are also close enough, the NW heterostructure shows the epitaxial single crystalline shell layer of NiO. This approach is quite interesting since it allows growing nanostructures with controlled structure, where the size can be easily controlled by varying the oxide core-NW diameter and the oxide shell thickness. Other oxide NWs for ReRAM have been fabricated by utilizing this core–shell NW strategy, including CoO_x and TiO_x [11, 47, 49, 51]. However, such ideal material combinations are rather limited since, in general, the crystal structures and the lattice constants of oxides are quite different [50]. Moreover, in this approach, the core oxide serves as a template for the shell growth and cannot be used as electrode or interconnection line as proposed for crossbar architectures demonstrated by other core/shell approaches to form a metal(core)/shell(oxide) [13, 45, 46].

23.2.2

Template-Assisted Fabrication Methods of NWs

Ordered arrays of oxides, metals, carbon nanotube (CNT) NWs, and metal–oxides or metal–CNT heterojunction NWs can be fabricated by using a porous AAO membrane as template [12, 15, 52–55]. Indeed, in the last years an increasing attention has been devoted to the use of AAO templates for nanostructure fabrication, due to the possibility of controlling the pore density (from 10¹⁰ to 10¹² cm⁻²) and pore diameter in the nanometric range (from 10 to 200 nm) by anodization techniques, and obtaining regular (hexagonal) arrangements of the pores. A regular arrangement of pores in AAO templates can be achieved by means of two different techniques: multiple self-organizing anodization [56] and pattern-guided anodization [57, 58]. Exhaustive literature reviews on AAO templates fabrication and NWs electrodeposition can be

found in Refs. [59, 60]. The peculiar structure of AAO templates makes them suitable for investigating high-density and nanoscaled resistive and rectifying elements, fabricated using different materials (oxide, metal, and carbon) and deposition techniques, such as electrolytic and chemical vapor deposition (CVD)/atomic layer deposition (ALD) methods. All-oxide NWs, metal–oxide (MO) heterojunctions, and segmented metal–oxide–metal (MOM) NWs for ReRAM applications have been synthesized by combining electrodeposition and selective oxidation into the nanopores of AAO [12, 15, 53–55, 61–64]. Even if the latter method can be extended to other systems, it is limited to those oxides whose precursor metal can be electrodeposited easily, and selectively oxidized by thermal treatment. Alternatively, the synthesis of the oxide layer can be obtained directly by electrodeposition using proper aqueous solution [63]. We should also mention that the AAO templates have been alternatively used as shadow mask to form regular metal nanodot array defining the nanoscale top electrode of a ReRAM device [26, 65].

Finally, other methods that have been proposed to fabricate NWs or nanodots for ReRAM devices, even if less explored up to know, are those based on solution methods exploited for Cu_{2-x}Se [66], NSL, and block-copolymer self-assembly [24, 25, 31, 38].

23.3

Resistive Switching in Single (All-Oxide) NW/Nanoisland ReRAM

Resistive switching in single all-oxide NWs has been reported by various authors and mainly based on NiO and ZnO materials [9, 12, 15, 36], with NW length up to tens of micrometers and with NW diameter in the 20–200 nm range. Furthermore, radial core/shell heterostructures such as MgO/NiO, MgO/CoO_x have been proposed to achieve single crystal oxide NW and to engineering the oxide structure [10, 11]. ReRAM devices based on all-oxide NWs as well as oxide nanodots/nanopillar have been fabricated in combination with various passive or active electrodes to exploit switching mechanisms based on thermal effects [7] as well as cation transport, as in electrochemical metallization (ECM) cell, or anion transport, as in valence change memory (VCM) [1–3, 6].

The fabricated ReRAM devices based on single NWs share the common feature that electrodes are patterned by electron beam or optical lithography after dispersion of the NWs onto a substrate (typically oxidized SiO₂). Then, switching properties are analyzed through a long oxide segment (i.e., the distance between the two parallel electrodes), from hundreds of nanometers to several micrometers. This approach often leads to high forming voltage as well as set/reset voltages above 10 V, values that are not useful for memory circuit implementation, even if low switching current down to <10 nA in NWs has been demonstrated. Furthermore, this easy configuration has the advantage of assessing resistive switching properties in nanostructures and to check variations with respect to thin films by confining the switching operation in a small region, defined by the NW diameter

and/or by the shell film thickness. Furthermore, this planar structure allows to easily investigate switching mechanisms also by dedicated experiments with multiprobes and/or under selected atmosphere control [47, 49].

23.3.1

Resistive Switching in Single NiO NWs and Nanoislands

Resistive switching in NW ReRAM has been first demonstrated for the NiO-based system [12, 34]. The interest in this material was driven by the results achieved for thin-solid film-based ReRAM devices, showing unipolar type of switching [1, 7]. Indeed, after the preliminary results achieved in the 1960s [67] there was a renewed interest in the 2000s [68] toward the implementation of NiO ReRAM for crosspoint high-density nonvolatile memories.

In the work by Kim *et al.* [12], arrays of polycrystalline cubic and vertically aligned NiO NWs with 70 nm of diameter and maximum length of 25 μm were prepared by electrodeposition of Ni into the pores of an AAO membrane, followed by full oxidation of the Ni at 450 °C in air (Figure 23.1). In order to

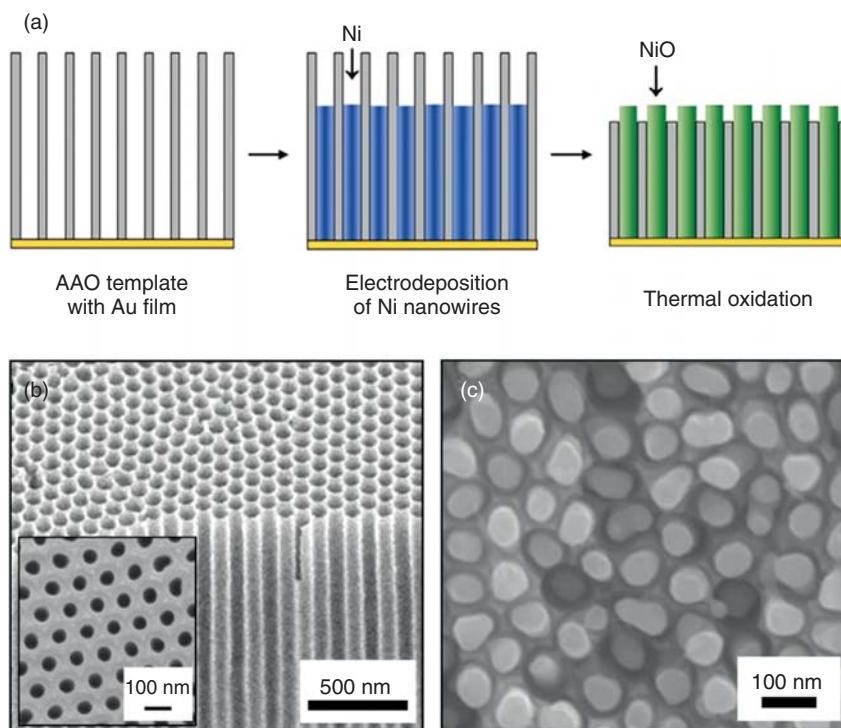


Figure 23.1 (a) Schematic of the NiO NW array fabrication procedure. (b) SEM image of a fabricated AAO membrane and (c) of the protruded NiO NWs above the AAO membrane. (Reprinted with permission from Ref. [12], © 2008, AIP Publishing LLC.)

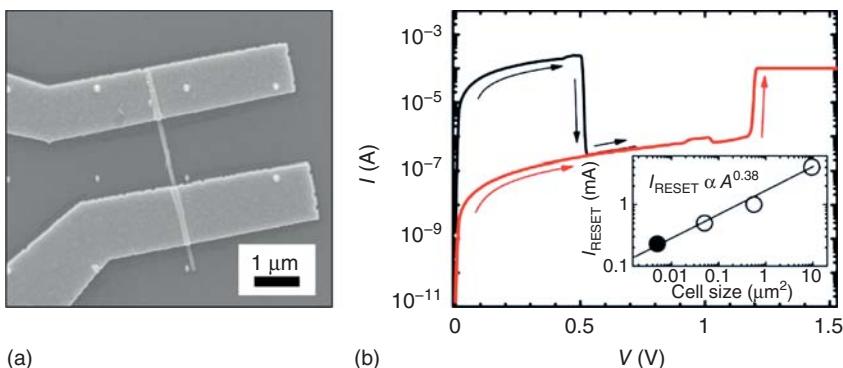


Figure 23.2 (a) SEM image of an individual NiO NW contacted with Au/Ti electrodes. (b) I - V curves of the NiO NW ReRAM device showing set and reset transition. (Reprinted with permission from Ref. [12], © 2008, AIP Publishing LLC.)

characterize the resistive switching of an individual NW ReRAM device, the NiO NWs were released by the matrix and dispersed on an oxidized silicon substrate. Then, parallel Au/Ti electrodes were patterned on the NW with a spacing of 1 μm by electron beam lithography and lift-off techniques (Figure 23.2a). The device is initially in a high-resistance state; by a forming process at voltage V_{FORM} of about 2.5 V, the device switches from the initial high-resistance state (HRS) to a low-resistance state (LRS) by electrically induced growth of a CF. HRS and LRS are also called ON and OFF states, respectively. Figure 23.2b shows the reset (black curve, switch from LRS to HRS) and set (red curve, switch from high to LRS) transitions, which correspond, respectively, to the dissolution and formation of a CF inside the NiO NW [12]. The reset transition occurs at a reset voltage (V_{RESET}) of 0.5 V and reset current (I_{RESET}) of 230 μA, while the set voltage (V_{SET}) is around 1.3 V. It should be noted that the forming, set and reset voltages of 1 μm long NiO NW are quite similar to MIM devices with typical NiO thickness in the 10–50 nm range [1, 7, 68, 69]. This finding can be understood by taking into account the high concentration of defects and the large number of grain boundaries in the NW microstructure, as evidenced by transmission electron microscopy (TEM) [12]. In addition, Ni defects could be another origin of a low V_{FORM} since it was reported that a decrease in defect concentration in NiO films results in an increase of switching voltages [70].

The effect of the microstructure on resistive switching is also evidenced by the experiment of Oka *et al.* [34] who measured NiO single-crystalline heterostructured NW fabricated in a radial core (10 nm diameter MgO)/shell (10 nm thick NiO) structure (Figure 23.3a). To measure the switching properties, the NiO NWs were deposited onto an oxidized Si substrate, then a Pt electrode was deposited as one of the two contacts. Finally, the conductive tip of an atomic force microscopy (C-AFM) was used as second electrode and positioned at 1 μm of distance from the Pt. In this configuration, the current flows through the cross-sectional area of the NiO shell layer of about $10^{-3} \mu\text{m}^2$. The device exhibits bipolar resistive switching

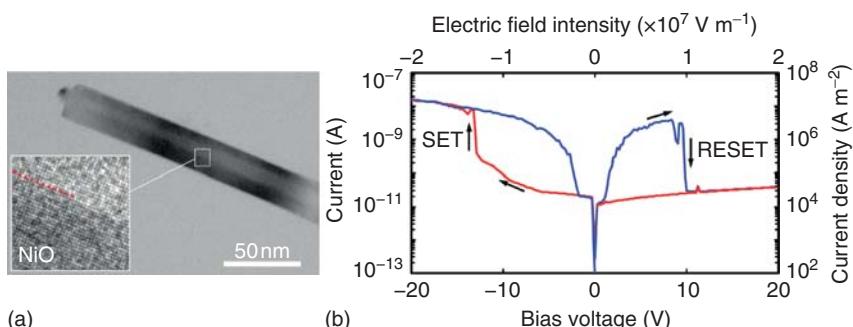


Figure 23.3 (a) TEM image of MgO(core)/NiO(shell) heterostructured NW. (b) Measured I - V curves for a heterostructures NiO NW obtained by C-AFM. (Reprinted and adapted with permission from Ref. [34]. © 2009 American Chemical Society.)

with relatively large set and reset voltages in the range between 10 and 15 V and switching currents below 10 nA (Figure 23.3b). Although the low switching currents are interesting for applications, the switching voltages are exceeding the requirements for practical use in memory circuits. It should be also noted that the large voltages are mainly due to the micrometer length of the oxide layer used in this work. However, the oxide microstructure could play a role in the switching properties since, considering the same NW length of 1 μm , the single crystalline NiO NWs exhibit a bipolar switching at high voltages [34], while the polycrystalline and highly defective NiO NWs show unipolar switching at low voltages [12].

Various attempts were used to improve the switching properties of single NiO NWs by engineering the structure of the NW and by reducing the length of the active oxide between the two electrodes down to hundreds or tens of nanometers (see also Section 23.4). Huang *et al.* [71] demonstrated that it is possible to reduce the switching voltages and narrow the switching distributions by using multilayered NiO/Pt NWs instead of monolithic polycrystalline NiO NWs (Figure 23.4). The NWs were synthesized with a similar approach used in [12] by cyclic electroplating of Ni/Pt in AAO template with a mean pore diameter of 70 nm, followed by *in situ* oxidation of the Ni segment into NiO at 800 °C for 6 h in air. The length of the NiO segment was 50 or 100 nm, while the thickness of Pt and the total length of NWs were fixed at 10 nm and 6 μm , respectively. The TEM images of polycrystalline monolithic NiO (used as reference) and multilayered NiO/Pt NWs with different NiO thicknesses (t_{NiO}) in each segment are shown in Figures 23.4a–c. The resistive switching properties were measured for NW arrays by using large area top electrodes of 200 μm in diameter and without dissolving the AAO matrix. After a forming process at 15 V and current compliance of 0.5 mA, the measured devices exhibited nonpolar resistive switching (coexistence of unipolar and bipolar resistive switching). Figure 23.4d shows the cumulative probabilities of V_{SET} and V_{RESET} for each sample. For the multilayered structures, lower switching voltages and narrowed distributions are obtained;

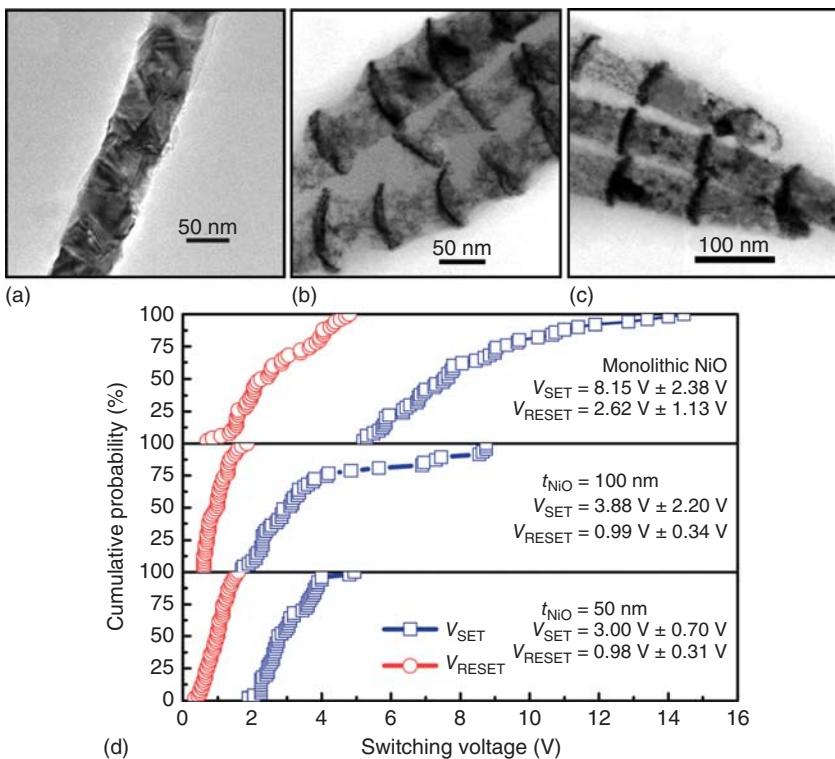


Figure 23.4 TEM images of polycrystalline monolithic NiO (a) and multilayered NiO/Pt NWs with different NiO thicknesses (t_{NiO}) of 50 nm (b) and 100 nm (c). Cumulative

probabilities of V_{SET} and V_{RESET} for each sample (d). (Reprinted with permission from Ref. [71], © 2012, AIP Publishing LLC.)

furthermore, with reducing t_{NiO} , the average V_{SET} and V_{RESET} as well as the switching voltage distributions are decreased in both unipolar and bipolar operations. As a result, the average operating electrical field of multilayered NWs ($\sim 0.5 \text{ MV m}^{-1}$) is much lower than the single crystalline NiO NW case ($\sim 10 \text{ MV m}^{-1}$) in Figure 23.3. Although the achieved results are interesting for applications, the switching currents, especially reset currents above 10^{-4} A , should be further reduced for implementation in memory circuits.

An interesting approach to reduce V_{SET} and V_{RESET} voltages as well as to achieve low switching currents was presented by Son *et al.* [26] by fabricating NiO ReRAM capacitor array structure on graphene sheet as bottom electrode. NiO/Pt nanopillars, having diameter of about 30 nm and height of about 33 nm for NiO and 5 nm for Pt, were fabricated by using an AAO membrane with thickness of 400 nm as shadow mask, while NiO and Pt were subsequently deposited by radio frequency (RF) sputtering. To measure electrical data, the conducting Pt tip of AFM was used to contact the Pt top electrode of the NiO ReRAM nanocapacitor. The devices exhibited typical unipolar ReRAM switching behaviors with a very low

ON current below 1 nA. The V_{SET} was about 1.5 V, and the V_{RESET} was about 0.5 V. The authors claims that the achieved results might be related also to the work function of graphene (4.66 eV), which is lower than that of noble metal such as Pt, and by the formation of a well-defined Schottky contact between NiO and graphene sheet. The dependence of resistive switching on the electrode material has also been reported for thin films [72, 73].

Finally, recently Kim *et al.* [74] demonstrated the possibility to integrate on a silicon substrate ordered NiO NW array exhibiting unipolar switching and fabricated by the AAO template method. The 400 nm thick AAO membrane was achieved following the double anodization process by starting from a Si/Ti/Au(50 nm)/Al(1.3 μm) substrate. This is an interesting approach compared with most of the work relying on free standing AAO membranes and/or isolated NW dispersed in oxidized Si surface, and it could represent a step forward to the use of bottom-up approaches for devices integrated on silicon. In the same work, the authors by comparing AAO membrane with different pore size (from 30 to 200 nm) and different number of NWs in the array (by defining electrodes of different area) demonstrated that the scaling behavior of NiO NWs is different from thin films. In particular, by considering the relationship $I_{\text{RESET}} \propto R_0^{-\alpha}$, where R_0 is the initial ON resistance, also observed for NiO thin films [75], the authors reported that the α exponent increases by decreasing the NW diameter. This finding might be explained by taking into account the increased localization of the conducting filaments into the NW.

23.3.2

Resistive Switching in Oxide NWs Alternative to NiO

Although NiO has been the active oxide mainly investigated to study the unipolar or bipolar type of switching in NWs and nanoislands, NW and nanoisland ReRAMs fabricated by similar methods but using alternative oxide materials, such as CoO_x , TiO_x , ZnO , and SiO_x [11, 24, 25, 35–37, 48, 49, 51, 76] have been also investigated.

Figure 23.5a shows the scanning electron microscopy (SEM) image of a $\text{MgO}(\text{core})/\text{Co}_3\text{O}_4(\text{shell})$ NW ReRAM device fabricated on Si substrate [11]. The core–shell NWs were grown by an in situ NW template method [11, 34, 50]. The diameter and length of MgO NWs were, respectively, 10 nm and 5 μm , while the Co_3O_4 shell was 5 nm thick. The single oxide NW was electrically connected by Pt electrodes to avoid detrimental interfacial oxidized layers. Figure 23.5b shows the typical I - V characteristics of the fabricated NW ReRAM. The device exhibits a bipolar resistive switching behavior, where the resistance change can be operated by applying positive/negative electric voltage. Although the device size is extremely small (the cross-sectional area of the Co_3O_4 shell layer is $10^{-4} \mu\text{m}^2$), the resistive switching could be operated for many cycles at least up to 10^8 times, as shown in Figure 23.5c. The operating voltage for set/reset operation is large, mainly due to the long gap distance between electrodes (~ 250 nm). Furthermore, the corresponding electric field (Figure 23.5b)

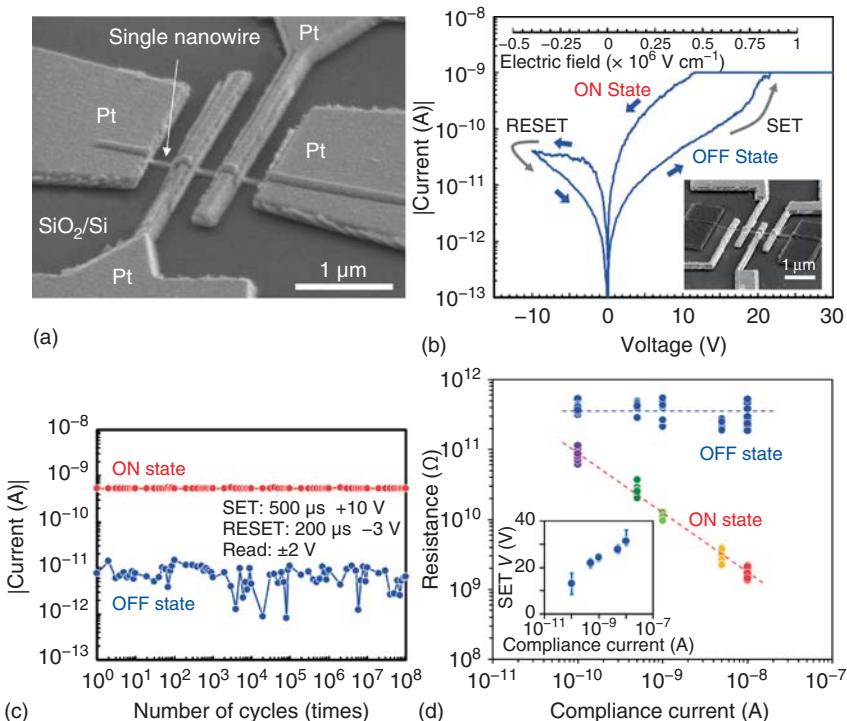


Figure 23.5 (a) SEM image of NW ReRAM device on Si substrate ($\text{MgO}(\text{core})/\text{Co}_3\text{O}_4(\text{shell})$ heterostructured NW). Typical I - V data (b) and endurance data (c) of the fabricated NW ReRAM device. (d) Resistance values of the ON and OFF

states (read at 5 V) as a function of the compliance current, demonstrating a multi-memory operation. (Adapted with permission from Ref. [11]. © 2010, American Chemical Society.)

is comparable to that of thin film devices [11]. It is also worth noticing that the bipolar switching in these devices can be operated at low current values, less than 10^{-9} A. Based on the filamentary model [2, 6–8], the LRS resistance is determined by the number and/or size of nanoscale CFs. Therefore, the low current values for the $\text{MgO}/\text{Co}_3\text{O}_4$ core/shell NW ReRAM could be related to the confinement effect of conducting filaments within the thin (~5 nm) shell layer. The effect of confinement of resistive switching in NWs has been discussed also by other authors for the NiO NW system [74]. In addition, the multi-memory operation can be performed by controlling the compliance current during set operation, as seen in Figure 23.5d. The resistance values of the ON and OFF states were read at 5 V, and the same reset voltage of -20 V was applied in each measurements. The ON state resistance decreases as a function of compliance current, and the ON/OFF resistance ratio can be modulated from 3 to 300 by varying the compliance current from 10^{-10} to 10^{-8} A, demonstrating that it is possible to achieve a multilevel programming. Thus, these data proved the

excellent nonvolatile memory properties of bipolar resistive switching at 10 nm size scale.

23.3.3

Study of Switching Mechanisms in Oxide NW ReRAM

One of the unique features of planar-type NW ReRAM devices (Figure 23.5a) is that we can directly extract the intrinsic properties of ReRAM, which have been buried within conventional capacitor-type ReRAM. Figure 23.6 shows the features of the planar-type NW ReRAM devices when compared with conventional capacitor-type devices. Three different approaches can be used to study switching mechanisms and properties in NW ReRAM:

- (i) Multiprobe measurements, which enable to identify the active switching location of resistive switching,
- (ii) Controlled-atmosphere measurements intentionally utilizing the chemical reactions with surroundings to extract the characteristics of conduction paths, and
- (iii) Planar field effect transistor (FET) configuration allowing to identify the carrier type of nanoscale conduction paths.

The latter approach is shown in Figure 23.7 for a $\text{MgO}(\text{core})/\text{CoO}_x$ shell NW ReRAM to identify the carrier type of the LRS conduction [47]. It is worth

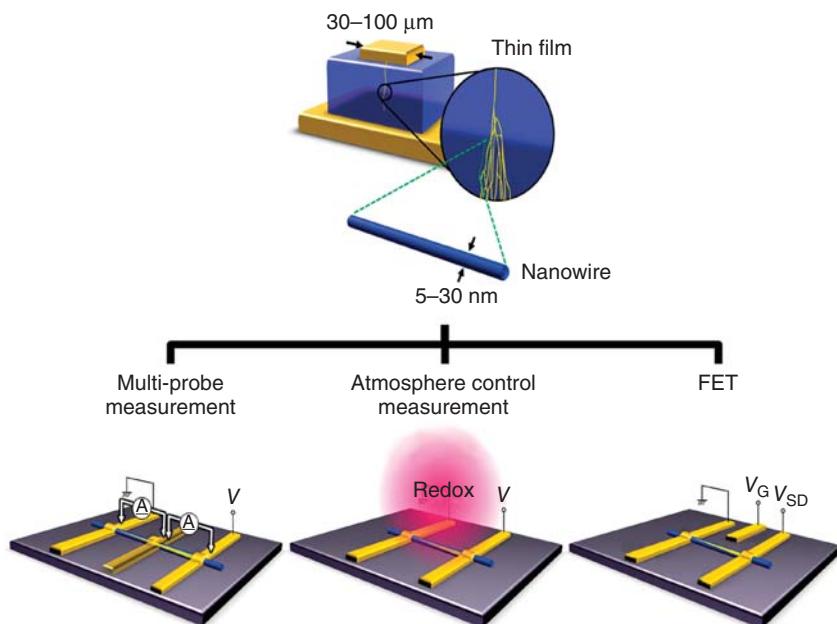


Figure 23.6 Unique features of planar-type NW ReRAM devices when compared with thin films capacitor-type ReRAM devices.

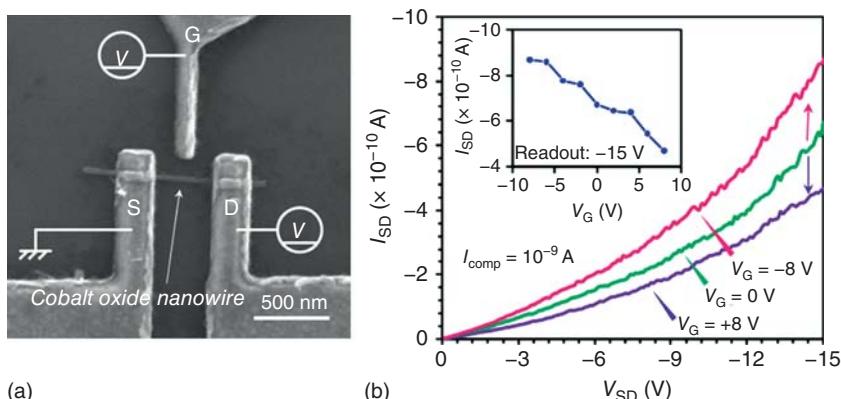


Figure 23.7 Identification of mobile carrier type by utilizing planar-type NW ReRAM devices comprised of CoO_x NW. (a) SEM image of the planar FET structure composed of source (S), drain (D), and gate (G) electrodes. The nanogap spacing of S–D (i.e., channel length) is $\sim 400 \text{ nm}$. (b) I_{SD} - V_{SD}

characteristics of the cobalt oxide NW when varying the gate voltage (V_G) ranged from -8 to $+8 \text{ V}$. The inset shows the V_G dependence of I_{SD} measured at -15 V . (Reprinted with permission from Ref. [47]. © 2011, American Chemical Society.)

noticing that such carrier type identification has been impossible up to now in conventional capacitor-type devices. In the FET experiments, a negative voltage was applied to the drain while the source was grounded. After performing the set process to switch the device to the LRS, a voltage bias was applied between the gate and the source electrodes. Figure 23.7b shows the I_{SD} - V_{SD} curves of the CoO_x NWs when varying the gate voltages (V_G), where I_{SD} and V_{SD} are the current and voltage between drain and source, respectively. I_{SD} decreased when applying the positive value $V_G = +8 \text{ V}$, whereas I_{SD} increased for $V_G = -8 \text{ V}$. Further, the I_{SD} measured at $V_{SD} = -15 \text{ V}$ systematically decreased with increasing V_G (inset of Figure 23.7b). These results indicate that the majority carriers of the LRS conduction in a CoO_x NW are holes (i.e., p-type). [47]. Based on these results, Nagashima *et al.* described a model for resistive switching in CoO_x NW based on p-type conduction with cation vacancies [47]. The set and reset processes are accomplished by, respectively, the formation and dissolution of a p-type conduction path consisting of cation vacancies [47]. The carrier type of bipolar resistive switching in CoO_x NW was therefore found to be consistent with the carrier type of bulk materials [47].

The same authors examined the active switching location of resistive switching in CoO_x NWs, by measuring the resistance between multi-electrodes. Figure 23.8 shows the experimental configuration based on three electrodes A, B, and C, and typical I - V characteristics. In particular, they utilized the electrode B to measure the resistance changes in A–B (anode side) and B–C (cathode side) during resistance switching, in order to identify the active switching location [47]. By utilizing this approach, the dominant switching location for p-type oxides based

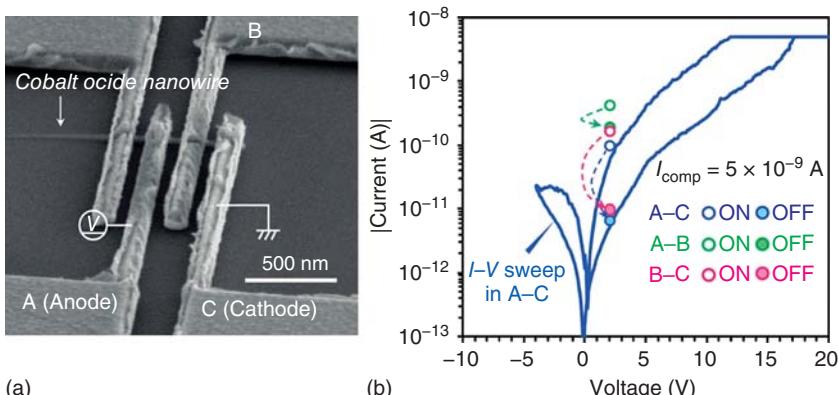


Figure 23.8 Identification of switching spatial location by utilizing planer-type NW ReRAM devices comprised of CoO_x NW. (a) SEM image of a single CoO_x NW device bridged between multielectrodes. The nanogap spacing in this device was ~ 100 nm. The voltage was applied to the electrode A, while the electrode C was grounded during resistive switching.

(b) Typical I - V characteristics; the compliance current was set to be 5×10^{-9} A. The opened and the filled circles show, respectively, the LRS and HRS currents measured in each configurations (A-C (blue), A-B (green), and B-C (pink)). (Reprinted with permission from Ref. [47]. © 2011, American Chemical Society.)

NW ReRAM, including NiO and CoO_x was found to be near the cathode side of forming process [47]. On the other hand, the switching location for n-type oxides such as TiO₂, was identified near the anode side of forming process [49, 51]. Thus, there is a clear dependence of the polarity of mobile carriers on the switching location of bipolar resistive switching. This trend was interpreted in terms of the motion of negatively-charged oxygen ions in presence of electric fields.

When the size of resistive switching device approaches the nanometer scale, the effect of the surroundings on the memory properties can be greatly amplified. Thus, in order to design reliable small resistive switching memory devices, it is important to design not only the memory body but also the surroundings, for example, the environment where the switching experiments are carried out. The NW planar-type device configuration (Figure 23.6) can be also utilized to examine such surrounding effects on the nanoscaled resistive switching devices. Figure 23.9 shows the effect of surrounding atmosphere on the resistive switching behaviors of TiO₂ NW ReRAM devices [77]. Figure 23.9 shows the pressure dependence on the I - V data of TiO_{2-x} planar-type device, as a function of ambient pressure. Interestingly, resistive switching and its characteristic hysteresis loop emerge when the pressure of surrounding atmosphere is decreased down to 5 Pa. If the pressure is further decreased, the hysteresis loop becomes smaller, and the OFF state current significantly increases. The I - V measurements were performed at least for five cycles to ensure the reproducibility (insets of Figure 23.9). The results of these type of experiments, also performed on other systems based on NiO and CoO_x NWs, demonstrated that the thermodynamical interaction with

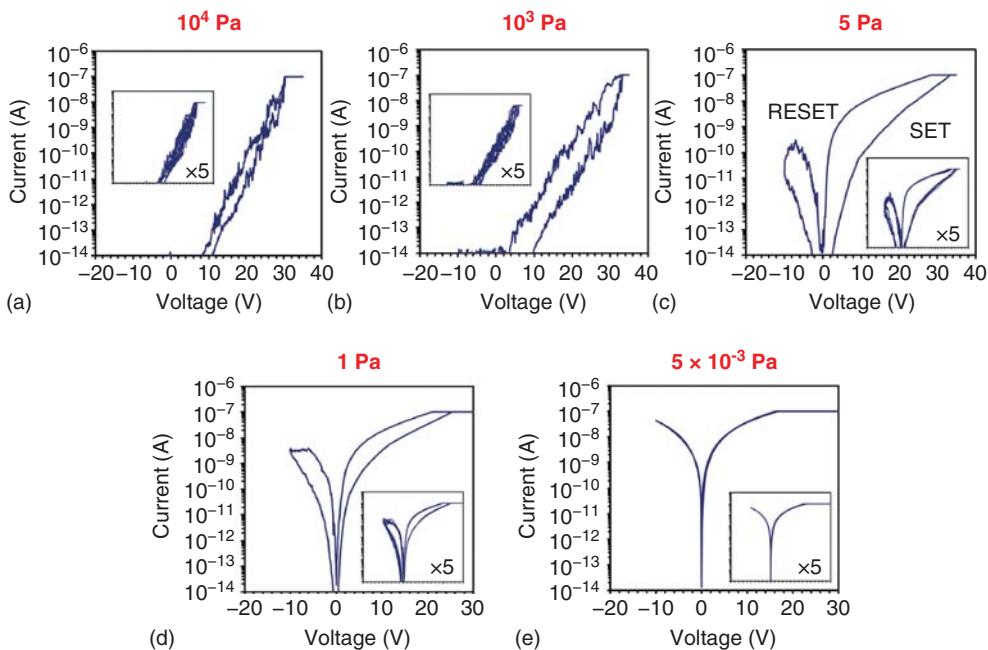


Figure 23.9 (a–e) Identification of effect of surroundings on resistive switching properties by utilizing planar-type NW ReRAM based on TiO_{2-x} . (Adapted with permission from Ref. [77]. © 2012, American Chemical Society.)

the environment critically impacts the bipolar resistive switching behavior via altering the thermodynamical stability of resistive switching [77]. In addition, this planar-type device allows to examine the effect of a solid passivation layer, such as SiO_2 , on the resistive switching properties, to tailor the surrounding conditions for small devices [77].

In the previously discussed NW ReRAMs, the resistive switching behaviors were only bipolar or unipolar type. However, there are many reports on the occurrence of resistive switching behaviors of different type in NW and thin film ReRAM, even using the same oxide. For instance, NiO based NW ReRAM can exhibit both unipolar and bipolar type of switching (see Section 23.3.1). Figure 23.10 shows the size effect on the occurrence of bipolar or unipolar resistive switching behaviors [76] for some oxides, such as CoO_x , NiO, and TiO_2 . Although the device size effect on the bipolar and unipolar switching types has been underestimated, these experimental data using small NW ReRAM devices clearly demonstrated the significant size effect on the polarity of resistive switching behaviors [76].

In conclusion, the planar-type NW ReRAM devices offer a novel way to study not only intrinsic nanoscale resistive switching properties, which have been buried in conventional capacitor devices, but also for designing improved memory devices by choosing appropriate passivation layers [47, 76, 77].

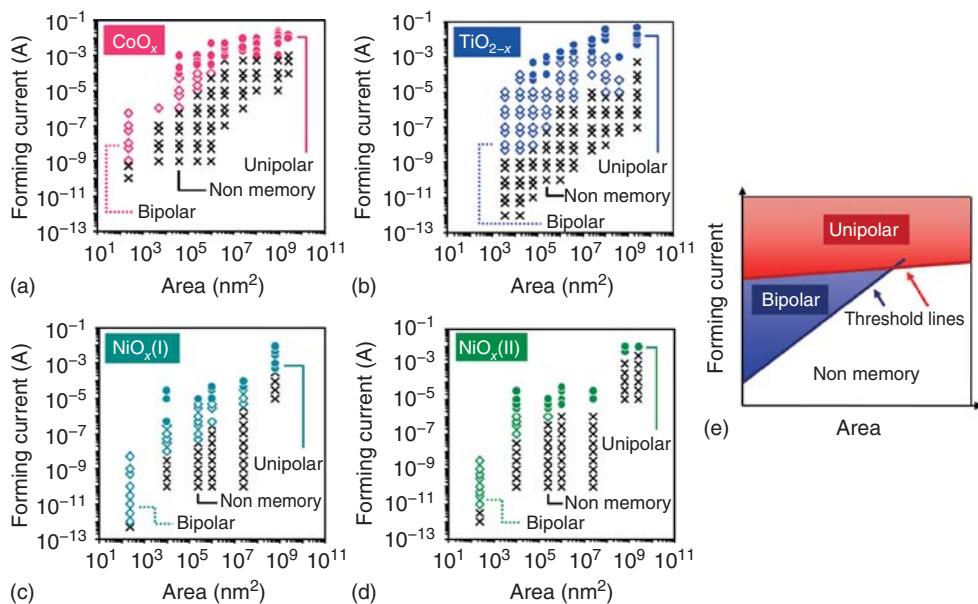


Figure 23.10 Emergence of electrical polarity characteristics in resistive switching behaviors as a function of cell area and forming current. Diagram of the variation of polarity characteristics in switching behaviors for CoO_x (a), TiO_{2-x} (b), and NiO_x (c,d). The data of two different NiO_x layers are

shown as a function of deposition conditions. (e) Schematic of observed scaling trends on the emergence of the two electrical polarity characteristics when varying a cell area and forming current. (Reprinted with permission from Ref. [76]. © 2013, Nature Publishing Group.)

23.3.4

Resistive Switching in NW ReRAM with Active Electrodes: ECM Mechanisms

In the previous paragraphs, we discussed the resistive switching properties of NW ReRAM based on oxides, in combination with noble or passive metal electrodes. The mechanisms of resistive switching properties of ReRAM based on these systems, as also reported for thin solid films, usually rely on thermal effect for unipolar switching and/or valence change mechanisms for bipolar switching [3]. On the other hand, another large class of ReRAM devices exploits the electrochemical concept (ECM) by using active electrode as Ag or Cu, where the switching is accomplished by formation and disconnection of metal filaments inside the oxide or electrolyte by cation migration [6].

Few works have been presented in the literature on this type of switching mechanism in ZnO single NWs in combination with Cu or Ag electrodes [36, 78, 79]. Yang *et al.* [36] studied the nonvolatile resistive switching in single-crystalline ZnO NWs grown on sapphire substrates by CVD. The growth process was catalyzed by gold particles, which were synthesized by selective wet-etching of gold film on self-assembled polystyrene microsphere monolayers. The sizes

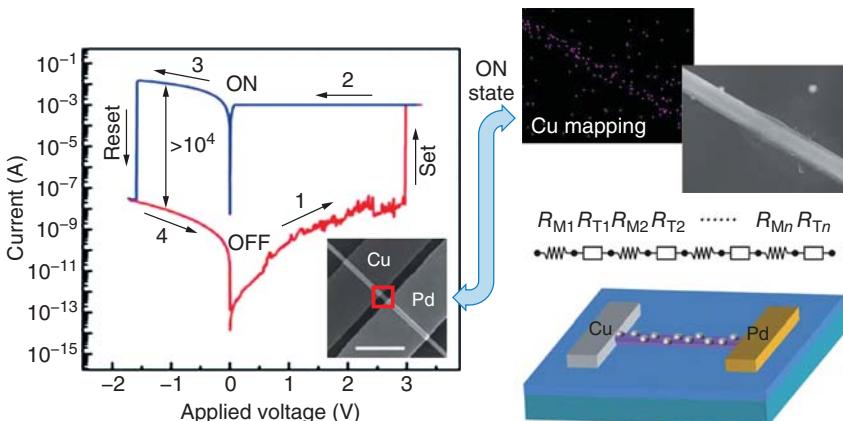


Figure 23.11 I - V curves showing bipolar resistive switching of Cu/ZnO-NW/Pd device (left), EDX analyses and schematic diagram for the formation of metallic filament on the

NW surface and its equivalent circuit (right). (Reproduced from Ref. [36] with permission of The Royal Society of Chemistry.)

of the ZnO NWs were typically $\sim 30\text{ }\mu\text{m}$ in length and $\sim 150\text{ nm}$ in diameter. In order to fabricate ReRAM NW devices, the as-grown ZnO NWs were dispersed in ethanol and then dropped onto a heavily doped Si substrate covered with a 300 nm-thick thermally-grown SiO_2 layer. The Pd/Au contacts were prepared via sequential processes of photolithography, Pd/Au deposition, and lift-off. Then Cu deposition (100 nm) and lift-off process were used to form Cu contacts. Reliable bipolar resistive switching with threshold voltages $\leq 3\text{ V}$ was achieved when the electrode spacing was $\sim 1\text{ }\mu\text{m}$ or below, and the switching was attributed to the formation/annihilation of metallic Cu filaments (Figure 23.11). It should be noted that, in contrast, NiO single crystalline NWs in combination with Pt electrodes exhibit bipolar switching at large voltages $> 10\text{ V}$ (Figure 23.3) [34]. Furthermore, the resistive switching in single crystalline ZnO NWs was explained by the formation of a metal island chain on the NW surface (Figure 23.11), differently from the mechanism of continuous metal filament formation along grain boundaries in polycrystalline thin films [3, 80]. This could be further explained also taking into account the extremely low solubility of Cu in single-crystalline ZnO [81]. Therefore, the Cu ions may drift along the NW surface forming a metal filament. The proposed mechanism was supported by energy-dispersive X-ray (EDX) chemical mapping in a scanning TEM. Further the authors studied the ON resistance as a function of temperature and modeled the results taking into account that the metal filament in the ON state is composed of discrete metal islands, and transport between neighboring islands is fulfilled by electron tunneling [46, 82] (Figure 23.11).

Qi *et al.* recently [79] reported a similar switching mechanism for Na-doped ZnO NWs with Ag electrodes, where the bipolar resistive switching is induced by the formation and rupture of an Ag nanoisland chain on the NW surface.

Finally, Wu *et al.* [66] reported the bipolar switching based on ECM mechanism in single Cu_{2-x}Se NWs in combination with Au and Cu electrodes. The NWs were synthesized using a modified solution method [83]. The as-synthesized Cu_{2-x}Se NWs were single-crystalline with face-centered cubic structure, 300–800 nm in diameter and 40 μm long. To fabricate the memory device, the synthesized NWs are dispersed uniformly onto $\text{SiO}_2/\text{p}^+ \text{-Si}$ substrates and Au and Cu electrodes as the opposite side of the NW are defined by photolithography, e-beam evaporation, and a lift-off processes. The distance between the two parallel electrodes is $\sim 15 \mu\text{m}$.

Electrical analysis of the device reveals bipolar resistive switching behavior (Figure 23.12) with very good reproducibility, a conductance ratio as high as 10^8 and low switching voltages (V_{SET} in the 0.3–0.6 V range and V_{RESET} in the $-0.3/-0.05$ V range) despite the 15 μm NW length. It should be noted that

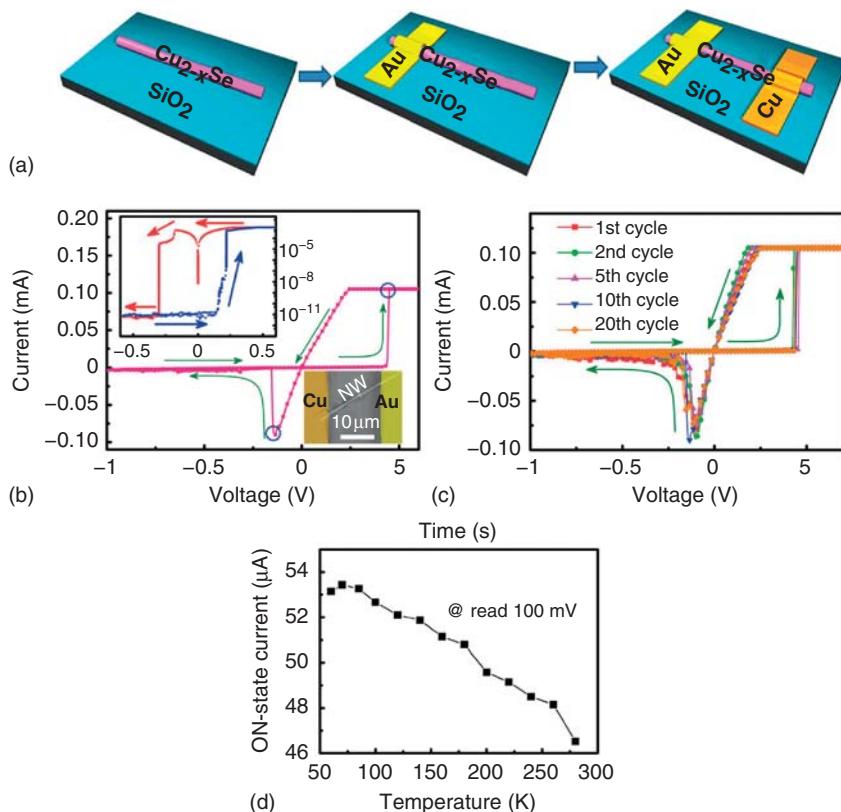


Figure 23.12 (a) Schematic illustration of the step-wise process for fabrication of ReRAM nanodevice based on a single Cu_{2-x}Se NW. (b,c) Typical I - V curve of the as-fabricated nanodevice. The arrows

indicate the voltage sweeping direction. (d) The change of the on-state current at various low temperatures. (Reprinted (and adapted) with permission from Ref. [66], © 2013, AIP Publishing LLC.)

although the set/reset voltages of the present device are comparable to previously reported Cu⁺-based resistive switching devices such as Cu₂S films [84], they are lower than that of ReRAM based on other nanostructures such as ZnS nanoribbon [85], Si/a-Si core/shell NWs [46], and ZnO NWs [36].

Regarding the switching mechanisms, the authors found that the switching behavior is due to the creation and annihilation of the Cu filament within the Cu_{2-x}Se NWs, and not on the surface as reported for ZnO single crystal NW. The difference could be related to the high-mobility of Cu⁺ cations in the Cu_{2-x}Se NWs in contrast with the low mobility in single crystalline ZnO NW. The conducting filament model was experimentally verified by the increase of ON state resistance with decreasing temperature, which is typically of metallic type of conduction (Figure 23.12d). Assuming that the conducting metallic filament formed in the Cu_{2-x}Se NW is pure copper and treating the resistance as a single filament for simplicity, the equivalent diameter of the copper filament can be estimated to be ~14 nm, comparable to the reported values in [86].

23.4

Resistive Switching in Axial Heterostructured NWs

Heterostructured NWs in the axial direction, such as for instance segmented MOM, have been proposed as prototypical ReRAM memory cells [15, 87–89]. In this approach, electrodes are directly integrated in the NW by self-assembly with the possibility to control the oxide/electrode interface and the oxide segment length down to few tens of nanometers [53]. The latter two characteristics are both fundamental to control the resistive switching properties and to reduce the switching power [15, 87]. The MOM structure represents a first step of a NW ReRAM which can be integrated in a resistive memory crossbar array [87].

Figure 23.13 show the Au–NiO–Au ReRAM devices based on a MOM NW structure synthesized by an AAO template-assisted method [61, 87]. The NW had a diameter of 250 nm and the length of the NiO segment was 900 nm. The all-metal Au–Ni–Au NWs were achieved by electroplating alternating segments of Au and Ni into the pores of an AAO matrix. Then the NWs were released from the AAO matrix, dispersed onto a SiO₂/Si substrate and annealed at 600 °C in air to convert the Ni segment into polycrystalline NiO with a grain size of 8 nm. High resolution TEM analyses showed a perfect Au/NiO interface. The single ReRAM NW cell was therefore connected to large area Au pads for electrical testing (Figure 23.13a). The *I-V* curves in Figure 23.13b shows the bipolar switching with set/reset voltages around 6 V and reset current around 250 μA.

Recently, it has been also demonstrated the fabrication of Au/NiO/Au NWs arrays embedded in an AAO matrix, with a diameter of 50 nm and a length of the NiO segment of 20 nm and below [15, 53]. The NW arrays were prepared by consecutive electrodeposition of Au–Ni–Au metal segments into the pores of the AAO matrix, polishing of the surface to expose the upper part of the Au

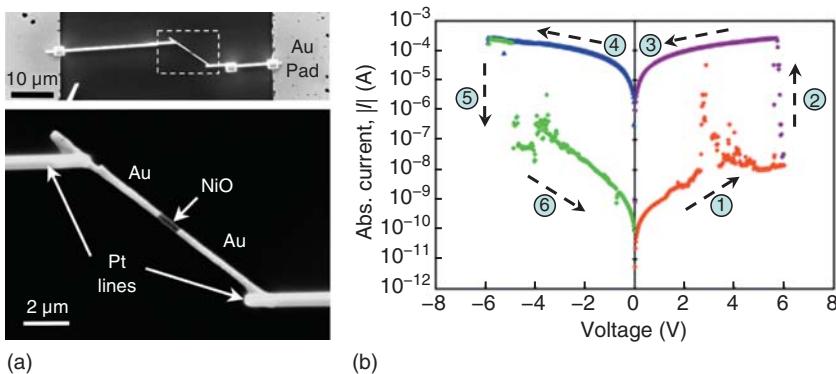


Figure 23.13 (a) SEM micrographs of a single Au–NiO–Au NW device showing the Au contact pads to the extreme left and right and the Pt lines connecting the Au contact pads to Au ends of

the NW. (b) I - V responses from individual Au–NiO–Au, showing bipolar resistive switching. (Reprinted (and adapted) with permission from Ref. [87], © 2009, AIP Publishing LLC.)

segment and *in situ* oxidation of the Ni segment to polycrystalline NiO, without removing the template. The resulting array consists of MOM heterostructured NWs embedded in the AAO matrix, where the electrical properties of the single NWs are accessed by using the C-AFM [15]. After the forming stage, the single Au/33 nm NiO/Au exhibits unipolar switching (Figure 23.14a). The electric field required for the forming stage ($1.2\text{--}2\text{ MV cm}^{-1}$) is comparable to the one reported for thin films [88], while V_{SET} and V_{RESET} voltage are around 3 and 1.5 V, respectively. The measured reset current was extremely low when compared to other works on NWs and it was found to vary from hundreds to few nanoamperes. The almost linear relationship between I_{RESET} and the initial current of the LRS, as reported for unipolar ReRAM based on thin film, is demonstrated also for NWs, confirming the role of Joule heating in the reset process also for very low (tens of nanoamperes) reset currents (Figure 23.14b). The extremely low switching current were ascribed by the authors to a confinement effect of the CF into the NWs further assisted by the radial oxidation profile of the NiO segment, possibly leading to a under-stoichiometric and defective region in the center of NW [15, 53].

The same authors reported the bipolar switching in nonsymmetric Au/200 nm NiO_x /400 nm Ni/Au NWs arrays still embedded in the AAO matrix (Figure 23.14c) [89]. This work demonstrated that it is possible to engineer the structures of an axial heterostructured NW to achieve either unipolar or bipolar switching by a fully bottom-up approach (Figure 23.14d).

Finally, besides MOM NWs, also NW arrays composed of metal chalcogenide-based mixed conductor (Ag_2S , Cu_2S , etc.) as active switching material and Ag/ Ag_2S axial heterostructures arrays have been synthesized into AAO NWs by electrochemical synthesis to explore bipolar resistive switching based on ECM mechanism [42, 43, 90]. Liang *et al.* [42, 90] fabricated Ag/ Ag_2S arrays by electrodeposition of Ag into the AAO pores followed by an anodic polarization

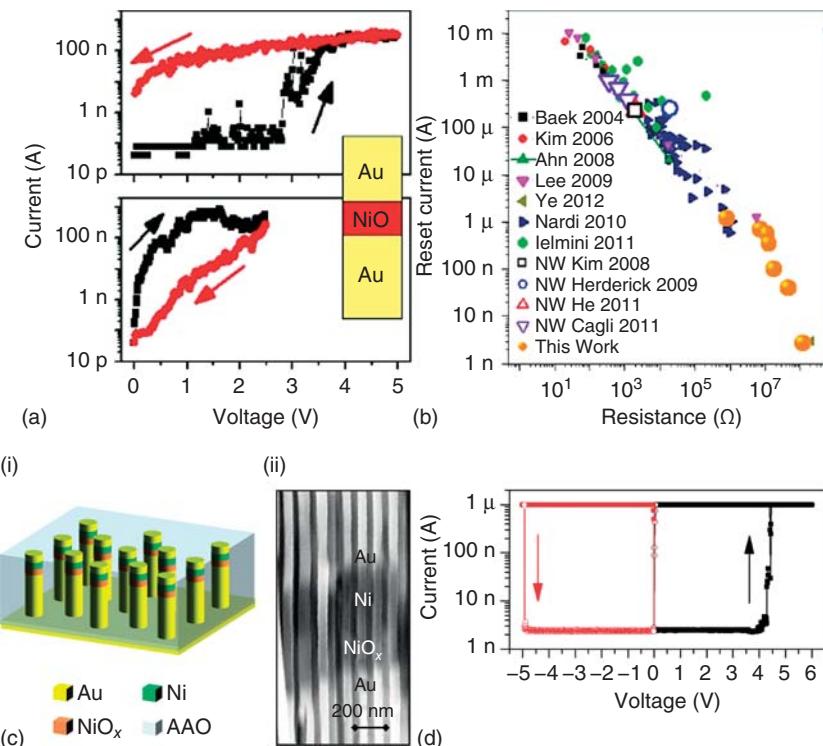


Figure 23.14 (a) Set and reset transition as measured by C-AFM in a single Au/NiO/Au NW. (b) Reset current as a function of the resistance of the ON state. (c) Schematic (i) and SEM cross-sectional image (ii) of the

Au/ NiO_x /Ni/Au NW array. (d) I - V curve showing bipolar switching of the Au/ NiO_x /Ni/Au device. (Reprinted with permission from Ref. [89], © 2013, AIP Publishing LLC.)

process to grow metal sulfide in aqueous hydrosulfide (HS-) solutions. Other groups reported the fabrication into AAO template of segmented Ni–Ag₂Se–Ni NWs (200 nm diameter) by electrodeposition [43] or CuTCNQ NW array via vacuum-vapor-induced reaction method [91].

23.5

Core–Shell NWs toward Crossbar Architectures

Core–shell NWs with a metallic core as electrode and an active material in the shell, such as an oxide, provide the most promising solution to obtain a nanoscale ReRAM. Indeed, heterostructured NWs in the radial direction can have various advantages on axially-heterostructures NWs, such as the segmented NWs, especially to build crossbar ReRAM arrays. The advantages of core–shell NWs are: (i) the presence of a conductive core, which acts as a metal electrode and provides the

interconnect line to be used in a crossbar array, and (ii) the possibility to control and engineer the properties and thickness of the shell layer acting as the active switching, for example, to achieve low-voltage and low-current switching. Therefore, when assembling NWs in a crossbar architecture, the bit size is defined by the diameters of the orthogonal NWs, while the electronic characteristics of the functional elements are defined by the compositions of the two NWs (e.g., coaxial core/shell materials). Various strategies have been proposed for the core–shell NW synthesis, such as (i) surface oxidation of a conductive NWs to form the shell layer as in the case of Ni/NiO_x [13, 14] or Ge/GeO_x [92] and (ii) sequential growth/deposition of the core and the shell as reported for Si/a-Si [46] and Au/Ga₂O₃ [45]. Even if the latter method is less explored and in principle more complex, it could allow more flexibility to engineer the core/shell structure and composition.

23.5.1

Crossbar Devices with Si(core)/a-Si(shell) NWs and Ag Electrodes

The concept of NWs ReRAM in a crossbar configuration were first proposed by Dong *et al.* [46], which demonstrated the bipolar switching in a crosspoint device based on a NW with monocrystalline Si core and a-Si shell (~ 5 nm) and an Ag crossing metal line (Figure 23.15a,b). The devices were fabricated by using a hybrid bottom-up/top-down approach. First, the core/shell NWs were fabricate by a two-step CVD process which involved (i) metal-nanocluster-catalyzed Si NW core growth followed by (ii) homogeneous deposition of the amorphous Si shell. Then, the NWs were assembled on a substrate using fluidic-based alignment [27], while Ni-metal contacts were defined at the NW ends to achieve ohmic junctions to Si NWs [93]. Finally, an additional lithography step was used to define one or more crossed metal Ag NWs. Figure 23.15c shows the representative current versus voltage (*I*-*V*) data obtained from a single Si/a-Si NW crossing a Ag NW. The device is initially in a high resistance OFF state, and when the voltage is increased from 0 to 4 V, the current abruptly increases at around 3 V, that is, the device switches to a low resistance ON state of about 300 k Ω . Second, as the voltage is subsequently reduced to a negative threshold value (-3 V in this case) the device switches back to the high-resistance OFF state. It should be noted that the devices exhibit intrinsic current rectification ($>10^6$ at ± 1.5 V), that is, the crossed NW devices show low conductance in the ON state when the applied voltage is negative. Current rectification is an attractive property, because it can minimize cross talk between individual elements in arrays [94]. Regarding the mechanism, the set transition can be ascribed to the formation of a conductive Ag filament by cation migration in the a-Si shell, and the current in the ON state is dominated by tunneling between metal islands forming the filaments. Temperature-dependent *I*-*V* measurements were carried out to obtain further insight into the switching mechanism. The turn-on threshold voltage increased from 3 to 4 V as the temperature is reduced

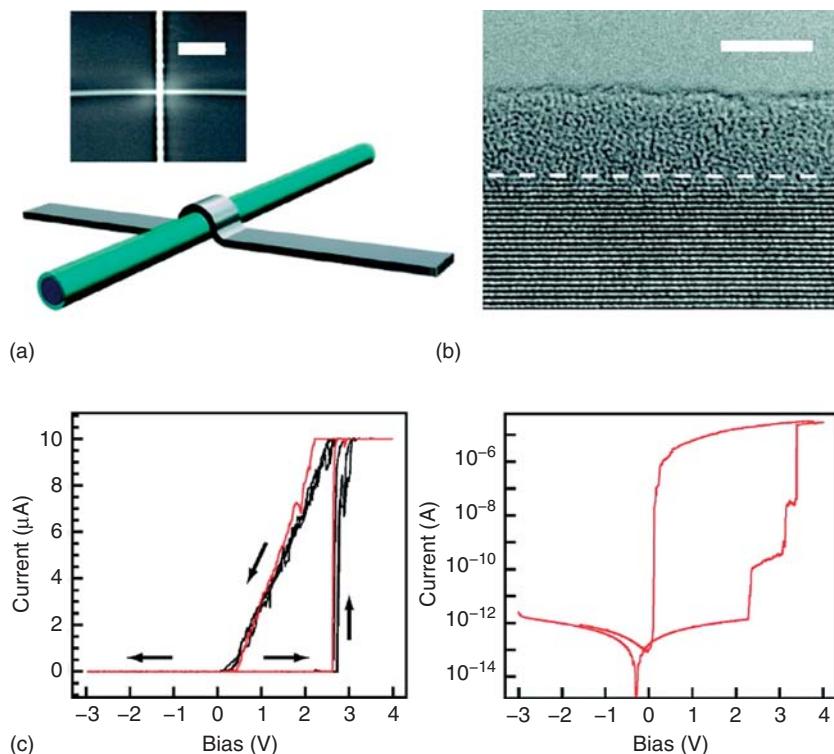


Figure 23.15 (a) A single switch is formed at the crosspoint of a Si (blue)/a-Si (cyan) core/shell NW and a metal NW (gray). (Inset) SEM image of a Si/a-Si NW (horizontal) crossed Ag-metal NW (vertical) device; scale bar is 1 μm. (b) HRTEM image of a Si/a-Si

core/shell NW. Dashed line indicates the interface between core and shell. Scale bar is 5 nm. (c) *I*-*V* sweeps in the linear and log scale. (Adapted with permission from Ref. [46]. © 2008, American Chemical Society.)

from 350 to 50 K, likely due to decrease of the Ag-ion diffusivity in the a-Si matrix [95]. As the temperature is reduced below 250 K, the current rectification is lost and a more conventional resistor-like behavior is observed in the ON state. These results indicate that the rectification behavior observed at high temperature is related to the insufficient stability of the CF formed under positive voltage. Because diffusion of the metal ions is a thermally-activated process, it is possible for the metal ions to diffuse away from SiNW core at high temperatures and thereby yield a high-resistance state at small negative bias, which is at the basis of the rectifying behavior. Finally, the authors demonstrated by systematic studies that in their NW devices: (i) the bit size can be as small as 20 nm × 20 nm, (ii) the writing time is <100 ns, (iii) the retention time is >2 weeks, and (iv) devices can be switched >10⁴ times without degradation in performance. In addition, they demonstrated also that crossbar array of Si/a-Si NW and Ag NW can be fabricated on plastic substrates without compromising the functionality.

23.5.2

Crossbar Devices with Ni(core)/NiO(shell) NWs and Ni Electrodes

Core–shell metal–oxide NW ReRAM based on Ni/NiO system has been proposed for crossbar ReRAM with low-power operation [15] and a better stability at high temperature of ON state [9, 13].

He *et al.* [14] studied the resistive switching properties of single Ni/NiO core/shell NWs fabricated by electrodepositing Ni in the pores of AAO templates and then oxidizing them in air for 3 days after release from the matrix. To fabricate the NW ReRAM device (Figure 23.16), Au electrodes consisting of wires of about 2–3 μm diameter were patterned (by electron beam lithography and lift-off) before the oxidation procedure to contact the Ni core and after the oxidation to contact the shell NiO_x layer. The NW has an average diameter of 75 nm and an amorphous NiO layer with a thickness of 5 nm. Figure 23.16 shows the high-resolution TEM image of part of the NW, which shows the crystalline Ni and amorphous NiO shell. The atomic ratio between Ni and O is about 1 : 0.87 by EDX analysis, indicating the presence of the oxygen vacancies in the NiO shell. The proposed structure displays controllable memory and threshold resistance switching depending on the compliance current used during the set operation at room temperature. Figure 23.16 shows the I - V curves of the device during the set operation at various current compliance I_{CC} fixed at 10 μA (red curve) and 100 μA (black curve), while the blue curve represents the reset process. The threshold switching occurs by setting a low I_{CC} , while the high I_{CC} triggers the memory switching with low threshold voltages thanks to the thin shell layer. The authors attributed the physical mechanisms for the threshold resistive switching to the electrical-field driven transition without forming a CF [14, 96, 97]. Instead, the memory resistive switching is due to the formation of a chain of oxygen vacancies

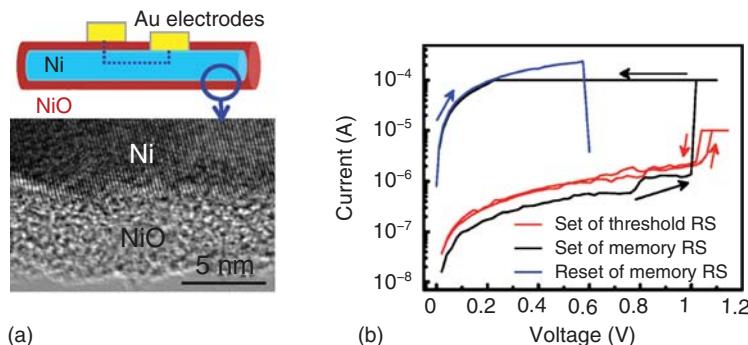


Figure 23.16 (a) Sketch of the device and TEM image of an individual Ni/NiO NW evidencing the amorphous structure of the shell layer. (b) The I - V curves showing threshold (red) and memory resistive switching (RS) with set (black) and reset

(blue) transitions. The threshold or memory switching were achieved by setting different compliance currents. (Adapted with permission from Ref. [14]. © 2011, American Chemical Society.)

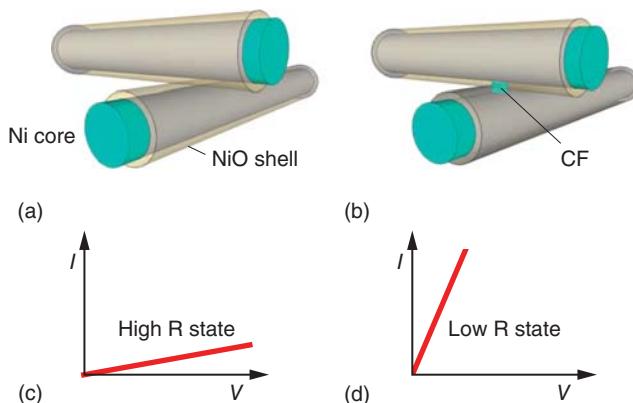


Figure 23.17 Schematic illustration of the crossbar memory structure formed from two NiO–Ni core–shell NWs, in high- (a) and low-resistance states (b), differing by the presence of a CF, and the consequent

I-*V* curves (c,d). Ni cores act as word line/bit line interconnects, while the NiO shell is the active switching layer. (Reprinted with permission from Ref. [13]. © 2011, John Wiley and Sons.)

(V_o) triggered by electric field and Joule heating which activate oxygen vacancies migration at high I_{CC} . During the reset process, the filament is disrupted by Joule heat causing the rupture of V_o – V_o interactions.

Cagli *et al.* [13] expanded the work on Ni (crystalline core)/NiO (amorphous shell) NWs to build crossbar switches of resistive switching devices (Figure 23.17). The authors proposed a fully bottom-up crossbar array, where the memory elements are positioned at the crossing point of perpendicular lines. Figure 23.17 shows the schematic of the proposed concept, where each NW consists of a Ni wire (core) acting as electrode and metal interconnection, and a NiO shell as the active switching layer. Resistive switching is accomplished by formation/dissolution of a CF connecting/disconnecting the two Ni cores through the NiO shell (low R state/high R state). The Ni/NiO NWs were fabricated following the approach of [12], starting from the electro-deposition of polycrystalline Ni NWs of about 30–40 μm within an AAO template with 200 nm pore diameter. After dissolution of the template, the NWs were deposited on an oxidized Si substrate and then aligned by using a magnetic field, to achieve the crossbar structure (Figure 23.18) with perpendicular NWs. The NiO shell of the bottom NW was achieved by thermal oxidation in air at 300 °C for 3 h leading to a 15 nm thick amorphous layers, with a nonstoichiometric composition Ni_xO with $x > 1$. The top Ni NWs oxidized spontaneously in air before the alignment step. For electrical testing the NWs were connected to external Au pads. The single Ni/NiO core/shell devices exhibit unipolar type of switching with low threshold voltages of around 1.5 V for the set and 0.5 V for the reset (not shown), which is ascribed to the formation and dissolution of a CF. Figure 23.18 shows *I*-*V* curves collected in four sequential cycles for a crossbar device consisting of two crossing core–shell NWs. To analyze the data, the authors considered an electrical model

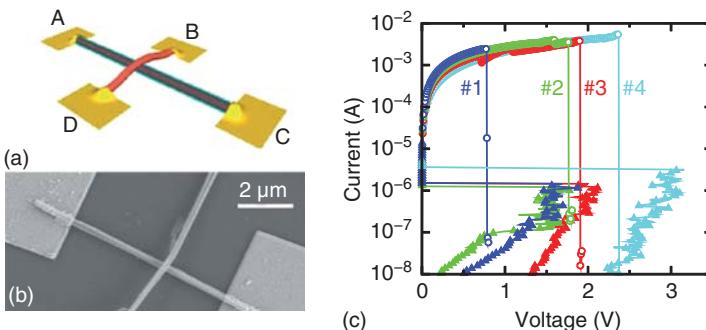


Figure 23.18 Schematic (a) and SEM images (b) of the ReRAM device fabricated by a full bottom-up crossbar between two perpendicular core–shell NWs, and (c) measured *I*-*V*

curves showing set/reset switching across AB for four sequential cycles. (Reprinted and adapted with permission from Ref. [13]. © 2011, John Wiley and Sons.)

that describes the crossbar cell as a network of five resistances, that is, the resistance at the NW–NW junction and the series resistances from the CF to one of the four electrical pads [13]. Using this model they demonstrated that forming, set and reset transitions occur in the NiO shell layer at the NW–NW crosspoint junction. Moreover, the *I*-*V* characteristics prove the possibility to achieve low voltage operation, even if with reset current still in the milliampere range. The latter finding can be related to a current overshoot during the set transient, leading to a CF overgrowth and a correspondingly high reset current [98]. This problem can be avoided by using a select transistor in the memory array to limit the supply current. The cycling endurance was only four cycles and further optimization of the system is necessary, anyway the results are relevant since they represent a first proof of concept of a full bottom-up scheme fabrication of functional NW-based crossbar memory.

More recently, bipolar resistive switching was demonstrated in single gold (core, 40 nm diameter)/ Ga_2O_3 (shell, 40 nm thickness) NW, as a result of the formation and dissolution of an oxygen vacancy filaments in the Ga_2O_3 shell connecting the Au core with the top titanium electrode defined by electron beam lithography [45]. Based on these results the authors proposed the possible integration in crossbar arrays. Indeed, the localized switching at each electrode crosspoint leads to the invariance of the V_{SET} and V_{RESET} against different distance between top electrodes crossing a single bottom electrode. Therefore, it is possible to fabricate numerous resistive switching cells in a single core/shell NW by depositing repetitive top crossing metal electrodes.

Recently the group of Stanford proposed a 3D architecture for ReRAM based on vertical core/shell pillar structure crossing a plane electrode, where the memory cell is formed at their intersection [5]. Based on circuit simulation, the best structure for vertical pillar can be composed of three layers such as a core formed by metal with low resistivity (such as W) serving as interconnecting lines, an a two shell structures, namely the active metal electrode (such as TiN) and the active

oxide layers. Even if no proof of concepts are reported in literature for vertical type of crossbar using NWs, core/shell NWs can be in principle used also for high density 3D ReRAM by combining top down and bottom up fabrication strategies.

23.6

Emerging Bottom-Up Approaches and Applications

This section reviews some emerging patterning strategies based on a combination of bottom-up and top-down approaches to fabricate one-diode one-resistor (1D1R) and 3D-based ReRAM memories.

23.6.1

1D1R Nanopillar Array

Ji *et al.* [24] reported the development of 1D1R nanopillar device architecture using nonconventional lithography. The SiO_x based ReRAM memory element is patterned by NSL and is vertically self-aligned with a Si-based PN diode by deep silicon etching of a $\text{P}^{2+}/\text{N}^+/\text{N}^{++}$ epitaxial Si wafer (Figure 23.19). The self-aligned process forms a high-density, large-scale nanopillar array architecture. The

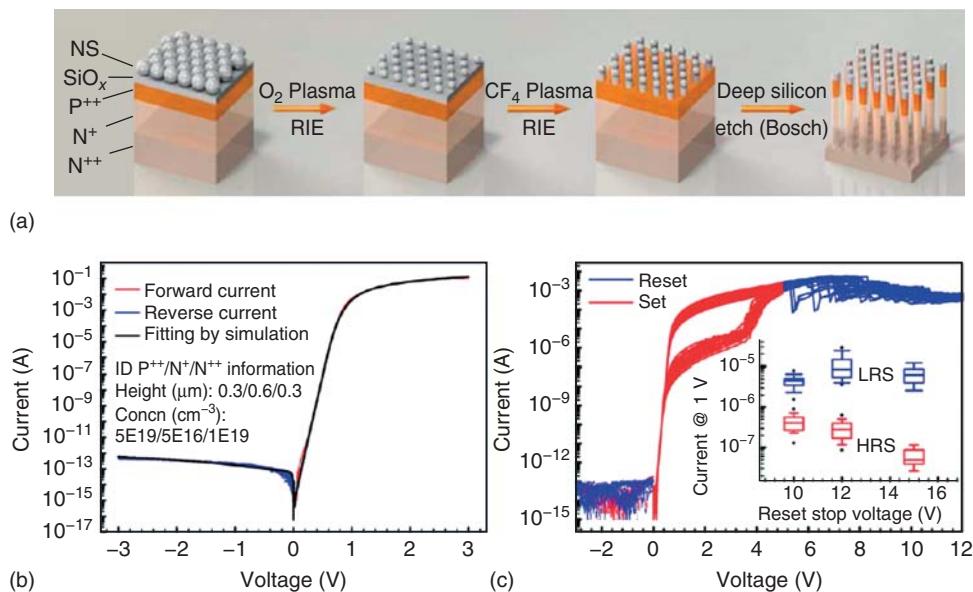


Figure 23.19 (a) The 1D1R fabrication procedure: nanosphere lithography to define the SiO_x based ReRAM memory element is followed by deep silicon etching to form the self-aligned PN diode. (b) Hundred measurement cycles of diode I - V curve of the diode (100 cycles) and simulation results

of 1D structure. (c) Resistive switching I - V cycles of the 1D1R structure. The inset shows the readout current in the low resistance (LRS) and high resistance (HRS) states at 1 V as a function of reset stop voltage. (Adapted with permission from Ref. [24]. © 2013, American Chemical Society.)

nanostructures fabricated using NSL can be well-controlled in shape, size, and interpillar spacing through direct assembly of polymer nanospheres on the wafer-scale. The average nanopillar height and diameter are 1.3 μm and 130 nm, respectively.

Figure 23.19 shows the *I-V* response of 100 V sweeps from -3 to +3 V for one diode in the nanopillar configuration. The forward current can reach 100 mA at 3 V, which is high enough to support the reset process. The reverse current is below 10^{-12} A at -3 V, which is significantly lower than the reported value in literature for Schottky diodes. Figure 23.19 shows the *I-V* curves of 30 switching cycles for a SiO_x -based 1D1R cell. After a forming process (~12 V), bipolar switching is measured with set voltage values of 3–4 V and reset voltage above 10 V. By controlling the reset stop voltage, the high-resistance state level can be controlled and the potential for multi-bit operation can be demonstrated for nanopillar SiO_x -based ReRAM.

For circuit-level applications, the set/reset voltages are still too high and should be optimized. Anyway, the low reverse-current of the selector can reduce the sneak-path leakage issue and provide larger readout margin, especially in large arrays [99].

23.6.2

Block-Copolymer Self-Assembly for Advanced ReRAM

In the last decade, the self-assembly phenomena of BCPs have received an increasing deal of attention due to the possibility to overcome the challenges of traditional nanofabrication technologies [29]. As an example of the capability of BCP, high density nanoscopic patterns for magnetic storage have been fabricated [100, 101]. Usually two main approaches can be used to fabricate functional nanostructures: the first one relies on self-assembled block-copolymers used as etch-masks or lift-off templates to transfer the pattern on another material [102]. Alternatively, selective incorporation or decoration of the block-copolymers with inorganic elements through solution or vacuum processes can be used to form ordered nanostructure, such as metal nanoparticles [103, 104]. Following the first approach, Frascaloli et al. [38] fabricated nanoscaled Pt/Ti top electrodes arranged in periodic arrays over the HfO_2/TiN surface, building ReRAM devices with a diameter of 28 nm and a density of 5×10^{10} devices/cm². The second approach has the advantage of avoiding the pattern-transfer process and therefore cost and process complexity are reduced. Recently, Park *et al.* [25] demonstrates unipolar switching in high-density silicon oxide nanostructures (6–18 nm in diameter) on graphene and metal electrodes fabricated by self-assembly of a Si-containing BCP. The devices were measured with C-AFM and could switch at relatively low power with a reset current in the order of 10 μA and V_{RESET} of 2 V. The latter approach represents a fully bottom up approach to fabricate ReRAM nanodevices, even if it is difficult to contact each single devices. On the other hand it is possible to use self-assembled BCP to define small structures in a combined top-down/bottom-up approach. Following this strategy, recently Wu *et al.* [31] proposed the first demonstration of a 3D

patterned HfO_x -based ReRAM by BCP self-assembly with area scaling down to 12 nm. The work illustrates the potential of using self-assembly lithography for manufacturing low-cost memory with ultra-high density. Further improvements of the work can be achieved with directed self-assembly strategies to form high-density structures in predefined location [29].

Finally, emerging applications for the resistive switching phenomena in NWs are related to the development of synaptic electronics, where NWs could represent the building blocks for high-density crossbar arrays to emulate the biological synaptic functions [4, 105, 106]. Currently, a fully bottom-up approach in this direction has not been demonstrated yet. On the other hand, preliminary results on the memristive properties of WO_3 NWs exhibiting a memristive hysteresis loop, similar to the ones reported on WO_x thin film used for neuromorphic application [107], have been recently reported by He *et al.* [108]. Further, the group of Boland [109] recently demonstrated the possibility to manipulate the connectivity in metallic NW networks with resistive nanojunctions by applying an electric field. The Ag and Ni NW networks were formed by spray deposition of NWs onto Si substrates coated with 300 nm of thermal oxide, while Pd and Ti/Au contact electrodes enabled transport measurement. The authors claim that by using NW networks it is possible to create nanomaterials with tunable electrical conductivity following external stimuli, and which can have various applications such as, for instance, computer networks that mimic the functions of the human brain.

23.7

Conclusions

The current status of bottom-up approaches for ReRAM fabrication has been reviewed by discussing the proposed fabrication strategies, the emerging self-assembly nanopatterning methods and the electrical characteristics of the devices. ReRAM devices have been fabricated starting from single NWs connected by external contact as electrodes in a planar type configuration, or heterostructured axial or radial (core/shell) type of NWs, also in an array configuration. The single ReRAM in a planar-type configuration is useful to assess intrinsic nanoscale resistive switching properties and mechanisms. On the other hand, the approach based on core/shell NWs is promising for applications since it is possible to tailor the properties and size of the device by controlling materials and diameter/thickness of the core and shell layer, respectively. Furthermore, core-shell NWs are suitable for both crossbar array or 3D ReRAM. Currently the performance of ReRAM devices fabricated by bottom-up approaches still does not match the ones of standard devices fabricated by top-down methods. On the other hand, there is still room for improvements toward scaling beyond 10 nm, possibly by combining top-down and bottom-up approaches, or for emerging applications such as memristive devices toward neuromorphic computations.

References

- Jeong, D.S., Thomas, R., Katiyar, R.S., Scott, J.F., Kohlstedt, H., Petraru, A., and Hwang, C.S. (2012) Emerging memories: resistive switching mechanisms and current status. *Rep. Prog. Phys.*, **75**, 076502.
- Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833.
- Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) Redox-based resistive switching memories – nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21**, 2632.
- Park, S., Noh, J., Choo, M.-L., Sheri, A.M., Chang, M., Kim, Y.-B., Kim, C., Jeon, M., Lee, B.-G., Lee, B.H., and Hwang, H. (2013) Nanoscale RRAM-based synaptic electronics: toward a neuromorphic computing devices. *Nanotechnology*, **24**, 384009.
- Deng, Y., Chen, H.-Y., Gao, B., Yu, S., Wu, S.-C., Zhao, L., Chen, B., Jiang, Z., Liu, X., Hou, T.-H., Nishi, Y., Kang, J., and Philip Wong, H.-S. (2013) Design and optimization methodology for 3D RRAM arrays. IEDM Technical Digest, pp. 629–632
- Yang, Y. and W, L. (2013) Nanoscale resistive switching devices: mechanism and modelling. *Nanoscale*, **5**, 10076–10092.
- Ielmini, D., Bruchhaus, R., and Waser, R. (2011) Thermochemical resistive switching: materials, mechanisms and scaling projections. *Phase Transitions*, **84**, 570.
- Kim, K.M., Jeong, D.S., and Hwang, C.S. (2011) Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook. *Nanotechnology*, **22**, 254002.
- Ielmini, D., Cagli, C., Nardi, F., and Zhang, Y. (2013) Nanowire-based resistive switching memories: devices, operation and scaling. *J. Phys. D: Appl. Phys.*, **46**, 074006.
- Oka, K., Yanagida, T., Nagashima, K., Kawai, T., Kim, J.-S., and Park, B.H. (2010) Resistive-switching memory effects of NiO nanowire/metal junctions. *J. Am. Chem. Soc.*, **132**, 6634–6635.
- Nagashima, K., Yanagida, T., Oka, K., Taniguchi, M., Kawai, T., Kim, J.-S., and Park, B.H. (2010) Resistive switching multistate nonvolatile memory effects in a single cobalt oxide nanowire. *Nano Lett.*, **10**, 1359–1363.
- Kim, S.I., Lee, J.H., Chang, Y.W., Hwang, S.S., and Yoo, K.-H. (2008) Reversible resistive switching behaviors in NiO nanowires. *Appl. Phys. Lett.*, **93**, 033503.
- Cagli, C., Nardi, F., Harteneck, B., Tan, Z., Zhang, Y., and Ielmini, D. (2011) Resistive-switching crossbar memory based on Ni-NiO core-shell nanowires. *Small*, **7**, 2899–2905.
- He, L., Liao, Z.-M., Wu, H.-C., Tian, X.-X., Xu, D.-S., Cross, G.L.W., Duesberg, G.S., Shvets, I.V., and Yu, D.-P. (2011) Memory and threshold resistance switching in Ni/NiO core-shell nanowires. *Nano Lett.*, **11**, 4601–4607.
- Brivio, S., Tallarida, G., Perego, D., Franz, S., Deleruyelle, D., Muller, C., and Spiga, S. (2012) Low-power resistive switching in Au/NiO/Au nanowire arrays. *Appl. Phys. Lett.*, **101**, 223510.
- Cui, Y. and Lieber, C.M. (2001) Functional nanoscale electronic devices assembled using silicon nanowire building blocks. *Science*, **291**, 851–853.
- Huang, Y., Duan, X.F., Cui, Y., Lauhon, L.J., Kim, K.H., and Lieber, C.M. (2001) Logic gates and computation from assembled nanowire building blocks. *Science*, **294**, 1313–1317.
- Lieber, C.M. (2003) Nanoscale science and technology: building a big future from small things. *MRS Bull.*, **28**, 486–491.
- Lu, W. and Lieber, C.M. (2007) Nanoelectronics from the bottom up. *Nat. Mater.*, **6**, 841–850.
- Cui, Y., Zhong, Z.H., Wang, D.L., Wang, W.U., and Lieber, C.M. (2003) High performance silicon nanowire field effect transistors. *Nano Lett.*, **3**, 149–152.

21. Kolmakov, A. and Moskovits, M. (2004) Chemical sensing and catalysis by one-dimensional metal-oxide nanostructures. *Annu. Rev. Mater. Res.*, **34**, 151–180.
22. Li, Y., Qian, F., Xiang, J., and Lieber, C.M. (2006) Nanowire electronic and optoelectronic devices. *Mater. Today*, **9**, 18–27.
23. Huang, Y., Duan, X.F., and Lieber, C.M. (2005) Nanowires for integrated multicolor nanophotonics. *Small*, **1**, 142–147.
24. Ji, L., Chang, Y.-F., Fowler, B., Chen, Y.-C., Tsai, T.-M., Chang, K.-C., Chen, M.-C., Chang, T.-C., Sze, S.M., Yu, E.T., and Lee, J.C. (2014) Integrated one diode–one resistor architecture in nanopillar SiOx resistive switching memory by nanosphere lithography. *Nano Lett.* doi: 10.1021/nl404160u
25. Park, W.I., Yoon, J.M., Park, M., Lee, J., Kim, S.K., Jeong, J.W., Kim, K., Jeong, H.Y., Jeon, S., No, K.S., Lee, J.Y., and Jung, Y.S. (2012) Self-assembly-induced formation of high-density silicon oxide memristor nanostructures on graphene and metal electrodes. *Nano Lett.*, **12** (3), 1235–1240.
26. Son, J.Y., Shin, Y.-H., Kim, H., and Jang, H.M. (2010) NiO resistive random access memory nanocapacitor array on graphene. *ACS Nano*, **4** (5), 2655–2658.
27. Huang, Y., Duan, X., Wei, Q., and Lieber, C.M. (2001) Directed assembly of one-dimensional nanostructures into functional networks. *Science*, **291** (5504), 630–633.
28. Whang, D., Jin, S., Wu, Y., and Lieber, C.M. (2003) Large-scale hierarchical organization of nanowire arrays for integrated nanosystems. *Nano Lett.*, **3** (9), 1255–1259.
29. Kim, B.H., Kim, J.Y., and Kim, S.O. (2013) Directed self-assembly of block copolymers for universal nanopatterning. *Soft Matter*, **9**, 2780–2786.
30. Shin, H.W., Park, J., Chung, H.Y., Kim, K.H., Kim, H.-D., and Kim, T.G. (2014) Highly uniform resistive switching in SiN nanorod devices fabricated by nanosphere lithography. *Appl. Phys. Express*, **7**, 024202.
31. Wu, Y., Yi, H., Zhang, Z., Jiang, Z., Sohm, J., Wong, S., and Wong, H.-S. (2013) First demonstration of RRAM patterned by block copolymer self assembly. *IEDM Technical Digest*, pp. 550–553
32. Park, W.I. et al. (2013) Self-Assembled incorporation of modulated block copolymer nanostructures in phase-change memory for switching power reduction. *ACS Nano*, **7** (3), 2651–2658.
33. Hobbs, R.G., Petkov, N., and Holmes, J.D. (2012) Semiconductor nanowire fabrication by bottom-up and top-down paradigms. *Chemistry of materials*, **24**, 1975–1991.
34. Oka, K., Yanagida, T., Nagashima, K., Tanaka, H., and Kawai, T. (2009) Nonvolatile bipolar resistive memory switching in single crystalline NiO heterostructured nanowires. *J. Am. Chem. Soc.*, **131**, 3434–3435.
35. Cheng, B., Ouyang, Z., Chen, C., Xiao, Y., and Lei, S. (2013) Individual Zn_2SnO_4 -sheathed ZnO heterostructure nanowires for efficient resistive switching memory controlled by interface states. *Sci. Rep.*, **3**, 3249.
36. Yang, Y.C., Zhang, X.X., Gao, M., Zeng, F., Zhou, W.Y., Xie, S.S., and Pan, F. (2011) Nonvolatile resistive switching in single crystalline ZnO nanowires. *Nanoscale*, **3**, 1917–1921.
37. Chiang, Y.D., Chang, W.Y., Ho, C.Y., Chen, C.Y., Ho, C.H., Lin, S.J., Wu, T.B., and He, J.H. (2011) Single-ZnO-nanowire memory. *IEEE Trans. Electron Devices*, **58**, 1735–1740.
38. Frascari, J., Brivio, S., Ferrarese Lupi, F., Seguini, G., Boarino, L., Perego, M., and Spiga, S. (2015) Resistive Switching in High-Density Nanodevices Fabricated by Block Copolymer Self-Assembly. *ACS Nano*, **9**, 2518 (DOI: 10.1021/nn505131b).
39. Wu, Y.Y. and Yang, P.D. (2001) Direct observation of vapor-liquid-solid nanowire growth. *J. Am. Chem. Soc.*, **123**, 3165–3166.
40. Yang, P.D., Yan, H.Q., Mao, S., Russo, R., Johnson, J., Saykally, R., Morris, N., Pham, J., He, R.R., and Choi, H.J. (2002) Controlled growth

- of ZnO nanowires and their optical properties. *Adv. Funct. Mater.*, **12**, 323–331.
41. Barth, S., Hernandez-Ramirez, F., Holmes, J.D., and Romano-Rodriguez, A. (2010) Synthesis and applications of one-dimensional semiconductors. *Prog. Mater. Sci.*, **55**, 563–627.
 42. Liang, C., Terabe, K., Hasegawa, T., and Aono, M. (2007) Resistance switching of an individual Ag₂S/Ag nanowire heterostructure. *Nanotechnology*, **18**, 485202.
 43. Lee, N.J., An, B.H., Koo, A.Y., Ji, H.M., Cho, J.W., Choi, Y.J., Kim, Y.K., and Kang, C.J. (2011) Resistive switching behaviour in a Ni-Ag₂Se-Ni nanowire. *Appl. Phys. A*, **102**, 897–900.
 44. Kolar, J., Macak, J.M., Terabe, K., and Wagner, T. (2014) Down-scaling of resistive switching to nanoscale using porous alumina membranes. *J. Mater. Chem. C.*, **2**, 349–355.
 45. Hsu, C.-W. and Chou, L.-J. (2012) Bipolar resistive switching of single gold-in-Ga₂O₃ nanowire. *Nano Lett.*, **12**, 4247–4253.
 46. Dong, Y., Yu, G., McAlpine, M.C., Lu, W., and Lieber, C.M. (2008) Si/a-Si core/shell nanowires as non-volatile crossbar switches. *Nano Lett.*, **8**, 386.
 47. Nagashima, K., Yanagida, T., Oka, K., Kanai, M., Klamchuen, A., Kim, J.S., Park, B.H., and Kawai, T. (2011) Intrinsic mechanisms of memristive switching. *Nano Lett.*, **11**, 2114–2118.
 48. Nagashima, K., Yanagida, T., Oka, K., and Kawai, T. (2009) Unipolar resistive switching characteristics of room temperature grown SnO₂ thin films. *Appl. Phys. Lett.*, **94**, 242902.
 49. Nagashima, K., Yanagida, T., Kanai, M., Celano, U., Rahong, S., Meng, G., Zhuge, F., He, Y., Park, B.H., and Kawai, T. (2013) Carrier type dependence on spatial asymmetry of unipolar resistive switching of metal oxides. *Appl. Phys. Lett.*, **103**, 173506.
 50. Oka, K., Yanagida, T., Nagashima, K., Tanaka, H., and Kawai, T. (2008) Growth atmosphere dependence of transport properties of NiO epitaxial thin films. *J. Appl. Phys.*, **104**, 013711.
 51. Nagashima, K., Yanagida, T., Kai, M., Oka, K., Klamchuen, A., Rahong, S., Meng, G., Horprathum, M., Xu, B., Zhuge, F., He, Y., and Kawai, T. (2012) Switching properties of titanium dioxide nanowire memristor. *Jpn. J. Appl. Phys.*, **51**, 11PE09.
 52. Pan, C.F. and Zhu, J. (2009) The synthesis, properties and applications of Si, ZnO, metal, and heterojunction nanowires. *J. Mater. Chem.*, **19**, 869.
 53. Perego, D., Franz, S., Bestetti, M., Cattaneo, L., Brivio, S., Tallarida, G., and Spiga, S. (2013) Engineered fabrication of ordered arrays of Au-NiO-Au nanowires. *Nanotechnology*, **24**, 045302.
 54. Zhao, X. *et al.* (2008) Field-induced semiconductor-metal transition in individual NiO-Ni Schottky nanojunction. *Appl. Phys. Lett.*, **93**, 152107.
 55. Luo, J., Zhu, J., Huang, Z., and Zhang, L. (2007) Arrays of Ni nanowire/multiwalled carbon nanotube/amorphous carbon nanotube heterojunctions containing Schottky contacts. *Appl. Phys. Lett.*, **90**, 033114.
 56. Masuda, H. and Fukuda, J. (1995) Ordered nanohole arrays made by a two-step replication of honeycomb structures of anodic alumina. *Science*, **268**, 1446.
 57. Masuda, H., Yamada, H., Satoh, M., Asoh, H., Nakao, M., and Tamamura, T. (1997) Highly ordered nanochannel-array architecture in anodic alumina. *Appl. Phys. Lett.*, **71**, 2770.
 58. Stasi, V., Cattaneo, G., Franz, S., Bestetti, M., Ubaldi, M.C., Piccinin, D., and Pietralunga, S.M. (2007) Aluminium pre-patterning for highly ordered nanoporous anodized alumina. *Photonics Nanostruct. Fundam. Appl.*, **5**, 136–139.
 59. Wehrspohn, R.B. (ed) (2005) *Ordered Porous Nanostructures and Applications*, Springer.
 60. Bestetti, M. *et al.* (2007) Porous anodic alumina templates: synthesis and application. *AIFM Galvanotecnica e Nuove Finiture*, **17** (1), 8.
 61. Tresback, J.S., Vasiliev, A.L., Padture, N.P., Park, S.-Y., and Berger, P.R. (2007) Characterization and electrical properties of individual Au–NiO–Au

- heterojunction nanowires. *IEEE Trans. Nanotechnol.*, **6**, 676.
62. Herderick, E.D., Tresback, J.S., Vasiliev, A.L., and Padture, N.P. (2007) Templated-directed synthesis, characterization and electrical properties of Au-TiO₂-Au heterojunction nanowires. *Nanotechnology*, **18**, 155204.
63. Oha, J., Taka, Y., and Leeb, J. (2005) Electrochemically deposited nanocolumnar junctions of Cu₂O and ZnO on Ni nanowires. *Electrochem. Solid-State Lett.*, **8** (6), C81–C84.
64. Sun, J.-L., Zhao, X., and Zhu, J.-L. (2009) Metal-insulator transition in Au-NiO-Ni dual Schottky nanojunctions. *Nanotechnology*, **20**, 455203.
65. Lyu, S.-H. and Lee, J.S. (2012) Highly scalable resistive switching memory cells using pore-size controlled nanoporous alumina template. *J. Mater. Chem.*, **22**, 1852–1861.
66. Wu, C.-Y., Wu, Y.-L., Wang, W.-J., Mao, D., Yu, Y.-Q., Wang, L., Xu, J., Hu, J.-G., and Luo, L.-B. (2013) High performance non-volatile memory devices based on Cu_{2-x}Se nanowires. *Appl. Phys. Lett.*, **103**, 193501.
67. Gibbons, J.F. and Beadle, W.E. (1964) Switching properties of thin NiO films. *Solid State Electron.*, **7**, 785.
68. Baek, I.G., Lee, M.S., Seo, S., Lee, M.J., Seo, D.H., Suh, D.-S., Park, J.C., Park, S.O., Kim, H.S., Yoo, I.K., Chung, U-In, and Moon, J.T. (2004) Highly scalable non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. IEEE IEDM Technical Digest, pp. 587–590.
69. Seo, S., Lee, M.J., Seo, D.H., Jeoung, E.J., Suh, D.-S., Joung, Y.S., Yoo, I.K., Hwang, I.R., Kim, S.H., Byun, I.S., Kim, J.-S., Choi, J.S., and Park, B.H. (2004) Reproducible resistance switching in polycrystalline NiO films. *Appl. Phys. Lett.*, **85**, 5655–5657.
70. Park, J.W., Park, J.W., Kim, D.Y., and Lee, J.K. (2005) Reproducible resistive switching in nonstoichiometric nickel oxide films grown by rf reactive sputtering for resistive random access memory applications. *J. Vac. Sci. Technol., A*, **23**, 1309.
71. Huang, Y.-C., Chen, P.-Y., Cin, T.-S., Liu, R.-S., Huang, C.-Y., and Lai, C.-H. (2012) Improvement of resistive switching in NiO-based nanowires by inserting Pt layers. *Appl. Phys. Lett.*, **101**, 153106.
72. Russo, U., Cagli, C., Spiga, S., Cianci, E., and Ielmini, D. (2009) Impact of electrode materials on resistive-switching memory programming. *IEEE Electron Device Lett.*, **30** (8), 817–819.
73. Seo, S., Lee, M.J., Kim, D.C., Ahn, S.E., Park, B.-H., Kim, Y.S., Yoo, I.K., Byun, I.S., Hwang, I.R., Kim, S.H., Kim, J.-S., Choi, J.S., Lee, J.H., Jeon, S.H., Hong, S.H., and Park, B.H. (2005) Electrode dependence of resistance switching in polycrystalline NiO films. *Appl. Phys. Lett.*, **87**, 263507.
74. Kim, S.I., Sa, Y.H., Kim, J.-H., Chang, Y.W., Kim, N., Kim, H., and Yoo, K.-H. (2014) Scaling behaviours for resistive memory switching in NiO nanowire devices. *Appl. Phys. Lett.*, **104**, 023513.
75. Lee, J.S., Lee, S.B., Chang, S.H., Gao, L.G., Kang, B.S., Lee, M.-J., Kim, C.J., Noh, T.W., and Kahng, B. (2010) Scaling theory for unipolar resistance switching. *Phys. Rev. Lett.*, **105**, 205701.
76. Yanagida, T., Nagashima, K., Oka, K., Kanai, M., Klamchuen, A., Park, B.H., and Kawai, T. (2013) Scaling effect on unipolar and bipolar resistive switching of metal oxides. *Sci. Rep.*, **3**, 1657.
77. Nagashima, K., Yanagida, T., Oka, K., Kanai, M., Klamchuen, A., Rahong, S., Meng, G., Horprathum, M., Xu, B., Zhuge, F., He, Y., Park, B.H., and Kawai, T. (2012) Prominent thermodynamical interaction with surroundings on nanoscale memristive switching of metal oxides. *Nano Lett.*, **12**, 5684–5690.
78. Dong, H., Zhang, X., Zhao, D., Niu, Z., Zeng, Q., Li, J., Cai, L., Wang, Y., Zhou, W., Gao, M., and Xie, S. (2012) High performance bipolar resistive switching memory devices based on Zn₂SnO₄ nanowires. *Nanoscale*, **4**, 2571–2574.
79. Qi, J., Huang, J., Paul, D., Ren, J., Chu, S., and Liu, J. (2013) Current self-compliant and self-rectifying resistive

- switching in Ag-electroded single Na-doped ZnO nanowires. *Nanoscale*, **5**, 2651–2654.
80. Son, J.Y. and Shin, Y.-H. (2008) Direct observation of conducting filaments on resistive switching of NiO thin films. *Appl. Phys. Lett.*, **92**, 222106.
 81. Jin, Z., Fukumura, T., Kawasaki, M., Ando, K., Saito, H., Sekiguchi, T., Yoo, Y.Z., Murakami, M., Matsumoto, Y., Hasegawa, T., and Koinuma, H. (2001) High throughput fabrication of transition-metal-doped epitaxial ZnO thin films: a series of oxide-diluted magnetic semiconductors and their properties. *Appl. Phys. Lett.*, **78**, 3824.
 82. Zhang, X., Zhang, L., Gao, M., Zhou, W., and Xie, S. (2010) High-resolution nanosphere lithography (NSL) to fabricate highly-ordered ZnO nanorod arrays. *J. Nanosci. Nanotechnol.*, **10**, 7432–7435.
 83. Zhang, Y., Hu, C., Zheng, C., Xi, Y., and Wan, B. (2010) Synthesis and thermoelectric property of Cu_{2-x}Se nanowires. *J. Phys. Chem. C*, **114**, 14849–14853.
 84. Sakamoto, T., Sunamura, H., Kawaura, H., Hasegawa, T., Nakayama, T., and Aono, M. (2003) Nanometer-scale switches using copper sulfide. *Appl. Phys. Lett.*, **82**, 3032.
 85. Yu, Y., Jiang, Y., Jiang, P., Zhang, Y., Wu, D., Zhu, Z., Liang, Q., Chen, S., Zhang, Y., and Jie, J. (2013) Large conductance switching nonvolatile memories based on p-ZnS nanoribbon/n-Si heterojunction. *J. Mater. Chem. C*, **1**, 1238.
 86. Chen, L., Xia, Y., Liang, X., Yin, K., Yin, J., Liu, Z., and Chen, Y. (2007) Nonvolatile memory devices with Cu₂S and Cu-Pc bilayered films. *Appl. Phys. Lett.*, **91**, 073511.
 87. Herderick, E.D., Reddy, K.M., Sample, R.N., Draskovic, T.I., and Padture, N.P. (2009) Bipolar resistive switching in individual Au–NiO–Au segmented nanowires. *Appl. Phys. Lett.*, **95**, 203505.
 88. Spiga, S., Brivio, S., Tallarida, G., Perego, D., Franz, S., Deleruyelle, D., and Muller, C. (2012) Resistive switching in NiO based nanowire array for low power ReRAM, in *Frontiers in Electronic Materials Technical Digest* (eds J. Heber, D. Schlom, Y. Tokura, R. Waser, and M. Wuttig), Wiley.
 89. Brivio, S., Perego, D., Tallarida, G., Bestetti, M., Franz, S., and Spiga, S. (2013) Bipolar resistive switching of Au/NiO_x/Ni/Au heterostructure nanowires. *Appl. Phys. Lett.*, **103**, 153506.
 90. Liang, C., Terabe, K., Hasegawa, T., Negishi, R., Tamura, T., and Aono, M. (2005) Ionic-electronic conductor nanostructures: template-confined growth and nonlinear electrical transport. *Small*, **1** (1), 971–975.
 91. Zheng, K., Shen, H., Li, J., Sun, D., and Chen, G. (2008) Fabrication and electrical properties of a Cu-tetracyanoquinodimethane nanowire array in a porous anodic alumina template. *Nanotechnology*, **19**, 015305.
 92. Prakash, A., Maikap, S., Rahaman, S.Z., Majumdar, S., Manna, S., and Ray, S.K. (2013) Resistive switching memory characteristics of Ge/GeO_x nanowires and evidence of oxygen ion migration. *Nanoscale Res. Lett.*, **8**, 220.
 93. Wu, Y., Xiang, J., Yang, C., Lu, W., and Lieber, C.M. (2004) Single-crystal metallic nanowires and metal/semiconductor nanowire heterostructures. *Nature*, **430**, 61.
 94. Scott, J.C. (2004) Is there an immortal memory? *Science*, **304**, 62.
 95. Owen, A.E., Le Comber, P.G., Spear, W.E., and Hajto, J. (1983) Memory switching in amorphous silicon devices. *J. Non-Cryst. Solids*, **59–60**, 1273.
 96. Ielmini, D., Cagli, C., and Nardi, F. (2009) Resistance transition in metal oxides induced by electronic threshold switching. *Appl. Phys. Lett.*, **94**, 063511.
 97. Ielmini, D., Nardi, F., and Cagli, C. (2011) Physical models of size-dependent nanofilament formation and rupture in NiO resistive switching memories. *Nanotechnology*, **22**, 254022.
 98. Goux, L., Lisoni, J.G., Wang, X.P., Jurczak, M., and Wouters, D.J. (2009) Optimized Ni oxidation in 80-nm contact holes for integration of forming-free and low-power Ni/NiO/Ni memory cells. *IEEE Trans. Electron Devices*, **56** (10), 2363–2368.

99. Cho, B., Kim, T.-W., Song, S., Ji, Y., Jo, M., Hwang, H., Jung, G.-Y., and Lee, T. (2010) Rewritable switching of one diode–one resistor nonvolatile organic memory devices. *Adv. Mater.*, **22**, 1228–1232.
100. Park, S., Lee, D.H., Xu, J., Kim, B., Hong, S.W., Jeong, U., Xu, T., and Russell, T.P. (2009) Macroscopic 10-terabit-per-square-inch arrays from block copolymers with lateral order. *Science*, **323** (5917), 1030–1033.
101. Xiao, S., Yang, X., Park, S., Weller, D., and Russell, T.P. (2009) *Adv. Mater.*, **21** (24), 2516–2519.
102. Frascaroli, J., Seguini, G., Spiga, S., Perego, M., Boarino, L. (2015) Fabrication of periodic arrays of metallic nanoparticles by block copolymer templates on HfO₂ substrates. *Nanotechnology*, **26**, 215301. (DOI: 10.1088/0957-4484/26/21/215301)
103. Chai, J., Wang, D., Fan, X., and Buriak, J.M. (2007) Assembly of aligned linear metallic patterns on silicon. *Nat. Nanotechnol.*, **2**, 500–506.
104. Nunn, A., Gwyther, J., and Manners, I. (2013) Inorganic block copolymer lithography. *Polymer*, **54**, 1269–1284.
105. Jo, S.H., Chang, T., Ebong, I., Bhadviya, B.B., Mazumder, P., and Lu, W. (2010) Nanoscale memristor device as synapse in neuromorphic system. *Nano Lett.*, **10**, 1297–1301.
106. Kuzum, D., Yu, S., and Wong, H.-S.P. (2013) Synaptic electronics: materials, devices and applications. *Nanotechnology*, **24**, 382001.
107. Chang, T., Jo, S.-H., and Lu, W. (2011) Short-term memory to long-term memory transition in a nanoscale memristor. *ACS Nano*, **5** (9), 7669–7676.
108. He, X., Yin, Y., Guo, J., Yuan, H., Peng, Y., Zhou, Y., Zhao, D., Hai, K., Zhou, W., and Tang, D. (2013) Memristive properties of hexagonal WO₃ nanowires induced by oxygen vacancy migration. *Nanoscale Res. Lett.*, **8**, 50.
109. Nirmalraj, P.N., Bellew, A.T., Bell, A.P., Fairfield, J.A., McCarthy, E.K., O'Kelly, C., Pereira, L.F.C., Sorel, S., Morosan, D., Coleman, J.N., Ferreira, M.S., and Boland, J.J. (2012) Manipulating connectivity and electrical conductivity in metallic nanowire networks. *Nano Lett.*, **12**, 5966–5971.

24

Switch Application in FPGA

Toshitsugu Sakamoto, S. Simon Wong, and Young Yang Liauw

24.1

Introduction

Resistive random-access memory (ReRAM) with nonvolatility and small footprint is greatly beneficial to programmable logic devices such as field programmable gate arrays (FPGAs). As reconfigurable hardware, FPGAs achieve programmability by utilizing a large number of programmable switches composed of static random access memories (SRAMs) and pass transistors. When the SRAM is replaced with ReRAM, the power consumption reduces, and the performance enhances dramatically.

The FPGA market has been growing since the invention by Xilinx in 1984 [1]. FPGAs have gone from being simple programmable logic devices to replacements for application-specific integrated circuits (ASICs). They offer many advantages including reduced nonrecurring engineering (NRE) cost and shorter time to market. The NRE cost of ASICs (i.e., mask cost and design cost) has been rapidly increasing as transistors are scaled down and circuit complexity increases. Furthermore, FPGAs reduce the NRE cost since the chip is already fabricated and the circuitry is already verified and characterized for mapping various applications. Moreover, FPGAs do not need lead time for layout, masks, or manufacturing steps, which are required for ASIC. Short time-to-market and rapid prototyping capability are essential due to shortened product lifetime cycles.

These advantages come at the cost of increased silicon area, delay, and power consumption compared with ASICs. The existence of these FPGA inefficiencies has been investigated by Kuon and Rose [2], who qualified the area, performance, and power-consumption gaps between ASICs and FPGAs. An FPGA is composed of configurable logic blocks (CLBs), routing, and IOs. A CLB includes multiple pairs of look-up tables (LUTs) and flip-flops (FFs), and the routing includes interconnects, buffers, and multiplexers (MUXes). Both the LUTs in a CLB and the MUXes in routing contain configuration memories (e.g., SRAMs). Kuon *et al.* compared the size of core programmable logic (namely, the silicon area except hard cores and IOs) with the size of ASICs having the same functionality by using

design automation tools. They concluded that the area difference is estimated to be 33 times on average for a 90-nm node.

The excess circuitry for the cause of programmability is the fundamental reason for the area gap. The area difference is reasonable when considering the area ratio ($= 33$) of functional logic to core programmable logic in FPGAs, as reported in Ref. [2]. The LUTs and FFs, which perform the basic logic operations in an FPGA, only take up 3.4% of the total core programmable logic area. The 3.4% figure is equivalent to the logic area difference between FPGA and ASIC. The area difference results in a signal delay gap of three to four times. Power consumptions of FPGA are also larger than those of ASIC. The power consumption ratio is similar for different applications, designs, and temperatures. There is about 14 times difference in dynamic power consumption, most of which consumes in the routing. As for static power, 88% of it is consumed by the SRAMs, MUXes, and LUTs [3]. Gate leakage or subthreshold leakage in transistors is a dominant source for static power. To reduce gate leakage the gate oxide should be thicker, and to reduce the subthreshold leakage the threshold voltage should be higher. In turn, however, the transistor performance becomes worse. The gate thickness and the subthreshold are optimized in terms of the toggle rate of the transistors, resulting in mitigating the power consumption without degrading the performance [4].

24.2

Monolithically 3D FPGA with BEOL Devices

Stacking the configuration memory over the core programmable logic significantly increases the area ratio of functional-block and reduces interconnect lengths (Figure 24.1) [5]. The area gap between FPGA and ASIC can be explained by the fact that the functional logic area only takes up 3.4% of the programmable logic area, as discussed in the previous section. The rest of the area is occupied by

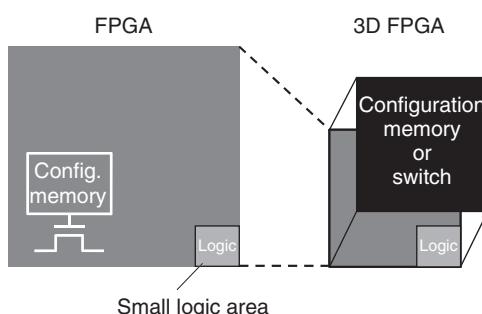


Figure 24.1 Conventional FPGA and 3D monolithic FPGA. Configuration memory and routings occupy large area in conventional FPGA. Area reduction is realized by three-dimensional FPGA, where

configuration memory or pass transistor is integrated in Cu interconnects. Concept of “Memory or switch over logic” by using resistive-change memory has been proposed.

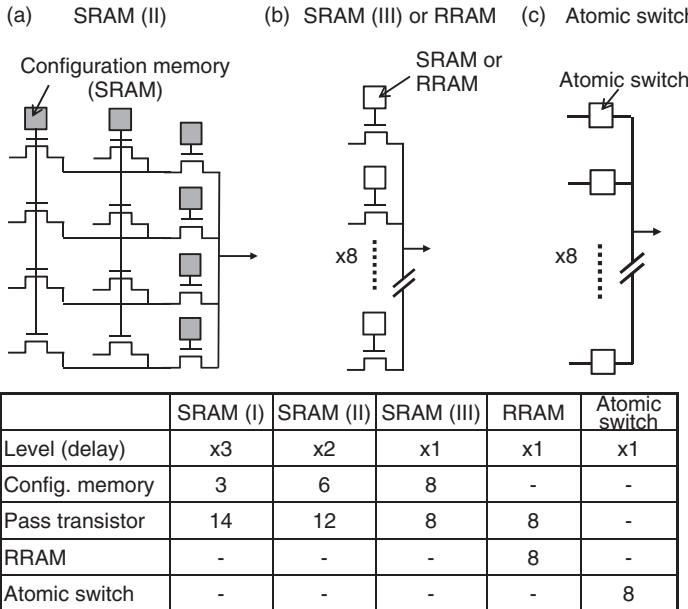


Figure 24.2 An 8-to-1 routing multiplexer (MUX) composed of (a) SRAM, (b) ReRAM or SRAM, and (c) atomic switch. Bottom table shows number of elements required for 8-to-1 MUX.

SRAM, interconnects, a MUX, and buffers. A particularly large portion of the rest of the area is occupied by the SRAM. Lin *et al.* estimated that SRAM occupies up to 43% of an FPGA [5]. Replacing the SRAM or MUX with a nonvolatile and compact memory has been an attractive approach to reducing the area gap, as shown in Figure 24.1. A nonvolatile memory such as FeRAM [6], PRAM [7], MRAM [8, 9], or ReRAM [10], formed in the metal interconnect layers (or the back-end of line (BEOL)), has been proposed to replace the SRAM in an FPGA. An atomic switch [11, 12], which works as a programmable switch, also replaces an SRAM and MUX.

Usage of compact configuration memory in a MUX can also improve the signal delay by reducing both wire length and switch levels in MUX. Figure 24.2 shows an 8-to-1 MUX. The MUX in the conventional FPGA is composed of dozens of SRAMs and pass transistors. Since the area occupied by SRAM ($\sim 200F^2$) is much larger than that by a pass transistor ($\sim 30F^2$), the usage of SRAM is controlled, where F is the minimum feature size. The three-level structure (SRAM(I) in Figure 24.2), which is the most area-efficient, results in increased delay. To achieve a compromised deal between area and delay, a typical 8-input routing MUX is realized by a 2-level structure comprising 12 pass transistors sharing 6 SRAMs (SRAM(II)). When the compact configuration memory (ReRAM [10] or atomic switch [11, 12]) in the BEOL is applied, the configuration bit size is negligible and a single-level structure, which has minimum delay, becomes possible.

The nonvolatility also saves the stand-by power arising from the leakage current in SRAM in the conventional FPGA. These technologies promise to close the FPGA-ASIC performance gap.

A nonvolatile LUT using MRAM or PRAM has been proposed by several groups. The nonvolatile two-input LUT proposed in Ref. [9] is composed of four spin-transfer torque type MRAMs. It also has a combinational logic circuit, a sense amplifier, a selector transistor, and a reference cell. An FPGA with a nonvolatile LUT was fabricated by a 0.14 μ m process and achieved 33% area reduction. A nonvolatile three-input LUT combined with a PRAM-based configuration memory for both LUT and MUX was also demonstrated [7].

An ReRAM-based configuration memory for both LUT and MUX was also demonstrated [10]. The memory cell consists of two resistive switching elements and a shared select transistor. One of the two elements is set to the ON state and the middle node drives the pass transistor. The author of [10] estimated that the programmable cell area is reduced by 40% for a 65-nm node. An FPGA with a resistive switching element will be covered in Section 24.3, while further area reduction is achieved by replacing both the configuration memory and the pass transistors with an atomic switch [11, 12]. The atomic switch plays both memory and switch roles since it has a large ON/OFF conductance ratio. This aspect is covered in Section 24.6.

24.3

Resistive Memory Replacing Configuration Memory

The area efficiency problem in conventional FPGAs is to replace the SRAM-based configuration memory with resistive-switching-element-based configuration memory that stacked directly on top of FPGA's programmable logic and routing.

24.3.1

Architecture

The resistive-switching-element-based FPGA can be viewed as having two functional layers: the CMOS logic layer at the bottom and the configuration memory layer on top. The CMOS layer (see Figure 24.2) contains programmable logic and peripheral circuitry. It consists of an array of tiles, row and column decoders and programming drivers for the configuration memory, and configurable I/O blocks (IOBs). Each tile comprises a CLB with a cluster of LUTs, horizontal and vertical segmented routing channels, two connect blocks (CBs) for interfacing the CLB to the channels, and a switch block (SB) for signal routing between channels. The LUTs and routing switches are implemented using pass gate MUX controlled by configuration memory cells. The configuration memory layer comprises an array of resistive switching elements that are stacked on top of the transistors and interconnects. Additional interconnect layers are used to connect the configuration memory cells to the programmable resources underneath.

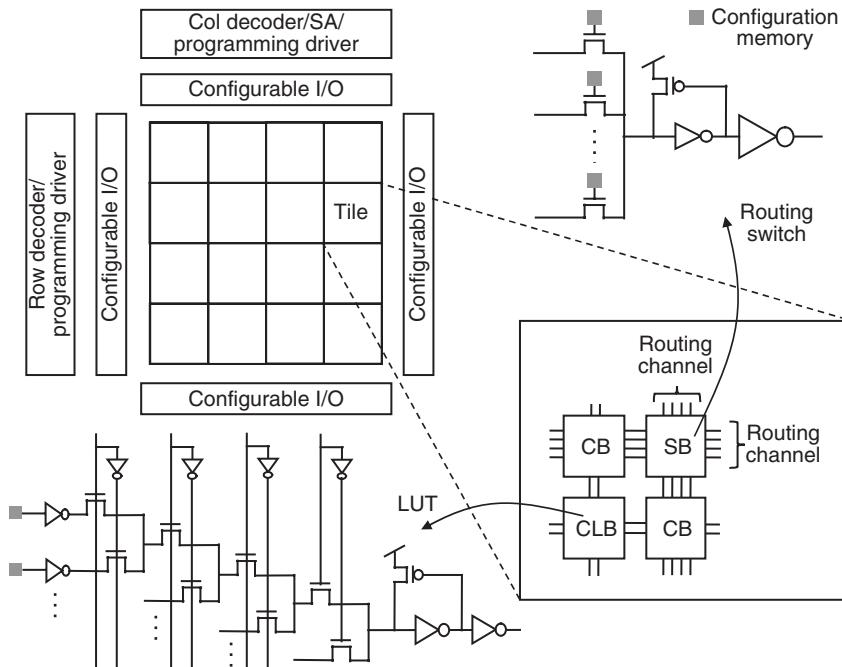


Figure 24.3 Architecture of 3D-FPGA's CMOS layer.

24.4

Resistive Configuration Memory Cell

A new configuration memory cell was proposed that consists of two resistive switching elements and a shared access transistor (referred to as 1T2R cell) [10]. The cell's schematic is shown in Figure 24.3. The configuration bit is stored in both switching elements. During normal FPGA operation, the access transistor is turned off, and the pair of bit lines, LBL and RBL, is connected to power supply (VDD) and ground (GND), respectively. The two resistive switching elements then form a resistive divider between the bit lines. The middle node's voltage represents the configuration bit. The two switching elements in the cell are programmed to opposite resistance states to prevent undetermined logic. If the cell controls a routing switch, the configuration bit can be directly connected. If the cell drives a LUT, the configuration bit needs to be buffered. Although the configuration bit can be stored in a single resistive switching element, additional transistors would be needed to turn the passive information into active static logic.

The resistive switching element needs to exhibit certain electrical and physical characteristics for the new configuration memory cell to be efficient and effective. The switching element needs to be sandwiched at the crosspoint where two metal wires of different layers intercept. The programming current also needs to be low, in few microamperes range and free of electroforming, such that the

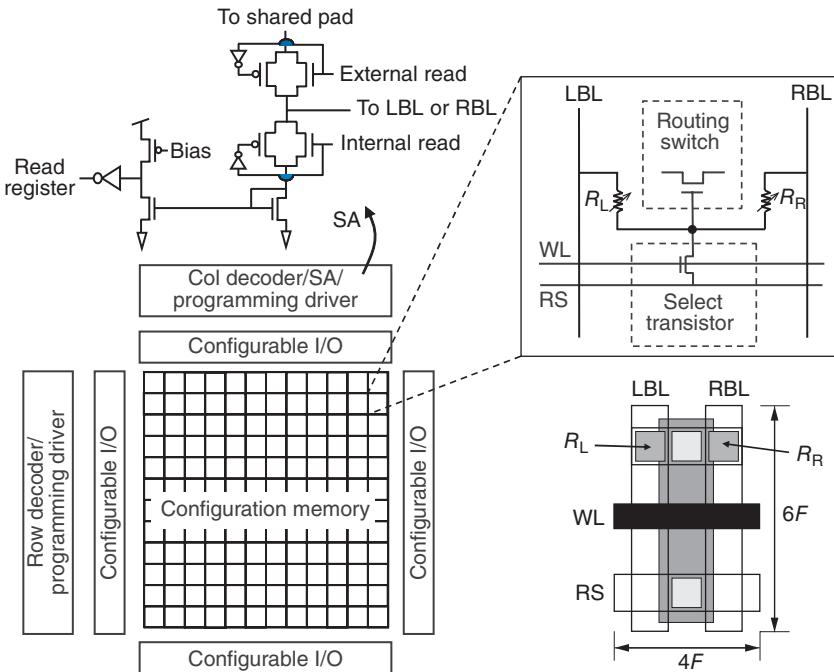


Figure 24.4 Architecture of 3D-FPGA's configuration memory layer.

selection transistor can be made small enough to conduct the necessary current. The element needs to exhibit bipolar switching behavior so that common electrodes, such as the bottom electrode, can be shared. These factors allow the 1T2R cell to be laid out as compact as $24F^2$ (see Figure 24.4), where F is the minimal feature size of the interconnect wire. The switching element also needs to exhibit >10 resistance ratio between opposite states so that there is sufficient headroom to resolve the logic level in the resistive divider. It also needs to exhibit $>1G\Omega$ resistance in the off state to limit the static current through the resistive divider. The switching element needs to be able to withstand DC stress at level near FPGA's operating voltage. It also needs to exhibit long retention period so that the resistive divider can operate reliably. The fabrication process needs to be compatible with standard CMOS and BEOL process. AlO_xN_y -based resistive switching element is one of many that appear to meet these requirements [13].

24.5

Resistive Configuration Memory Array

Similar to conventional SRAM-based configuration memory, the resistive switching element-based configuration memory has several access modes to support

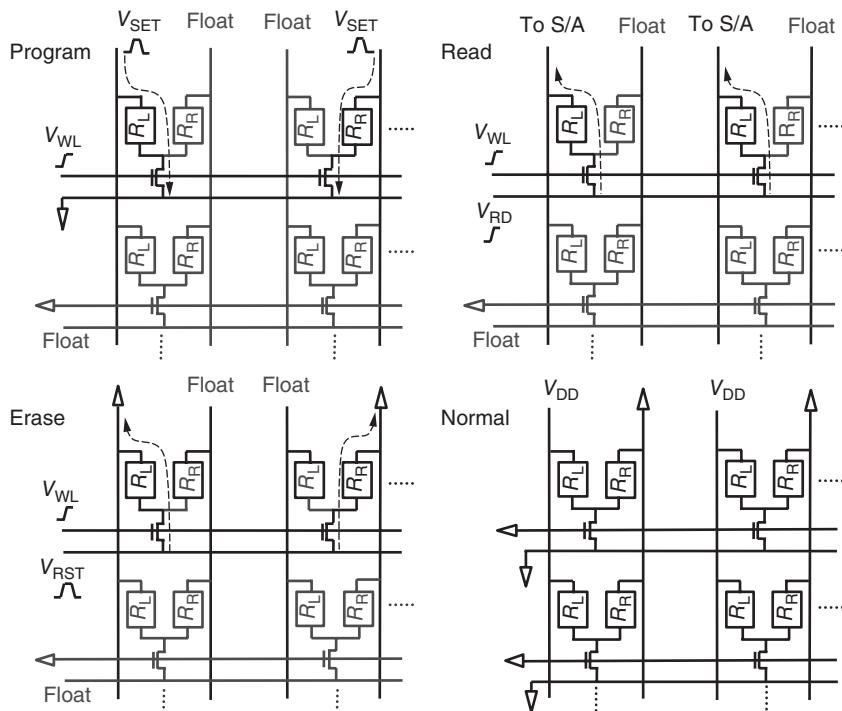


Figure 24.5 Four operation modes of the resistive configuration memory.

FPGA's operation. The memory array is organized to support the switching element's bipolar switching behavior.

The left bit line (LBL) and right bit line (RBL) can be connected to programming voltages, VDD, or GND. The source nodes of select transistors are connected to a shared row line signal (RS), which can be connected to programming voltages or GND. Because pristine resistive switching elements start in the high-resistance state, the stored configuration logic is unknown. One of the two switching elements must be set to the low-resistance state for the resistive divider to function properly. To write logic “1” into the fresh 1T2R cell at the top left, as shown in the program diagram in Figure 24.5, RL is set to the low resistance state. After word line is asserted and RS is connected to GND, a VSET pulse is applied to LBL, while RBL is left floating. The set current flows from the top to bottom of RL. The select transistor limits the current as RL changes from HR to LR. Similarly, to write “0” into the top right 1T2R cell, RR is set to the LR state by floating LBL and applying the VSET pulse to RBL.

A written cell must be erased and restored to the unknown state before the opposite logic can be written. Resetting the switching element back to high-resistance state erases the cell's logic. To reset RL in the top left 1T2R cell to the high-resistance state, as shown in the erase diagram in Figure 24.5, LBL is connected to GND, and RBL is left floating. After word line is asserted, a VRST

pulse is applied to RS. The reset current flows from the bottom to the top of RL. Similarly, to reset RR in the top right 1T2R cell, RBL is grounded and LBL is left floating. Both RL and RR can be read to verify the memory cell's logic state. The read mode is similar to the erase mode. A VRD pulse (< VRST) is applied to RS and the selected bit line is connected to a current mirror-based sense amplifier that detects the resistance. During normal FPGA operation, all word lines and RSs are connected to GND, and LBLs and RBLs are connected to VDD and GND, respectively. This activates the resistive divider logic in each 1T2R cell. The configuration bits that drive LUTs are buffered to prevent accidental programming. Bits that configure routing switches are not buffered, however. The configuration memories are not in the signal paths during normal FPGA operation.

24.5.1

Prototype

The concept is demonstrated via an FPGA prototype fabricated in 180 nm process that integrates AlO_xN_y -based resistive switching elements over CMOS logic. A micrograph of the FPGA prototype is shown in Figure 24.6. The FPGA follows the architecture outlines in Figures 24.3 and 24.4. The 289 logic tiles are organized into a 17×17 array. The 21-Kb configuration memory is organized into a 144×152 array and is stacked on top of CMOS logic. The test chip measures $2.5 \text{ mm} \times 2.5 \text{ mm}$. The total active area measures $1.72 \text{ mm} \times 1.84 \text{ mm}$.

Transistors and the lower metal interconnect layers (M1–M5) are fabricated first in standard CMOS process. The resistive switching elements, each measuring $2 \mu\text{m} \times 2 \mu\text{m}$, are monolithically integrated on top of the wafer subsequently.

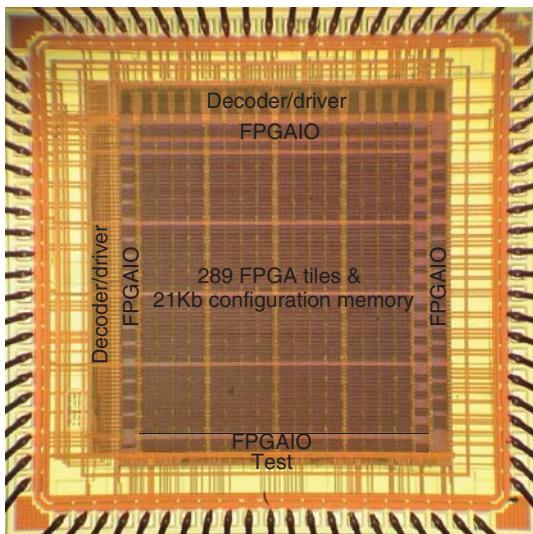


Figure 24.6 Prototype micrograph.

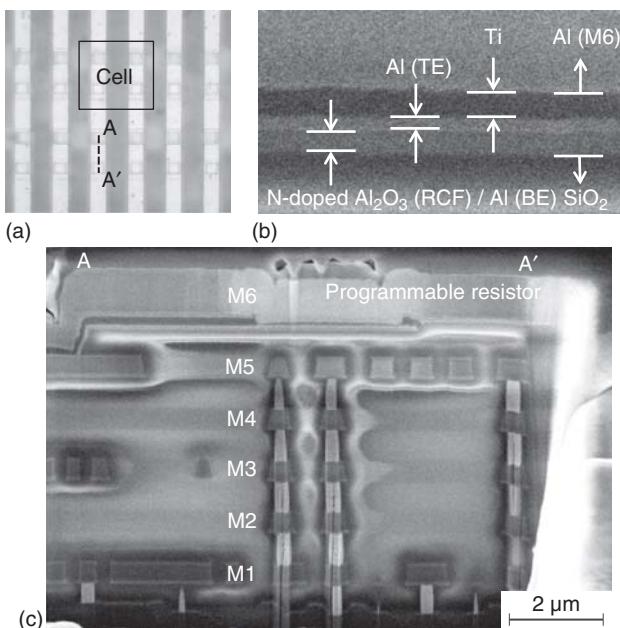


Figure 24.7 Top (a) and cross-section (b, c) views of bit lines in prototype.

Afterward, a sixth metal layer (M6) is deposited. M6 is used as the bit lines for the configuration memory array, as well as connecting the array to the programming circuits at the edges. Fabrication and integration of the resistive switching elements require four additional photolithographic masks. Figure 24.7a shows the top view of a configuration memory cell from a finished test chip. The visible vertical wires are BLs in M6. The two squares along the BL are vias that connect the resistive switching elements to M5 (at marking A) and M6 (at marking A'). A scanning electron microscopy (SEM) image of the corresponding cross-section of the memory cell along the BL direction, from A to A', is shown in Figure 24.7b. The height of the switching element is about 100 nm. A white rectangle is used to show its vertical location in the figure. A close-up of the switching element's cross section is given in Figure 24.7c with annotated layers. The thickness of the resistance change film is about 10 nm.

24.5.2

Measurement Results

The programmable resistors are tested and measured first. Shown in Figure 24.8 are measured $I - V$ curves of the resistor during program (Figure 24.8a) and erase (Figure 24.8b). Afterward, configuration bits are written, erased, and rewritten to test the memory's reconfigurability, as shown in Figure 24.9.

To test the FPGA's programmable logic functionality, benchmark circuits are mapped into the test chip. The mapped circuits also provide a way to extract

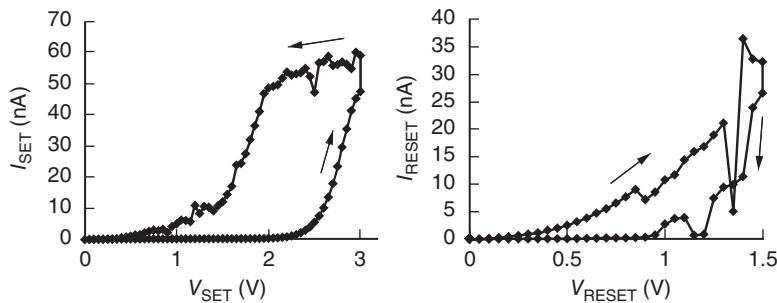


Figure 24.8 Measured I - V curves of a programmable resistor in program and erase mode.

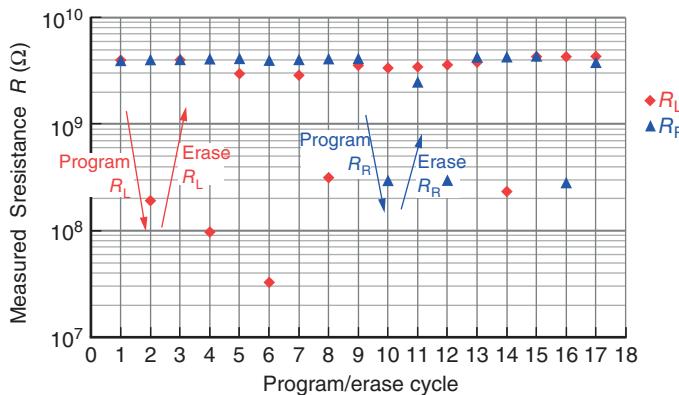


Figure 24.9 Measured resistances of both programmable resistors in a configuration memory cell before and after a sequence of program and erase operations.

performance metrics of the FPGA, such as maximum operating frequency (F_{max}) and power consumption. The mapping process contains two steps. First, benchmark circuits are converted into configuration bit stream. Next, logic analyzer then loads the bit stream into the test chip.

Two benchmark circuits are mapped into the test chip. The first one is a 4-bit linear feedback shifter register (LFSR), which tests the minimum delay between adjacent CLBs. The FPGA's configuration is shown in Figure 24.10a. The measured maximum operating frequency of the LFSR is 250 MHz at 1.8 V, and the test chip consumes 39.6 mW while operating at this frequency. This power includes the power of the FPGA, on-chip tester and IO buffers as they all shared the same supply. Figure 24.8b shows the measured supply voltage versus minimum LFSR delay. The second benchmark circuit is a 4-bit adder. Eight CLBs are used to implement the adder. The CLB placement is shown in Figure 24.10c. The measured minimum carryout delay is 9 ns at 1.8 V. The test chip consumes 46.8 mW while operating at 100 MHz.

To examine the benefits of resistive-switching-element-based configuration memory, an SRAM-based FPGA tile and a resistive-switching- element-based

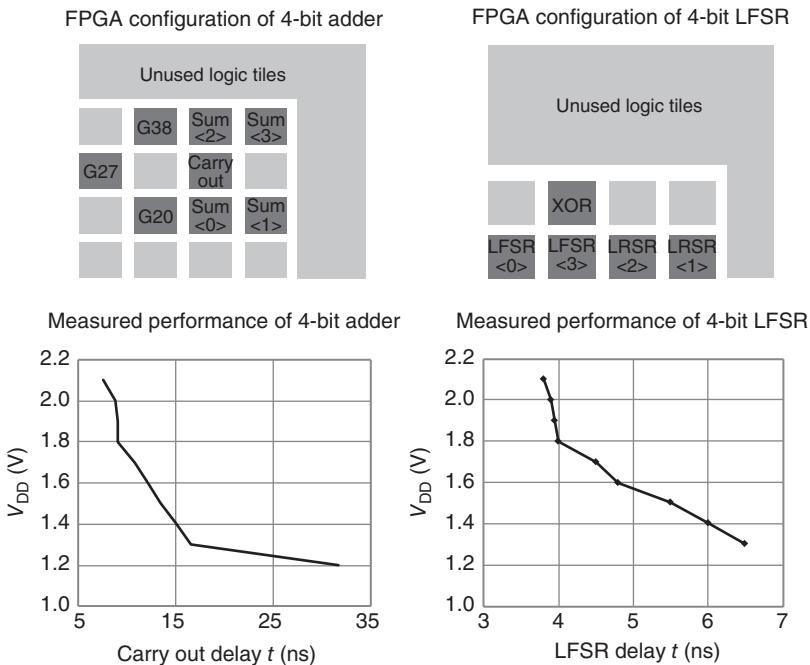


Figure 24.10 CLB placement of a mapped 4-bit LFSR (upper left) and its measured minimum LFSR delay (lower left); CLB placement of a mapped 4-bit adder (upper right) and its measured carryout delay (lower right).

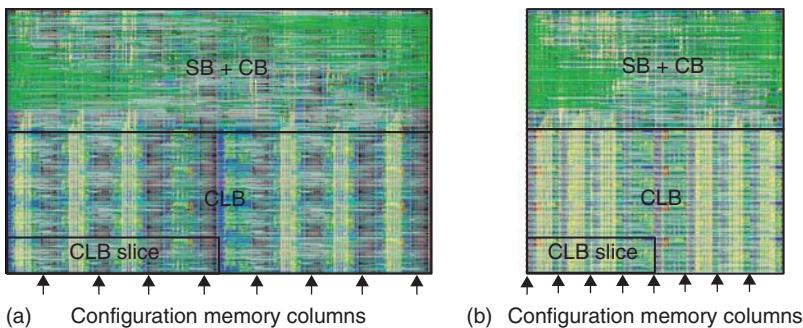


Figure 24.11 Layout of SRAM-based FPGA tile (a) and ReRAM-based FPGA tile (b).

FPGA tile are laid out and compared, as shown in Figure 24.11. Both tiles are laid out in 65 nm CMOS process with seven metal layers and implement a cluster of 8 six-input LUTs per CLB and 96-track segmented routing channels. The resistive-switching-element-based tile achieves 40% area reduction and 28% energy-delay-product reduction compared with the SRAM-based tile. The area saving is due to the reduced memory cell footprint and downsized routing switches.

24.6

Complementary Atomic Switch Replacing Configuration Switch

24.6.1

Complementary Atomic Switch (CAS)

Another approach to enhance the performance of FPGAs is an atomic switch. Since the atomic switch has a large ON/OFF, it replaces the programmable element, which is composed of SRAM and pass transistor (Figure 24.2c). The atomic switch satisfies all the necessary requirements of the programmable switch, namely (i) nonvolatility, (ii) high ON/OFF conductance ratio, (iii) OFF-state reliability against DC voltage stress, (iv) ON-state reliability, and (v) low programming voltage. To meet these requirements, the complementary atomic switch (CAS) is proposed. The CAS consists of two atomic switches connected in series in opposite direction (Figure 24.12) [14]. For the ON state of the CAS, both of the two atomic switches are set to ON state, and for the OFF state both are programmed to be OFF state (Figure 24.12b). The CAS has a large ON/OFF conductance ratio of 5 orders of magnitude (Figure 24.12c), which is suitable

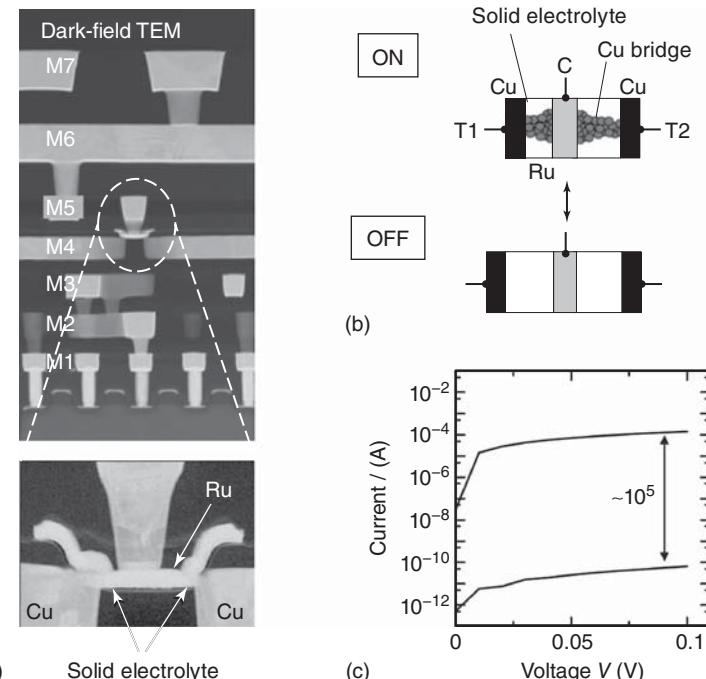


Figure 24.12 Features of complementary atomic switch (CAS). (a) Cross-sectional TEM image of CAS formed between M4 and M5 layers. (b) ON or OFF-state CAS. (c) Current–voltage characteristics of ON- or OFF-state CAS.

Each element turns on or off when voltage is applied between Ru and Cu electrodes.

for a programming switch. Each switch element is sequentially programmed by applying a voltage between Cu and Ru electrodes. Cu is an active electrode and Ru is an inert electrode. When a positive voltage is applied to the Cu electrode relative to the Ru electrode, a Cu bridge is formed in the solid electrolyte sandwiched between Cu and Ru [15]. When a positive voltage is applied to the Ru (or the switch is reversely biased), the switch maintains OFF state due to its bipolar nature until breakdown occurs.

The CAS also provides both high OFF-state reliability and low programming voltage. The programmable switch in FPGA is required to maintain its ON or OFF state during logic operations. The OFF-state switch should be durable against VDD (~ 1 V) stress during the operation. However, the OFF-state switch should be turned on immediately when the set voltage is applied to the CAS for the set operation. To achieve a highly reliable OFF state, the set voltage ($>VDD$) was controlled to be high (i.e., higher than 3.5 V) for a conventional two-terminal switch [12]. Hence, there is a trade-off between the high OFF-state reliability and the low set voltage. The CAS can break through the trade-off advantageously. Either of two elements of the CAS, namely a reverse-biased element, is highly durable against a stress voltage, since the breakdown voltage of the element is much larger than the set voltage [14]. When the stress voltage is applied to the OFF-state CAS, the voltage is mainly sustained by the reverse-biased element. It has been confirmed that the CAS has 10-year retention under a stress voltage of 1 V and the set voltage is reduced to be 2 V. Lowering the set voltage enhances scalability of CLB. The design with a 1.8 V-tolerant transistor for the programming transistor [16] leads to a cell layout reduced by half that of the 3.3 V-tolerant one [12].

The CAS is integrated on a 65-nm-node CMOS with 1P7M process (Figure 24.12a) [16]. The embedded process of the CAS is fully compatible with CMOS, since it was confirmed that no degradation occurs in CMOS transistor and interconnects. The CAS is integrated on fourth metal (M4) interconnects [16]. A polymer-solid electrolyte was used, thus allowing for a forming free operation and high breakdown voltage [17, 18].

24.6.2

Cell Architecture with CAS

The CAS-based CLB consists of two pairs of four-input LUTs and FFs [16]. Each data memory element in the LUTs is composed of a pair of CAS (Figure 24.13a). Each CAS is in ON or in OFF state alternately to output logic “1” or “0”. A 64-bit CAS is used for the data memory element of LUTs in CLBs. The function block is connected to the 19×16 crossbar switch by vertical and horizontal wires. Each column of the crossbar is composed of 19-bit CAS connected in parallel (Figure 24.13b), which is equivalent to a 19-to-1 MUX in SRAM-based CLB. A 304-bit CAS is used in the 19×16 crossbar switch. As discussed in the previous section, the architecture of the routing MUX must be highly optimized due to the trade-off between the area consumption and power and/or delay cost. This routing MUX can be replaced by multiple CAS to achieve a single-level structure.

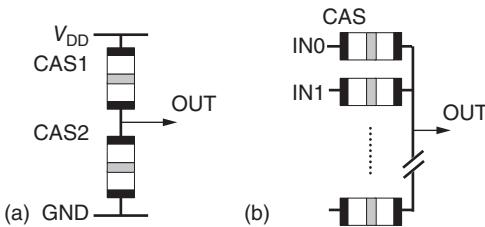


Figure 24.13 (a) Memory bit in LUT composed of two CAS. (b) MUX in routing with CAS.

To reduce a signal delay, it is essential to minimize the capacitance downstream of the switches. The delay can be improved due to the small input capacitance of CAS ($\sim 0.14 \text{ fF}$) and the single-level switch, although the fan-in of the cross-bar switch is large (~ 19). Moreover, due to the elimination of the configuration memory required for the SRAM-based CLB, the CAS-based MUX consumes significantly less area and this leads to small signal delay. To clarify the area benefit of a CAS-based cell, a conventional SRAM-based cell is also prepared. The programmable logic cell in both designs has equivalent routing and logic resources (i.e., each cell has two 4-bit LUTs and 16 directional wires). The SRAM-based cell utilizes standard-cell-based MUXes for the signal routing and a D-Latch for the configuration memory, which is suitable for low-voltage operation. Compared with an SRAM-based reference design, the area of the CAS-based cell is reduced by 78% (Figure 24.14).

- 65 nm 1P7M CMOS
- cell : 2x4 input LUT
- switch : multiplexer
- logic : core Tr

- 65 nm 1P7M CMOS
- cell : 2x4 input LUT
- CAS : 368 bit/cell
- logic : core Tr
- selector : HV 18

46.46 μm

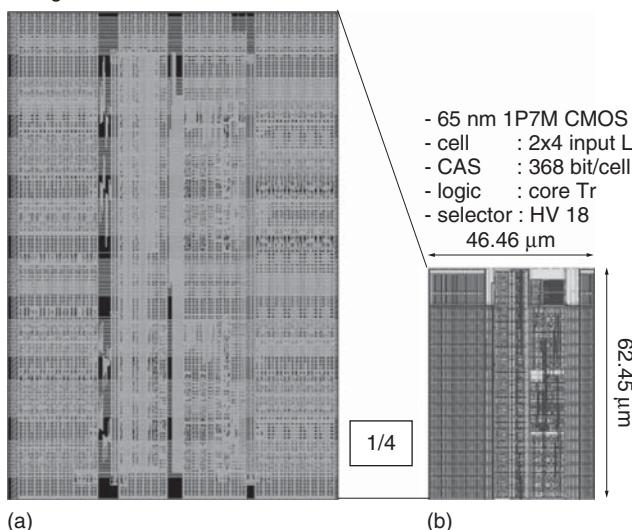


Figure 24.14 Tile layouts of SRAM-based programmable logic (a) and CAS-based logic (b).

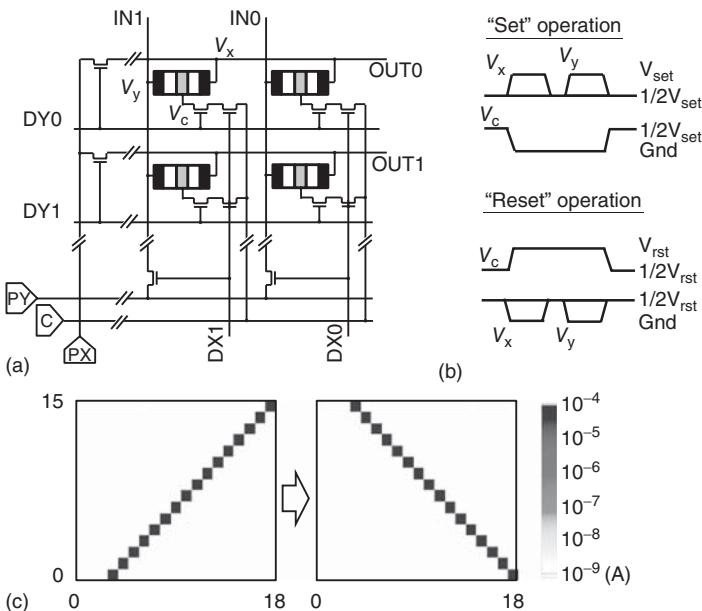


Figure 24.15 16×19 CAS-based crossbar. (a) Circuit and (b) biasing condition for set or reset operation. (c) ON/OFF current map after programming.

Figure 24.15 shows the schematic of the 16×19 crossbar circuit and the current map after programming. A CAS is placed at each crosspoint and the programming voltage is applied by programming drivers (PX , PY , and C) via select transistors. These drivers and the decoders are placed on the periphery of the logic cell array. The unit cell has a 2T2R type structure by sharing the last stage of select transistors [16]. In this schematic, the target control node of the CAS is uniquely asserted by the DX and DY decode signals, resulting in quite high immunity to write disturb compared with the half-selected programming scheme [12]. A high ON/OFF conductance ratio in the crossbar is imperative for signal integrity. In Figure 24.15c, the diagonal elements in the crossbar are programmed to be the ON state and the others are to be the OFF state. When a signal is transferred to a particular input, the signal via the programmed path is observed and there is no cross talk between the other paths. Moreover, the crossbar shows an excellent reprogrammability. Once the diagonal element is set as the OFF state by applying the reset voltage, the crossbar switch is correctly reprogrammed to be an opposite diagonal, as shown in Figure 24.15c. The high ON/OFF ratio after reprogramming indicates that the CAS is applicable for reprogrammable switches.

24.6.3

Demonstration of CAS-Based Programmable Logic

The 3×3 CAS-based programmable logic cell arrays are demonstrated. Various logics are realized by synthesizing resistor transistor logic (RTL) codes and

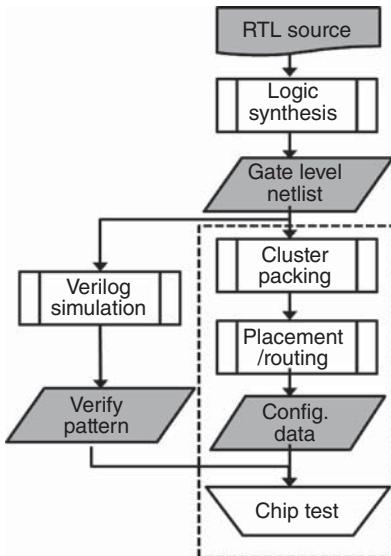


Figure 24.16 In-house mapping tool chain.

mapping the configurations into a CAS-based programmable cell array. The configuration data are obtained from an arbitrary RTL description utilizing in-house cluster packing and a placement/routing tool chain (Figure 24.16). A 2-bit full adder and 4-bit synchronous counter are synthesized and configured. The programmed CASs in the crossbar and the LUT are in ON state, whereas the others stay in high-resistive state. The output pattern of the programmable is successfully confirmed to be correct by comparing it with the expected pattern generated by the tool chain (Figure 24.17).

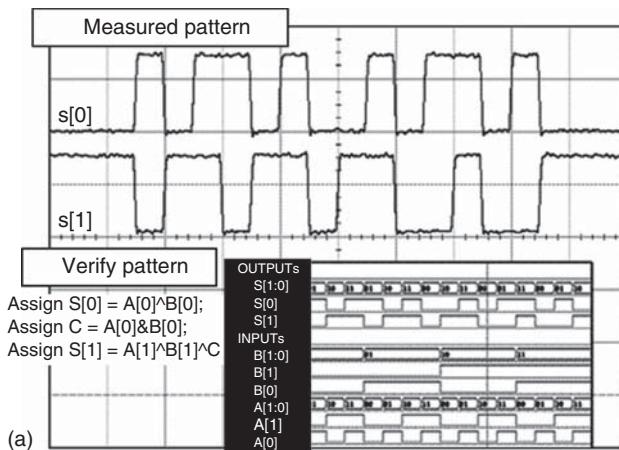
Dynamic power is a critical issue for a conventional programmable logic. Compared with the dynamic power of the SRAM-based cell in Figure 24.14, the CAS-based cell exhibits 61% reduction. This reduction originated from a low input capacitance in CAS and shorter wiring length owing to reduced cell area. CAS is a promising candidate to realize a nonvolatile programmable logic with low power consumption and high performance.

24.7

Energy Efficiency of Programmable Logic Accelerator

In addition to providing good energy efficiency, the atomic-switch-based programmable logic has nonvolatility. This achieves low standby power and instant on. These features are promising for energy constraint systems but requiring computational intensive processing. One of the applications of nonvolatile programmable logic is that of an off-loader for wireless sensor nodes (WSNs). Energy saving is one of the most important issues for battery-powered WSNs.

2 bit adder



4 bit counter

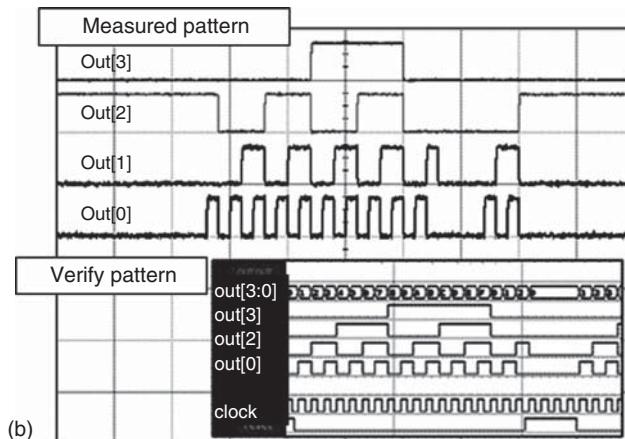


Figure 24.17 Measured output waveform and verification pattern of (a) 2-bit adder and (b) 4-bit counter.

Nonvolatility is essential for the intermittent operation of WSNs due to the requirement for power-on without loading configuration data.

A data compression algorithm for WSN is implemented in both atomic-switch-based programmable logic and low-power CPU to evaluate the power efficiency. Data compression is desired for WSNs since it reduces energy consumption during data transmission, but needs intensive processing in a low power CPU. A power simulation conducted using EDA tools shows the atomic-switch-based programmable logic can execute at 21 times faster than a CPU with comparable average power consumption. This means that in terms of power efficiency the atomic-switch-based programmable logic operates 22 times better the CPU [19].

24.8

Conclusion and Outlook

FPGAs have become key devices for communication network, industrial applications, and Internet-of-Things applications. Although FPGAs have several advantages against ASIC, the excess circuitry such as configuration memory and pass transistor for the reconfigurability introduces a significant performance gap with ASIC. When the configuration memory is replaced with ReRAM, the chip area and power consumption reduces dramatically.

The ReRAM-based programmable cell, where the configuration memory is replaced by two ReRAMs, reduces the cell area by 40% and the energy delay product by 28%, compared with the SRAM-based cell. The area saving is due to the reduced memory cell footprint and downsized routing switches.

Low-power programmable-logic array, which utilizes CAS instead of SRAM and pass transistor, is also demonstrated here. Compared with the conventional cell design using CMOS routing multiplexer (MUX), the proposed programmable-logic array exhibits 61% active power saving and 78% area reduction. A 2-bit full adder and a 4-bit synchronous counter are successfully mapped on 3×3 programmable logic cells using with CAS architecture.

Large-scale implementation of ReRAM-based programmable logic cell is necessary to clarify its feasibility in FPGA. The variation in the switching properties of ReRAM is the most important issue for this demonstration.

The 3D-FPGA presented here is only one of the many applications of ReRAM based switching elements. Integration of resistive switching elements with CMOS will enable a whole new range of innovative designs.

References

1. Carter, W.S., Duong, K., Freeman, R.H., Hsieh, H.C., Ja, J.Y., Mahoney, J.E., Ngo, L.T., and Sze, S.L. (1986) A user programmable reconfigurable logic array. *IEEE Custom Integrated Circuit Conference*, **15** (6), pp. 233–235.
2. Kuon, I. and Rose, J. (2007) Measuring the gap between FPGAs and ASICs. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, **26** (2), 203–215, doi: 10.1109/TCAD.2006.884574.
3. Tuan, T. and Lai, B. (2003) Leakage power analysis of a 90nm FPGA. *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*, pp. 57–60, doi: 10.1109/CICC.2003.1249359.
4. Telikepalli, A. (2005) Power vs. Performance: the 90 nm inflection point. Xilinx White Paper, WP223 (v1.1).
5. Lin, M., El Gamal, A., Lu, Y.C., and Wong, S. (2007) Performance benefits of monolithically stacked 3-D FPGA. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, **26** (2), 216–229, doi: 10.1109/TCAD.2006.887920.
6. Koga, M., Iida, M., Amagasaki, M., Ichida, Y., Saji, M., Iida, J., and Sueyoshi, T. (2010) First prototype of a genuine power-gatable reconfigurable logic chip with FeRAM cells. *Field Programmable Logic and Applications (FPL), 2010 International Conference on*, pp. 298–303, doi: 10.1109/FPL.2010.67.
7. Wen, C.Y., Li, J., Kim, S., Breitwisch, M., Lam, C., Paramesh, J., and Pileggi, L. (2011) A non-volatile look-up table design using PCM (phase-change memory) cells. *VLSI Circuits (VLSIC), 2011 Symposium on*, pp. 302–303.

8. Sekikawa, M., Kiyoyama, K., Hasegawa, H., Miura, K., Fukushima, T., Ikeda, S., Tanaka, T., Ohno, H., and Koyanagi, M. (2008) A novel SPRAM (Spin-transfer torque RAM)-based reconfigurable logic block for 3D-stacked reconfigurable spin processor. *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, pp. 1–3.
9. Suzuki, D., Natsui, M., Ikeda, S., Hasegawa, H., Miura, K., Hayakawa, J., Endoh, T., Ohno, H., and Hanyu, T. (2009) Fabrication of a non-volatile lookup-table circuit chip using magneto/semiconductor-hybrid structure for an immediate-power-up field programmable gate array. *VLSI Circuits, 2009 Symposium on*, pp. 80–81.
10. Liauw, Y.Y., Zhang, Z., Kim, W., Gamal, A., and Wong, S. (2012) Nonvolatile 3D-FPGA with monolithically stacked RRAM-based configuration memory. *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, pp. 406–408, doi: 10.1109/ISSCC.2012.6177067.
11. Kaeriyama, S., Sakamoto, T., Sunamura, H., Mizuno, M., Kawaura, H., Hasegawa, T., Terabe, K., Nakayama, T., and Aono, M. (2005) A nonvolatile programmable solid-electrolyte nanometer switch. *IEEE J. Solid-State Circuits*, **40** (1), 168–176, doi: 10.1109/JSSC.2004.837244.
12. Miyamura, M., Nakaya, S., Tada, M., Sakamoto, T., Okamoto, K., Banno, N., Ishida, S., Ito, K., Hada, H., Sakimura, N., Sugabayashi, T., and Motomura, M. (2011) Programmable cell array using rewritable solid-electrolyte switch integrated in 90nm CMOS. *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 228–229, doi: 10.1109/ISSCC.2011.5746296.
13. Kim, W., Park, S.I., Zhang, Z., Yang-Liauw, Y., Sekar, D., Wong, H.S.P., and Wong, S. (2011) Forming-free nitrogen-doped AlO_x RRAM with sub μA programming current. *VLSI Technology (VLSIT), 2011 Symposium on*, pp. 22–23.
14. Tada, M., Sakamoto, T., Miyamura, M., Banno, N., Okamoto, K., Iguchi, N., Nohisa, T., and Hada, H. (2011) Highly reliable, complementary atom switch (CAS) with low programming voltage embedded in Cu BEOL for non-volatile programmable logic. *Electron Devices Meeting (IEDM), 2011 IEEE International*, pp. 30.2.1–30.2.4, doi: 10.1109/IEDM.2011.6131642.
15. Sakamoto, T., Sunamura, H., Kawaura, H., Hasegawa, T., Nakayama, T., and Aono, M. (2003) Nanometer-scale switches using copper sulfide. *Appl. Phys. Lett.*, **82** (18), 3032–3034, doi: 10.1063/1.1572964.
16. Miyamura, M., Tada, M., Sakamoto, T., Banno, N., Okamoto, K., Iguchi, N., and Hada, H. (2012) First demonstration of logic mapping on nonvolatile programmable cell using complementary atom switch. *Electron Devices Meeting (IEDM), 2012 IEEE International*, pp. 10.6.1–10.6.4, doi: 10.1109/IEDM.2012.6479020.
17. Tada, M., Okamoto, K., Sakamoto, T., Miyamura, M., Banno, N., and Hada, H. (2011) Polymer solid-electrolyte switch embedded on CMOS for non-volatile crossbar switch. *IEEE Trans. Electron Devices*, **58** (12), 4398–4406, doi: 10.1109/TED.2011.2169070.
18. Sakamoto, T., Tada, M., Okamoto, K., and Hada, H. (2012) Electronic conduction mechanism in atom switch using polymer solid electrolyte. *IEEE Trans. Electron Devices*, **59** (12), 3574–3577, doi: 10.1109/TED.2012.2219051.
19. Nakaya, S., Miyamura, M., Sakimura, N., Nakamura, Y., and Sugabayashi, T. (2013) A non-volatile reconfigurable offloader for wireless sensor nodes. *IPSJ Trans. Syst. LSI Des. Methodol.*, **6**, 52–59, doi: 10.2197/ipsjtsldm.6.52.

25

ReRAM-Based Neuromorphic Computing

Giacomo Indiveri, Eike Linn, and Stefano Ambrogio

25.1

Neuromorphic Systems: Past and Present Approaches

Artificial neural networks are currently being used as a promising means to solve problems in machine learning and computer science [1–3]. These networks are loosely inspired by massively parallel biological neural systems, but they are often implemented on mostly sequential hardware computing platforms, ignoring key features of real neural processing systems, such as their ability to carry out robust and efficient computation using noisy components with limited resolution, which are highly variable and unreliable, and which have extremely low-power consumption characteristics. Neuromorphic systems represent a new and alternative class of neural network hardware architectures that attempt to implement these features by employing hybrid analog–digital electronic circuits that directly emulate the physics of neural processing elements by exploiting the physics of the devices and materials used [4, 5]. Recent developments in nanotechnologies are making available not only extremely compact and low-power but also highly variable and unreliable solid-state devices with memristive properties that can potentially extend the offerings of current neuromorphic systems [6–9]. In particular, memristors are considered as a promising solution for modeling key features of biological synapses due to their nanoscale dimensions, their capacity to store multiple bits of information per element and the low energy required to change their state.

25.2

Neuromorphic Engineering

Although the history of implementing electronic models of neural circuits extends back to the construction of perceptrons [10] and retinas [11],¹⁾ the modern wave of research utilizing very-large-scale integration (VLSI) technology and emphasizing the nonlinear current characteristics of the transistor began in

1) Text adapted from Ref. [5].

the mid-1980s with the collaboration that sprung up between prominent scientists Max Delbrück, John Hopfield, Carver Mead, and Richard Feynman [12]. Inspired by graded synaptic transmission in the retina, Mead sought to use the graded (analog) properties of transistors, rather than simply operating them as on–off (digital) switches. He showed that analog neuromorphic circuits share many common physical properties with proteic channels in neurons [4]. As a consequence, these types of circuits require far fewer transistors than digital approaches to emulating neural systems.

Through the *Physics of Computation* course at Caltech (led by Carver Mead, John Hopfield, and Richard Feynman), Mead's textbook *Analog VLSI and Neural Systems* [4], and the creation of the *Telluride Neuromorphic Engineering Workshop*, the field of Neuromorphic Engineering was established. Prominent in the early expansion of the field were scientists and engineers such as Christof Koch, Terry Sejnowski, Rodney Douglas, Andreas Andreou, Paul Mueller, Jan van der Spiegel, and Eric Vittoz, training a generation of cross-disciplinary students.

It has been argued that neuromorphic circuits are ideal for developing a new generation of computing technologies that use the same organizing principles of the biological nervous system [13–15]. In addition to the computations of a single neuron, many neuromorphic circuits also utilize spiking representations for communication, learning and memory, and computation. The use of asynchronous spike-based or digital event-based representations in electronic systems can be energy efficient and fault tolerant, making them ideal for building modular systems and creating complex hierarchies of computation. The most successful neuromorphic systems to date have been single-chip devices that emulate peripheral sensory transduction such as silicon retinas, visual motion sensors, and silicon cochleas for a wide variety of applications.

In recent years, many larger multichip neuromorphic systems have begun to emerge that have raised new issues and challenges. Neuromorphic engineering now aims to use these technologies for developing larger-scale neural processing systems and move from the predominantly feed-forward, reactive neuromorphic systems of the past to adaptive behaving ones that can be considered cognitive [5, 16].

25.3

Neuromorphic Computing (The Present)

The idea of linking the type of information processing that takes place in the brain with theories of computation and computer science (i.e., *neurocomputing*) dates back to the origins of computer science itself [17, 18]. Neurocomputing has been very popular in the past [10, 19], eventually leading to the development of the artificial neural networks field, with both software simulations and dedicated hardware implementations, applied to a wide variety of practical problems [1, 20–23].

Both artificial neural networks and neuromorphic computing architectures are now receiving renewed attention thanks to progress in information and communication technologies (ICTs) and to the advent of new promising nanotechnologies. Some of present day neurocomputing approaches attempt to model the fine details of neural computation using standard technologies. For example, the *Blue Brain* project, launched in 2005, made use of a 126 kW Blue Gene/P IBM supercomputer to run software that simulated with great biological accuracy the operations of neurons and synapses of a rat neocortical column [24]. Similarly, the *BrainScaleS* EU-FET FP7 project had the goal of developing a custom neural supercomputer by integrating standard CMOS analog and digital VLSI circuits on full silicon wafers to implement about 262 000 I&F neurons and 67 million synapses [25]. Although configurable, the neuron and synapse models were hardwired in the silicon wafers, and the electronic models ran about 10 000 times faster than real biology, with each wafer consuming about 1 kW of power, excluding all external components.

Another large-scale neurocomputing project based on conventional technology is the *SpiNNaker* project [26]. The SpiNNaker is a distributed computer, which interconnects conventional multiple integer precision multi-ARM core chips via a custom communication framework. Each SpiNNaker package contains a chip with 18 ARM9 CPU on it, and a memory chip of 128 Mbyte Synchronous DRAM. Each CPU can simulate different neuron and synapse models. If endowed with simple synapse models, a single SpiNNaker device ARM core can simulate the activity of about 1000 neurons in real time. More complex synapse models (e.g., with learning mechanisms) require more resources and decrease the number of neurons that can be simulated in real time. A complete SpiNNaker board contains 47 of these packages, and the goal is to assemble 1200 of these boards. A full SpiNNaker system of this size would consume about 90 kW.

The implementation of custom large-scale spiking neural network hardware simulation engines is being investigated also by industrial research groups. For example, IBM recently unveiled a multicore spiking neural network chip called “TrueNorth” [27], comprising one million neurons distributed across multiple “neurosynaptic” cores. Each of these cores comprises 256 digital I&F neurons, with 1024×256 binary valued synapses, configured via an SRAM cross-bar array, and uses an asynchronous event-driven design to route spikes from neurons to synapses. The goal is to simulate networks of simplified spiking neurons with human-brain dimensions (i.e., approximately 10^{10} neurons and 10^{14} synapses) in real time. In the mean time, IBM simulated 2.084 billion neurosynaptic cores containing 53×10^{10} neurons and 1.37×10^{14} synapses in software on the Lawrence Livermore National Lab Sequoia supercomputer (96 Blue Gene/Q racks), running 1542× slower than real time [28], and dissipating 7.9 MW.

A diametrically opposite approach is represented by the *Neurogrid* system [29]. This system comprises an array of sixteen $12 \times 14 \text{ mm}^2$ chips, each integrating mixed analog neuromorphic neuron and synapse circuits with digital asynchronous event routing logic. The chips are assembled on a $16.5 \times 19 \text{ cm}^2$ PCB, and the whole system can model over one million neurons connected by billions of synapses in real-time, and using only about 3 W of power [30]. As opposed

to the neurocomputing approaches that are mainly concerned with fast and large simulations of spiking neural networks, the Neurogrid has been designed following the original neuromorphic approach, exploiting the characteristics of CMOS VLSI technology to directly emulate the biophysics and the connectivity of cortical circuits. In particular, the Neurogrid network topology is structured by the data and results obtained from neuroanatomical studies of the mammalian cortex. Although offering less flexibility in terms of connectivity patterns and types of synapse/neuron models that can be implemented, the Neurogrid is much more compact and dissipates orders of magnitude less power than the other neurocomputing approaches described earlier.

25.4

Neuromorphic ReRAM Approaches (The Future)

As discussed in the previous section, most neuromorphic approaches are based on pure CMOS technology. However, the soon projected availability of novel nanodevices [31] offers some unique properties that are highly advantageous for implementing hardware synapses [32]. Especially, redox-based resistive switching random access memory (RAM) devices (ReRAM) [33], also called memristive devices [34, 35], are considered to be very promising candidates for enabling high-density and ultimately scaled synaptic arrays in neuromorphic architectures. ReRAM cells are two terminal devices, which can be arranged in passive crossbar arrays of minimum feature size or potentially integrated in 3D synaptic arrays [6, 36]. The scaling properties of ReRAM cells have been shown to be excellent and could pave the path to very energy-efficient neuromorphic hardware systems (see Section 2.2 for details on the scaling aspects).

In Section 25.4.1, we briefly review the available ReRAM-based neuromorphic approaches. Then, in the following sections, we discuss the key properties of ReRAM useful for realization of hardware synaptic functions in more detail, showing how, depending on the approach followed, one or more of the following properties are exploited:

- nonvolatility/volatility of resistive states
- nonlinear switching kinetics
- multilevel resistance behavior
- capacitive properties
- switching statistics.

25.4.1

ReRAM-Based Neuromorphic Approaches

The basic idea in most ReRAM-based neuromorphic approaches is to consider ReRAM devices, or small ReRAM-based circuits, as artificial synapses. The idea of using ReRAM devices for neuromorphic applications goes back

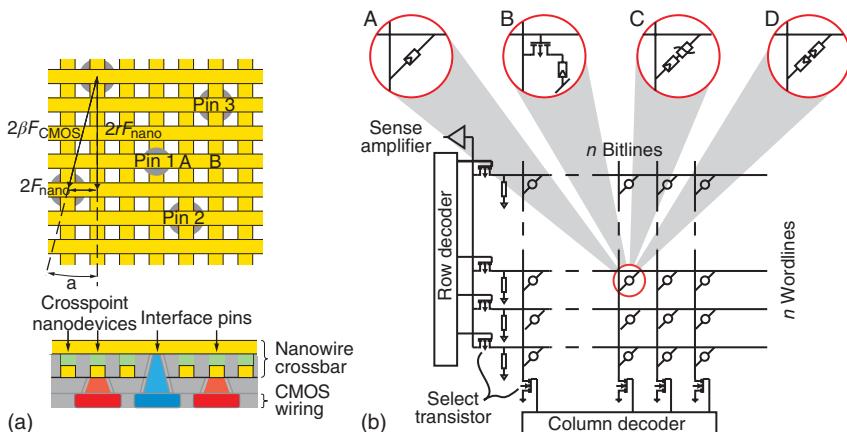


Figure 25.1 (a) “Crossnets” approach from Ref. [40], Chapter 22, Figure 9. (b) Generic ReRAM-based crossbar array indicating various selector options, including (A) a resistive switch offering a nonlinear I - V characteristic (1R), (B) a vertical transistor in series to

a resistive switch (1T1R), (C) a selector in series to a resistive switch (1S1R), and (D) a complementary resistive switch configuration (1CRS). (Reprinted with permission from Ref. [40].)

to Likharev [37], who introduced the concept of “Crossnets,” where ReRAM devices serve as programmable interconnects, that is, binary synapses. On top of CMOS-based neurons, a crossbar array of nanoscaled ReRAM devices is used for reconfigurable wiring of the neurons. An example is given in Ref. [38], and shown in Figure 25.1a, indicating at the top a crossbar array of ReRAM devices providing reconfigurable, binary connection to the CMOS front-end circuit underneath. Binary ReRAM synapses could also be directly used to implement associative memories [39]. Note that the ReRAM crossbar array requires select elements [31] as shown in Figure 25.1b (see Chapters 22 and 24 for more details).

In a more general approach, the tunable resistive state of ReRAM device is used as a synaptic weight, which can be adapted by suitable methods for changing ReRAM resistance in analogy with synaptic plasticity rules. This is schematically shown in Figure 25.2a: by updating the weight of the synapses, the electrical connection between a presynaptic neuron and a postsynaptic neuron changes, thus enabling the possibility of implementing a variety of learning models for pattern recognition and memory storage.

The weight of the synapses is typically updated following the learning rule called STDP, where the conductance change of the synapse depends on the relative timing of the electrical pulses delivered from the presynaptic neuron and the postsynaptic neuron. If the presynaptic spike precedes the postsynaptic one, the synapse conductance is enhanced, whereas if the postsynaptic spike precedes the presynaptic one, the synapse is depressed [24, 41, 43]. Experimental results measured from a rat-hippocampal neuron are shown in Figure 25.2b [41].

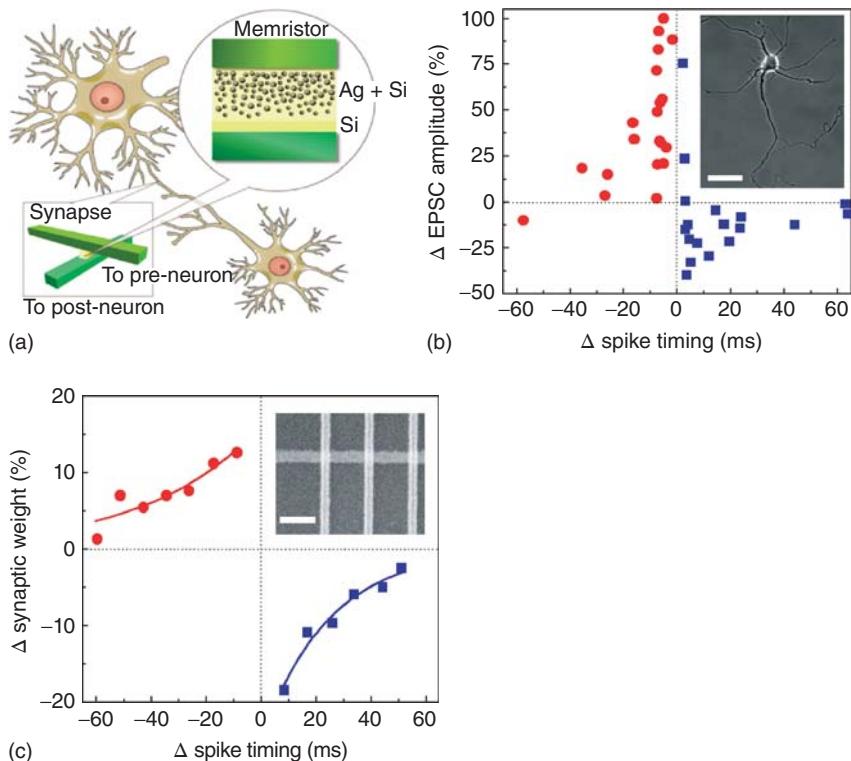


Figure 25.2 (a) Schematic description of the role of the ReRAM (or memristor) as a synapse between two neurons. (b) Experimental change of excitatory postsynaptic current (EPSC) of rat hippocampal neurons

[41] as a function of the relative spike timing. (c) Experimental ReRAM STDP curve. The exponential behavior is similar to biological measurements in (b). (Reprinted with permission from Ref. [42].)

The first experimental evidence of the feasibility of the STDP learning curve with ReRAM devices was reported by Jo *et al.* [42]. Figure 25.2c shows the experimentally measured STDP, namely the conductance update as a function of inter-spike delay. The STDP characteristic shows a conductance increase for a time delay $\Delta t = t_{\text{pre}} - t_{\text{post}} < 0$, that is, for the presynaptic spike preceding the postsynaptic spike. Conversely, when the presynaptic spike precedes the postsynaptic ($\Delta t < 0$), the ReRAM conductance is decreased. The resulting STDP characteristic shows an exponentially decaying behavior, which is consistent with biological data in Figure 25.2b.

The experimental demonstration of memristive STDP in Figure 25.2c was achieved by the approach presented by Snider [44]. Figure 25.3 schematically shows the ReRAM-based synaptic array (a) and the spiking scheme to achieve STDP behavior in the synapse (b). According to this idealized scheme, a single spike cannot alter the resistive state of the synapse, and a weight update requires the presence of a postsynaptic spike and a presynaptic spike overlapping in time.

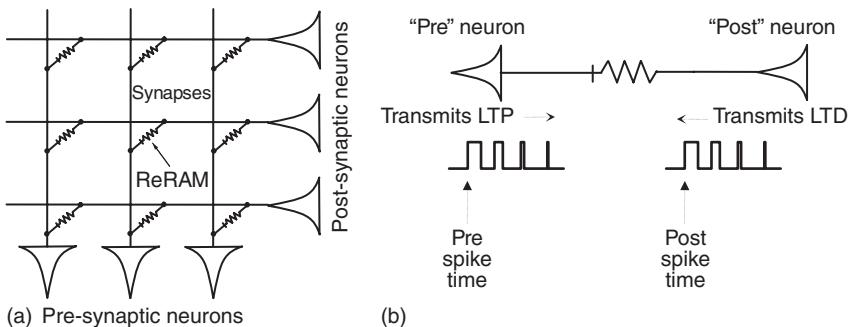


Figure 25.3 (a) ReRAM-based synaptic array and (b) spiking scheme to either induce long-term potentiation (LTP) or long-term depression (LTD) in the synapse. (Reprinted with permission from Ref. [44].)

Note that the pulse width of the spikes gradually decreases with time according to an exponential law. Moreover, the synaptic weight can be either increased or decreased, inducing either long-term potentiation (LTP) or long-term depression (LTD) changes in the synapse.

25.4.2

Nonvolatility and Volatility of Resistive States

The nonvolatility of the resistive states, that is, the hysteretic memristive behavior, is one of the most appealing ReRAM properties to enable synaptic functionality. In general, ReRAM devices offer long-term state retention up to 10 years at typical temperatures below 85 °C [45]. However, this is not true for all ReRAM devices and even the same ReRAM device can display volatile/nonvolatile retention depending on the programming/erasing conditions. In Reference [46], the feasibility of short-term plasticity (STP) and LTP was demonstrated for Ag_2S -based devices. Figure 25.4a shows the STP behavior of a ReRAM device under the application of a sequence of pulses of 0.5 s pulse width, separated by a relatively long interpulse delay $T = 20$ s. Note the volatile change of weights, where the conductance is only transiently increased with no stable potentiation even after nine pulses. In contrast, Figure 25.4b shows that a permanent weight change to high conductance can be achieved when the interpulse delay is reduced to $T = 2$ s. Thus, the volatility of high ohmic ON states enables implementation of STP within a single ReRAM device depending on the spike rate, whereas the nonvolatility of the permanent high-conductance state enables implementation of LTP. Similarly, internal nonequilibrium states that cause an emf (electromotive force) voltage [47] may also influence the resistance states [48]. This means that ReRAM states can offer exponential forgetting, that is, the resistive state is varied exponentially with time, depending on internal nonequilibrium states. In general, internal nonequilibrium states allow for using biorealistic nonoverlapping input spikes that enable

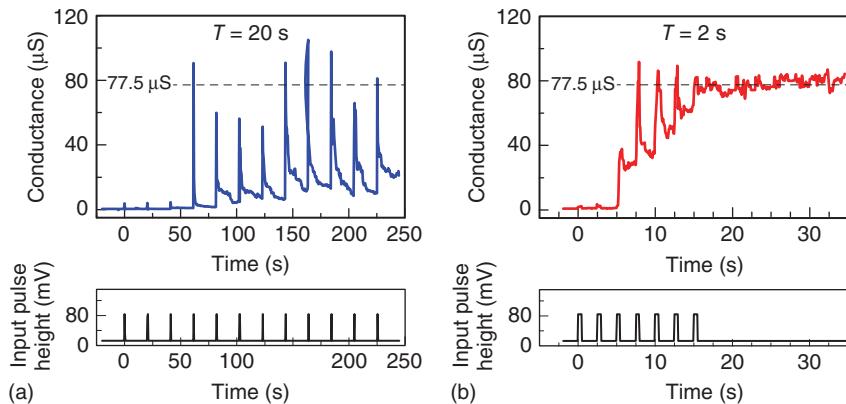


Figure 25.4 (a) Short-term potentiation (STP) and (b) long-term potentiation (LTP) of Ag_2S -based devices is induced for different spike rates. (Reprinted with permission from Ref. [46].)

the emulation of different synaptic plasticity effects, for example, the behavior of the cell's Ca^{2+} concentration as a second-order memristive effect [49, 50].

25.4.3

Nonlinear Switching Kinetics

ReRAM devices in general offer a highly nonlinear switching kinetics [51–54], where the set time t_{set} exponentially depends on the applied pulse height (see Chapter 11). This feature can be used to implement STDP in a simple manner [44, 55]. For a certain pulse length, one can define a threshold voltage below which no switching occurs for the specific pulse duration. If we select a voltage V_{pulse} as the presynaptic spike voltage and $-V_{\text{pulse}}$ as the postsynaptic spike voltage, which are both well below the threshold voltage, the application of either the presynaptic or postsynaptic pulse alone would lead to a negligible change of the synaptic weight. However, if pre- or postsynaptic signals occur simultaneously, the total voltage at the junction will be equal to $2V_{\text{pulse}}$, which shall be well above the threshold voltage. In this case, therefore, the synaptic weight will significantly be affected, thus enabling STDP. Figure 25.5 shows a corresponding STDP implementation according to [55].

To improve the accuracy of the spike shape in terms of biological signals, a similar STDP approach was proposed in Ref. [56]. Various STDP schemes using memristive devices are reviewed in Ref. [57], where the impact of ReRAM device models (see Chapters 13, 14, and 17) is also discussed.

25.4.4

Multilevel Resistance Behavior

Another useful property of ReRAM devices is the capability of multilevel resistances [53, 58–60]. Multiple levels offer the possibility to have multiple discrete

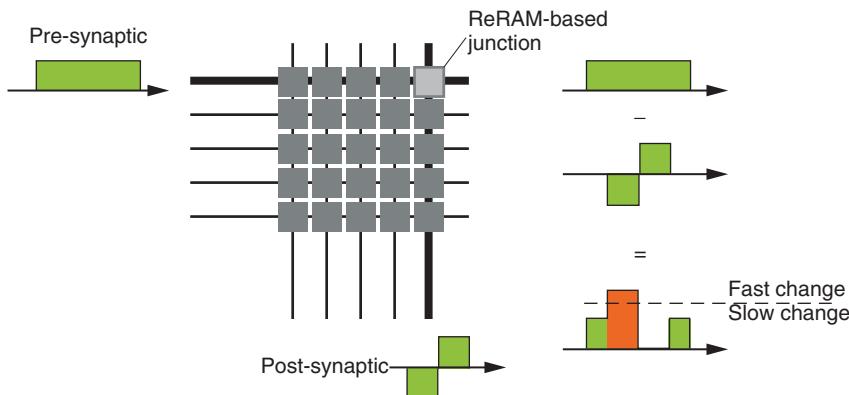


Figure 25.5 Implementation of simple STDP functionality using a ReRAM-based cross-bar array. In this straightforward approach, two types of pre- and postsynaptic signals, both being below the threshold voltage, are considered. Only if the postsynaptic voltage signal occurs while the

presynaptic one is still active, their differential voltage produces a signal large enough to induce a fast change of the resistive state. If the presynaptic signal is applied without a postsynaptic pulse, or vice versa, the resistance change is negligible. (Adapted from Ref. [55].)

weights of the synapses instead of binary weights [9, 32, 46, 60–65]. Gradual tunable ReRAM devices offer the potential of mimicking biological synaptic behavior more realistically than bistable devices (e.g., SRAM-based synapses). However, for typical ReRAM devices, the set process is an abrupt process. This is shown in the $I-V$ curve of an HfO_x ReRAM device in Figure 25.6a, where the set transition at positive voltage shows a vertical change of the current, while the reset transition is more gradual [60]. To reliably tune the resistance to a given analog value, either a current compliance (e.g., $I_c = 500 \mu\text{A}$ in Figure 25.6a) or a series resistor are required. Figure 25.6b shows the control of resistance state during set transition by using increasing compliance currents, from I_{c1} to I_{c3} , resulting in a decreasing value of resistance. The compliance current I_c can be controlled by the voltage applied to the gate of the transistor in a 1T1R structure [60]. Furthermore, the resistance during reset transition can be controlled by the value of the stop voltage as shown in Figure 25.6c.

To take advantage of the analog tuning of resistance in both the set and reset transitions, a 2-transistor/1-resistor (2T1R) synapse capable of both communication and plasticity was proposed [66]. In the 2T1R synapse, the resistor consists of a bipolar HfO_x ReRAM connected to two transistors in parallel, as displayed in Figure 25.7a. The synapse circuit has three access terminals, namely (i) the top electrode (TE), (ii) the communication gate (CG), and (iii) the fire gate (FG). The bottom electrode (BE) serves as output of the current from the synapse to a postsynaptic neuron (POST). The 2T1R synapse is inserted between two neurons, namely a presynaptic neuron (PRE) driving TE and CG with voltage spikes, and POST that receives the current resulting from a PRE spike and driving the

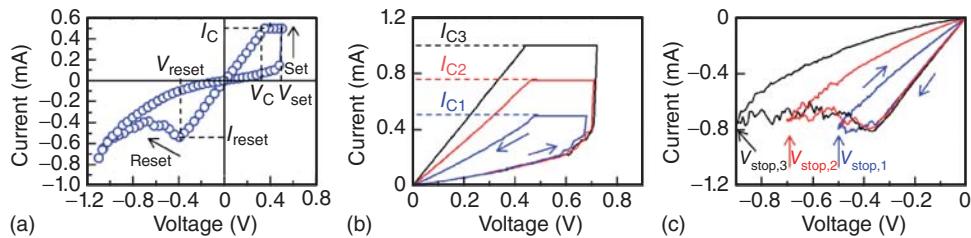


Figure 25.6 (a) Exemplary I - V curve of an HfO_2 ReRAM cell. (b) I - V curves at positive voltage, where resistance after set transition increases at increasing stop voltage ($V_{stop,1} = 0.5$ V, $V_{stop,2} = 0.7$ V and $V_{stop,3} = 0.9$ V). Reprinted with permission from Ref. [60].

voltage, where resistance after reset transition increases at increasing stop voltage ($V_{stop,1} = 0.5$ V, $V_{stop,2} = 0.7$ V and $V_{stop,3} = 0.9$ V). Reprinted with permission from Ref. [60].

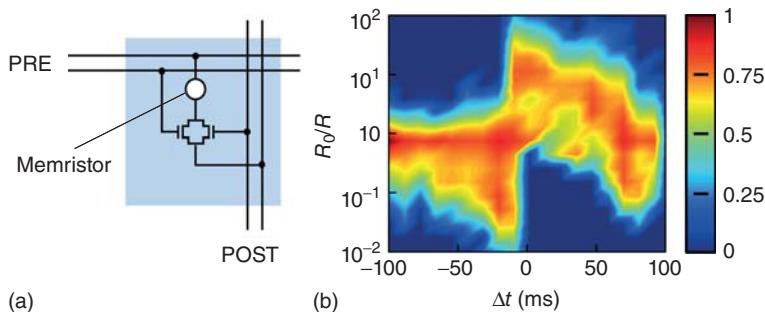


Figure 25.7 Schematic layout of the 2T1R synapse circuit (a) and experimental STDP characteristic (b), showing the probability for a conductance change R_0/R as a function of the delay Δt between PRE/POST synaptic pulses. (Reprinted with permission from Ref. [66].)

FG by a pulse (fire) generated in correspondence of the POST spike. Every neuron behaves as an integrate & fire circuit, capable of integrating all the current received at the input node and transforming the current in an internal (membrane) potential. Once the membrane potential reaches a characteristic threshold, a fire event is generated, where the POST delivers a spike pulse driving the FG terminal of all synapses. If overlapped with the TE pulse delivered by the PRE, this pulse can result in set or reset transitions in the ReRAM element, thus inducing synapse potentiation/depression and enabling learning in the neuromorphic network. The shape of the TE and spike pulses causes a specific STDP response; namely LTP/LTD is achieved when the pulse is following/preceding the PRE spike (delay $\Delta t > 0$ or $\Delta t < 0$, respectively).

Figure 25.7b shows the experimental STDP characteristic for the 2T1R synapse [66]. To characterize this behavior in the synapse, 1T1R samples with HfO_x ReRAM elements were subjected to random spikes with variable delay Δt and random initial state (LRS, HRS, and various intermediate states). The STDP was then reported as the probability to display a conductance change R_0/R for

a given delay Δt (see color scale of Figure 25.7). The statistical representation is needed due to the largely stochastic behavior of the ReRAM device, which can be explained both by the switching variability in the set/reset processes of the HfO_x memristor [67], and the dependence of STDP response to the initial resistance R_0 (i.e., set state, reset state, and intermediate state). Learning of a visual pattern using this scheme has been demonstrated by simulations of a 2-layer neural network (64 neurons in the first layer, 1 neuron in the second layer) linked by 64 2T1R synapses [66]. The results support the ability to learn visual/auditory patterns by gradually changing the synapse weight in a neural network, as opposed to the case of binary abrupt switching [68].

25.4.5

Capacitive Properties

Although ReRAM devices are generally considered as resistive devices, the metal/insulator/metal (MIM) structure of ReRAM also offers a capacitance in parallel [69]. This capacitance can be exploited when considering complementary resistive switch (CRS) structures as binary synapses as described in Figure 25.8a. The resistance of elements A and B can be switched between the high resistance state (HRS) and the low resistance state (LRS), whereas the capacitance is controlled by the device geometry and the permittivity of the insulator material (see Figure 25.8b). For the logic state 0, element A is in the LRS, and thus the capacitance C_A is short-circuited. In this case, the overall capacitance is determined by C_B . Similarly, for state 1, element B is short-circuited and capacitance C_A dominates the overall device behavior. As a result, the stored information of the CRS cell influences the detectable capacitance, which thus can be read out nondestructively. In Figure 25.8b a capacitive voltage divider is formed with C_{out} , and V_{out} is evaluated using a sense amplifier. The controllable capacitance feature of this device can be used to extend the functionality of the binary synapse and may enable pattern recognition tasks [70].

25.4.6

Switching Statistics

ReRAM devices are known to display a switching statistics where set/reset parameters (resistance, voltage, and current) randomly change from cycle to cycle [32, 67]. The stochastic switching can be used to emulate the nondeterministic update of synaptic weight, which has also been observed in biological synapses. From the biological point of view, synapses are characterized by a synaptic conductance or weight w [71, 72], which depends on three different processes, namely

$$w = abc, \quad (25.1)$$

where a represents the number of quantal neurotransmitter release sites, or channels, b is the probability of synaptic release for each site, and c represents a measure of the postsynaptic effect of the synapse [72]. The majority of the STDP models

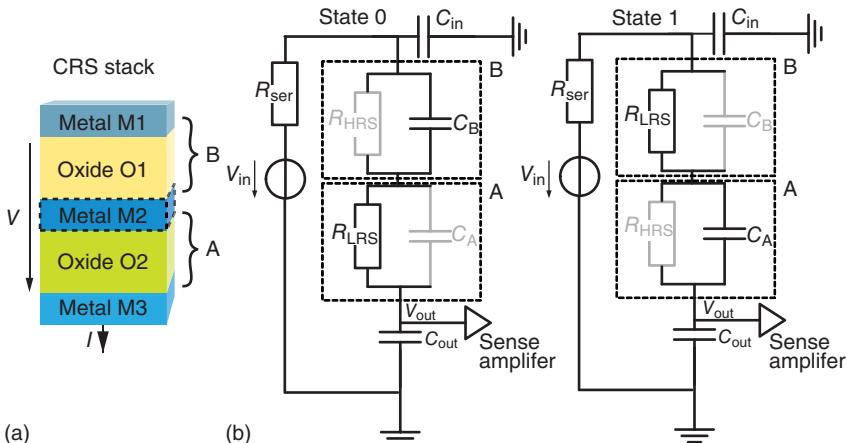


Figure 25.8 (a) Typical CRS stack. (b) State 0 (LRS/HRS) and state 1 (HRS/LRS) offer different capacitances. (Adapted from Ref. [70].)

and demonstrations consider a and b to be constant, leaving all the weight tunability in the parameter c , which gives the form of the STDP learning rule [73]. Analog STDP models thus neglect the natural variability of the number of release sites and probability of synaptic release. Another equivalent approach consists in emulating biological synapses by leaving c constant and varying the transmission probability b . This method fully exploits the variability of nanoscale ReRAM, since deterministic multilevel and stochastic binary synapses are functionally equivalent [71]. This enables synapses, hence ReRAM devices, to assume only two different resistance levels irrespective of their precise values.

Figure 25.9 shows the switching probability for reset (a) and set (b) for a Ag-based ECM cell in 1T1R configuration. By applying relatively weak switching conditions, a certain switching probability, for example, 50%, can be achieved. The switching probability is a function of both the pulse length and the applied voltage, as shown in Figure 25.9. For a given gate voltage ($V_g = 1.5$ V in the figure) and pulse length (500 ns and 1 us), the set/reset probability increases with the applied voltage across the 1T1R. For increasing pulse width, the voltage to achieve a given switching probability decreases due to the time–voltage relationship for set/reset processes [54]. For ReRAM devices that exhibit abrupt set/reset transitions, thus most suitable for binary synapses, stochastic switching allows to gradually tune the synapse weights for learning and recognition, even with no multilevel capabilities [68]. This method reveals that variability arising from scaling could provide some advantage in the design of new neuromorphic architectures.

Another way to practically deal with variability is similar to the previously described method and consists in directly emulating the single biological ion channels, which are affected by stochasticity, in the synaptic cleft through a resistive element. In this way, one resistive device can represent one ion channel, thus employing the intrinsic variability of the nanoelectronic cell to emulate

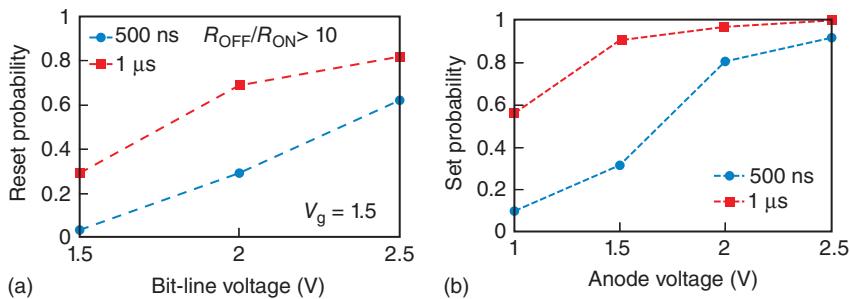


Figure 25.9 Switching probability of a 1T1R Ag-based ECM cell. Two different pulse lengths of 500 ns and 1 μ s are considered, and a gate voltage of $V_g = 1.5$ V is applied.

(a) RESET probability for bitline voltages in the range of 1.5–2.5 V. (b) SET probability for anode voltages of 1.5–2.5 V. (Reprinted with permission from Ref. [8].)

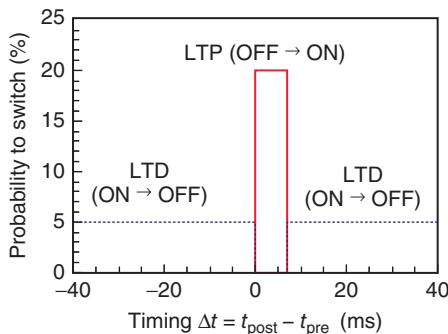


Figure 25.10 Probabilistic STDP learning rule used in Ref. [68]. Δt is defined as the difference between the postsynaptic and the presynaptic spike. (Reprinted with permission from Ref. [68].)

the intrinsic variability of the single-ion channel [32]. Therefore, one synapse is implemented through a number of parallel ReRAMs.

However, scaled ReRAM can allow for a deterministic STDP learning rule by means of a gradual resistance modulation. Yu *et al.*, [62] demonstrated the possibility of employing scaled devices with elevated variability by programming them with very short voltage pulses with a time width of 10 ns. The application of many short pulses allows to overcome the variability of the device since, averaging on many pulses, the device resistance follows a deterministic behavior. Figure 25.11 shows three different resistance states, which are gradually reset through the application of hundreds of pulses. It is thus possible to gradually vary and control the resistance change, also allowing for sub-pJ switching programming energy. With the described method, it has been simulated a robustness of an entire video pattern recognition system at typical experimental variability levels with resistance fluctuations up to 9%. Only a slight decrease of the selectivity properties has been observed with an additional increase of the variability of resistance values.

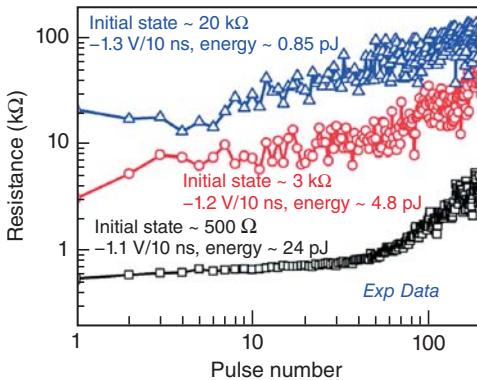


Figure 25.11 Three different resistance states gradually programmed through the application of 10-ns-width voltage pulses. (Reprinted with permission from Ref. [62].)

25.5

Scaling in Neuromorphic ReRAM Architectures

Nowadays, Moore's law is reaching its ultimate scaling limits, due to intrinsic variability and excessive off-state leakage issues in CMOS technology. To extend Moore's law in future generations, the semiconductor industry is considering new technologies, such as ReRAM devices, allowing for denser structures and more compact memory and logic architectures [74]. The same need for scaling is even more important for neuromorphic applications. Indeed, human brains have approximately 10^{11} neurons and 10^{14} synapses. Implementing brain-like processing architectures that comprise even a small fraction of these elements, without employing huge and expensive silicon areas, requires an aggressive scaling of devices and optimization of their organization in dense and low-power passive architectures. In this frame, ReRAM technology is a promising candidate for neuromorphic computing, thanks to its low power consumption (<10 fJ/bit) [75], fast switching (<10 ns) [76], multibit storage [59], extremely small size (<10 nm) [77], and capability of integration in a $4F^2$ area, where F is the minimum feature size. However, despite these promising aspects, several serious issues arise when scaling down the device area, which can be partly overcome with intelligent and alternative neuromorphic approaches.

One of the key issues for scaling down ReRAM devices is the variability of the switching parameters [67, 78]. Recent works have pointed out that switching variations are due to discrete injection of ionized defects [79], which can be modeled by a Poisson distribution [67, 78]. The scaling down of the current and of the CF size also gives rise to undesired random telegraph noise (RTN), which affects the device resistance during read operation [80]. Reducing the filament dimensions also leads to structural instabilities that, through chemical dissolution and electromigration, cause unwanted resistance increase from the set state [81, 82].

As the interconnect lines within the crossbar array are downscaled, the series resistance due to the metal lines might raise additional issues. For instance, the resistivity of Cu, which is generally used for low-resistance metal interconnect in memory circuits, is going to increase over $10 \mu\Omega\text{cm}$ [31] as the cross section is reduced to roughly $10 \times 10 \text{ nm}^2$. As a result, the effective voltage drop across the ReRAM device in the array might be significantly smaller than the applied voltage, due to the voltage drop across the interconnections. To solve this issue, the resistance of the device must be significantly higher than the metal lines, which can be achieved by small filaments obtained at small operation current during set [67]. This, in turn, leads to size-dependent reliability issues such as statistical fluctuations and RTN, as previously discussed.

The problems of device scaling can be mitigated when considering neuromorphic implementations: neuromorphic networks have the advantage of being able to deal with unreliable elements as real biological synapses, thanks to a highly parallel computation. For this reason, scaling drawbacks of ReRAM such as switching variability and noise do not represent critical issues for neuromorphic circuits.

25.6

Applications of Neuromorphic ReRAM Architectures

Potential applications that can best exploit the properties of neuromorphic ReRAM architectures are those that involve the processing of real-world signals and that require compact and low-power devices. Typical examples can be found in the robotic domain, in intelligent brain–machine interfaces, in sensor networks, embedded systems, and portable devices. As neuromorphic computing is mainly concerned with the implementation of the computational principles used by the nervous system, most existing neuromorphic systems adopt *learning* mechanisms in neural networks to solve computational problems [32, 56, 57, 71, 83–85]. In these systems, synapses represent the essential component for both learning and signal processing. Synaptic weight values can be stored in ReRAM devices and updated following the prescription of different types of learning algorithms that can depend on the pre- and postsynaptic neuron activity [61, 86, 87]. In general, within this context, much research has been dedicated to the use of resistive synaptic devices [32, 88] for implementing learning and neural computation, with classification and pattern recognition being the most explored applications [62, 63, 68, 89]. A typical example of a pattern recognition application is in the domain of auditory signals: for example, in Ref. [68] the authors first analyze auditory signals by an array of band-pass filters; then they connect each band-pass filter output to a corresponding input neuron, which is part of the first neural layer (see Figure 25.12). All neurons in this input layer are then connected to a single-output neuron, via CBRAM synapses (three per connection). The output neuron is then trained, following a classical perceptron learning scheme [10]. As training proceeds, the audio pattern is memorized, selectivity becomes very high, and the number of errors decreases. It is possible

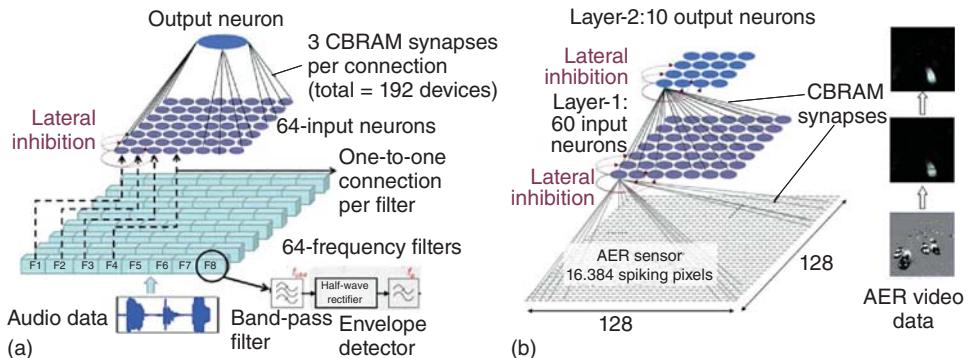


Figure 25.12 Schematic structures of an audio (a) and video (b) neural network pattern recognition architectures. (Reprinted with permission from Ref. [68].)

to imagine multiple applications in the auditory scene analysis domain, using this scheme.

Another common example of pattern recognition comes from image processing of natural scenes. In the example shown in Figure 25.12b from Suri *et al.*, [68], an address-event representation (AER) silicon retina [90] provides transient visual sensory signals to a neural network layer, via synapses implemented by resistive devices. A second neural layer, again connected through CBRAM synapses to the first layer, constitutes the output layer. A visualization of the visual information present at each level of the network is shown on the right, starting from the sensor's output up to the top layer output, in which only the trained pattern is highlighted (which was set to be the smaller pattern on the right, representing a car on the right lane of a freeway). Similar visual pattern recognition applications have been proposed by Zamarreño-Ramos *et al.* [56], Thorpe [89], Yu *et al.* [62], Park *et al.* [63].

As these examples demonstrate, the neuromorphic approach makes use of massively parallel architectures, that typically operate using hybrid analog and digital components, often with inhomogeneous or even unreliable outputs. This practically means that developing computing architectures that can operate using unreliable devices is potentially not a problem, as some studies have already pointed out [55]. Neuromorphic architectures can also exploit the probabilistic nature of the components they have and the signals they process, without using hard logic [32, 91]. Therefore, they are an ideal medium for exploiting the properties of memristive devices and applying them to the solution of real-world problems.

In conclusion, the neuromorphic approach with memristive devices is promising for a new class of applications and revolutionary logic perspectives. For the vast majority, these types of applications have only been simulated up to now. Many efforts are still needed to improve the memristive switching performances and integrate them onto standard CMOS processes, but real memristive neuromorphic future applications are not inconceivable.

References

1. Bishop, C. (2006) *Pattern Recognition and Machine Learning*, Springer-Verlag, New York.
2. Bengio, Y. (2009) Learning deep architectures for AI. *Found. Trends® Mach. Learn.*, **2** (1), 1–127, doi: 10.1561/2200000006.
3. Lee, H., Grosse, R., Ranganath, R., and Ng, A.Y. (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, New York, ICML '09, pp. 609–616, doi: 10.1145/1553374.1553453.
4. Mead, C. (1989) *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA.
5. Indiveri, G. and Horiuchi, T. (2011) Frontiers in neuromorphic engineering. *Front. Neurosci.*, **5** (118), 1–2, doi: 10.3389/fnins.2011.00118.
6. Gao, B., Bi, Y., Chen, H.Y., Liu, R., Huang, P., Chen, B., Liu, L., Liu, X., Yu, S., Wong, H.S.P., and Kang, J. (2014) Ultra-low-energy three-dimensional oxide-based electronic synapses for implementation of robust high-accuracy neuromorphic computation systems. *ACS Nano*, **8** (7), 6998–7004.
7. Jo, S.H., Chang, T., Ebong, I., Bhadviya, B.B., Mazumder, P., and Lu, W. (2010) Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.*, **10** (4), 1297–1301.
8. Suri, M., Bichler, O., Querlioz, D., and Palma, G. (2012) CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: auditory (Cochlea) and visual (Retina) cognitive processing applications. *International Electron Devices Meeting (IEDM)*, pp. 10.3.1–10.3.4.
9. Jackson, B.L., Rajendran, B., Corrado, G.S., Breitwisch, M., Burr, G.W., Cheek, R., Gopalakrishnan, K., Raoux, S., Rettner, C., Padilla, A., Schrott, A.G., Shenoy, R.S., Kurdi, B.N., Lam, C.H., and Modha, D.S. (2013) Nanoscale electronic synapses using phase change devices. *ACM J. Emerg. Technol. Comput. Syst.*, **9** (2), 1–20.
10. Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, **65** (6), 386–408, doi: 10.1037/h0042519.
11. Fukushima, K., Yamaguchi, Y., Yasuda, M., and Nagata, S. (1970) An electronic model of the retina. *Proc. IEEE*, **58** (12), 1950–1951, doi: 10.1109/PROC.1970.8066.
12. Hey, T. (1999) Richard Feynman and computation. *Contemp. Phys.*, **40** (4), 257–265, doi: 10.1080/001075199181459.
13. Boahen, K. (2005) Neuromorphic microchips. *Sci. Am.*, **292** (5), 56–63.
14. Douglas, R., Mahowald, M., and Mead, C. (1995) Neuromorphic analogue VLSI. *Annu. Rev. Neurosci.*, **18**, 255–281.
15. Sarpeshkar, R. (2006) Brain power—borrowing from biology makes for low power computing—bionic ear. *IEEE Spectr.*, **43** (5), 24–29.
16. Neftci, E., Binas, J., Rutishauser, U., Chicca, E., Indiveri, G., and Douglas, R. (2013) Synthesizing cognition in neuromorphic electronic systems. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (37), E3468–E3476, doi: 10.1073/pnas.1212083110.
17. McCulloch, W. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, **5**, 115–133.
18. von Neumann, J. (1958) *The Computer and the Brain*, Yale University Press, New Haven, CT.
19. Minsky, M. (1967) *Computation: Finite and Infinite Machines*, Prentice-Hall, Upper Saddle River, NJ.
20. Hopfield, J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, **79** (8), 2554–2558.
21. Rumelhart, D. and McClelland, J. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, vol. 1, MIT Press, Cambridge, MA.
22. Kohonen, T. (1988) *Self-Organization and Associative Memory*, Springer Series in Information Sciences, 2nd edn, Springer-Verlag.

23. Hertz, J., Krogh, A., and Palmer, R. (1991) *Introduction to the Theory of Neural Computation*, Addison-Wesley, Reading, MA.
24. Markram, H. (2006) The blue brain project, in *ACM/IEEE Conference on Supercomputing, SC 2006*, IEEE, ACM, New York, pp. 53–, doi: 10.1145/1188455.1188511.
25. Schemmel, J., Fieres, J., and Meier, K. (2008) Wafer-scale integration of analog neural networks. *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 431–438.
26. Jin, X., Lujan, M., Plana, L., Davies, S., Temple, S., and Furber, S. (2010) Modeling spiking neural networks on SpiNNaker. *Comput. Sci. Eng.*, **12** (5), 91–97.
27. Merolla, P., Arthur, J., Alvarez-Icaza, R., Cassidy, A., Sawada, J., Akopyan, F., Jackson, B., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S., Appuswamy, R., Taba, B., Amir, A., Flickner, M., Risk, W., Manohar, R., and Modha, D. (2014) A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, **345** (6197), 668–673, doi: 10.1126/science.1254642.
28. Wong, T., Preissl, R., Datta, P., Flickner, M., Singh, R., Esser, S., McQuinn, E., Appuswamy, R., Risk, W., Simon, H., and Modha, D. (2012) 10^{14} , IBM Research Report RJ10502 (ALM1211-004), IBM Research, San Jose, CA.
29. Benjamin, B.V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A.R., Bussat, J., Alvarez-Icaza, R., Arthur, J., Merolla, P., and Boahen, K. (2014) Neurorgrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proc. IEEE*, **102** (5), 699–716.
30. Choudhary, S., Sloan, S., Fok, S., Neckar, A., Trautmann, E., Gao, P., Stewart, T., Eliasmith, C., and Boahen, K. (2012) Silicon neurons that compute, in *Artificial Neural Networks and Machine Learning—ICANN 2012*, Lecture Notes in Computer Science, vol. **7552** (eds A. Villa, W. Duch, P. Erdi, F. Masulli, and G. Palm), Springer-Verlag Berlin / Heidelberg, pp. 121–128.
31. ITRS (2012) The international technology roadmap for semiconductors - ITRS, 2012 edn.
32. Indiveri, G., Linares-Barranco, B., Legenstein, R., Deligeorgis, G., and Prodromakis, T. (2013) Integration of nanoscale memristor synapses in neuromorphic computing architectures. *Nanotechnology*, **24**, 384010.
33. Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009) Redox-based resistive switching memories - nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, **21** (25-26), 2632–2663.
34. Chua, L. and Kang, S. (1976) Memristive devices and systems. *Proc. IEEE*, **64** (2), 209–223.
35. Strukov, D.B., Snider, G.S., Stewart, D.R., and Williams, R.S. (2008) The missing memristor found. *Nature*, **453** (7191), 80–83.
36. Prezioso, M., Merrikh-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K., and Strukov, D.B. (2015) Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nat. Lett.*, **521**, 61–64.
37. Turel, O. and Likharev, K. (2003) Cross-nets: possible neuromorphic networks based on nanoscale components. *Int. J. Circuit Theory Appl.*, **31** (1), 37–53.
38. Strukov, D. and Likharev, K. (2005) CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices. *Nanotechnology*, **16** (6), 888–900.
39. Alibart, F., Sherwood, T., and Strukov, D. (2011) Hybrid CMOS/nanodevice circuits for high throughput pattern matching applications. *2011 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pp. 279–286.
40. Waser, R (ed.) (2012) *Nanoelectronics and Information Technology*, 3rd edn, Wiley-VCH Verlag GmbH.
41. Bi, G.Q. and Poo, M.M. (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, **18** (24), 10 464–10 472.
42. Jo, S.H., Chang, T., Ebong, I., Bhadviya, B.B., Mazumder, P., and Lu, W. (2010) Nanoscale memristor device as synapse

- in neuromorphic systems. *Nano Lett.*, **10**, 1297–1301.
43. Abbott, L.F. and Nelson, S.B. (2000) Synaptic plasticity: taming the beast. *Nat. Neurosci.*, **3**, 1178–1183.
 44. Snider, G. (2008) Spike-timing-dependent learning in memristive nanodevices. *IEEE International Symposium on Nanoscale Architectures*, pp. 85–92.
 45. Wei, Z., Takagi, T., Kanzawa, Y., Katoh, Y., Ninomiya, T., Kawai, K., Muraoka, S., Mitani, S., Katayama, K., Fujii, S., Miyanaga, R., Kawashima, Y., Mikawa, T., Shimakawa, K., and Aono, K. (2012) Retention model for high-density ReRAM. *2012 4th IEEE International Memory Workshop, IMW 2012*.
 46. Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J.K., and Aono, M. (2011) Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.*, **10**, 591–595.
 47. Valov, I., Linn, E., Tappertzhofen, S., Schmelzer, S., van den Hurk, J., Lentz, F., and Waser, R. (2013) Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nat. Commun.*, **4**, 1771.
 48. Tappertzhofen, S., Linn, E., Böttger, U., Waser, R., and Valov, I. (2014) Nanobattery effect in RRAMs - implications on device stability and endurance. *IEEE Electron Device Lett.*, **35** (2), 208–210.
 49. Kim, S., Du, C., Sheridan, P., Ma, W., Choi, S., and Lu, W.D. (2015) Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.*, **15**, 2203–2211.
 50. Du, C., Ma, W., Chang, T., Sheridan, P., and Lu, W.D. (2015) Biorealistic implementation of synaptic functions with oxide memristors through internal ionic dynamics. *Adv. Funct. Mater.*, **25** (27), 4290–4299.
 51. Menzel, S., Tappertzhofen, S., Waser, R., and Valov, I. (2013) Switching kinetics of electrochemical metallization memory cells. *Phys. Chem. Chem. Phys.*, **15** (18), 6945–6952.
 52. Menzel, S., Waters, M., Marchewka, A., Böttger, U., Dittmann, R., and Waser, R. (2011) Origin of the ultra-nonlinear switching kinetics in oxide-based resistive switches. *Adv. Funct. Mater.*, **21** (23), 4487–4492.
 53. Russo, U., Kamalanathan, D., Ielmini, D., Lacaita, A.L., and Kozicki, M.N. (2009) Study of multilevel programming in programmable metallization cell (PMC) memory. *IEEE Trans. Electron Devices*, **56** (5), 1040–1047.
 54. Ielmini, D. (2011) Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Devices*, **58** (12), 4309–4317.
 55. Querlioz, D., Bichler, O., and Gamrat, C. (2011) Simulation of a memristor-based spiking neural network immune to device variations. *International Joint Conference on Neural Networks (IJCNN)*, pp. 1775–1781.
 56. Ramos, C.Z., Camuñas-Mesa, L.A., Pérez-Carrasco, J.A., Masquelier, T., Serrano-Gotarredona, T., and Linares-Barranco, B. (2011) On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front. Neurosci.*, **5** (26), 1–22.
 57. Serrano-Gotarredona, T., Masquelier, T., Prodromakis, T., Indiveri, G., and Linares-Barranco, B. (2013) STDP and STDP variations with memristors for spiking neuromorphic learning systems. *Front. Neurosci.*, **7** (2), 1–15.
 58. Menzel, S., Böttger, U., and Waser, R. (2012) Simulation of multilevel switching in electrochemical metallization memory cells. *J. Appl. Phys.*, **111** (1), 014 501/1–5.
 59. Balatti, S., Larentis, S., Gilmer, D.C., and Ielmini, D. (2013) Multiple memory states in resistive switching devices through controlled size and orientation of the conductive filament. *Adv. Mater.*, **25** (10), 1474–1478.
 60. Ambrogio, S., Balatti, S., Nardi, F., Facchinetto, S., and Ielmini, D. (2013) Spike-timing dependent plasticity in a transistor-selected resistive switching memory. *Nanotechnology*, **24**, 384 012.
 61. Brader, J., Senn, W., and Fusi, S. (2007) Learning real world stimuli in a neural network with spike-driven

- synaptic dynamics. *Neural Comput.*, **19**, 2881–2912.
62. Yu, S., Gao, B., Fang, Z., Yu, H., Kang, J., and Wong, H.S.P. (2012) A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: experimental characterization and large-scale modeling. *International Electron Devices Meeting IEDM*, pp. 239–242.
63. Park, S., Kim, H., Choo, M., Noh, J., Sheri, A., Jung, S., Seo, K., Park, J., Kim, S., Lee, W., Shin, J., Lee, D., Choi, G., Woo, J., Cha, E., Jang, J., Park, C., Jeon, M., Lee, B., and Hwang, H. (2012) RRAM-based synapse for neuromorphic system with pattern recognition function. *International Electron Devices Meeting IEDM*, pp. 231–234.
64. Wang, Z.Q., Xu, H.Y., Li, X.H., Yu, H., Liu, Y.C., and Zhu, X.J. (2012) Synaptic learning and memory functions achieved using oxygen ion migration/diffusion in an amorphous InGaZnO memristor. *Adv. Funct. Mater.*, **22**, 2759–2765.
65. Heittmann, A. and Noll, T.G. (2012) A hybrid CMOS/memristive nanoelectronic circuit for programming synaptic weights. *European Symposium on Artificial Neural Networks (ESANN)*, pp. 639–644.
66. Wang, Z., Ambrogio, S., Balatti, S., and Ielmini, D. (2015) A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning in neuromorphic systems. *Front. Neurosci.*, **8**, 438.
67. Ambrogio, S., Balatti, S., Cubeta, A., Calderoni, A., Ramaswamy, N., and Ielmini, D. (2014) Statistical fluctuations in HfO_x resistive-switching memory (RRAM): Part I - Set/Reset variability. *IEEE Trans. Electron Devices*, **61** (8), 2912–2919.
68. Suri, M., Querlioz, D., Bichler, O., Palma, G., Vianello, E., Vuillaume, D., Gamrat, C., and DeSalvo, B. (2013) Bio-inspired stochastic computing using binary CBRAM synapses. *IEEE Trans. Electron Devices*, **60** (7), 2402–2409.
69. Tappertzhofen, S., Linn, E., Nielsen, L., Rosezin, R., Lentz, F., Bruchhaus, R., Valov, I., Böttger, U., and Waser, R. (2011) Capacity based nondestructive readout for complementary resistive switches. *Nanotechnology*, **22** (39), 395203/1-7.
70. Kavehei, O., Linn, E., Nielsen, L., Tappertzhofen, S., Skafidas, S., Valov, I., and Waser, R. (2013) Associative capacitive network based on nanoscale complementary resistive switches for memory-intensive computing. *Nanoscale*, **5** (11), 5119–5128.
71. Goldberg, D.H., Cauwenberghs, G., and Andreou, A.G. (2001) Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons. *Neural Netw.*, **14**, 781–793.
72. Koch, C. (1999) *Biophysics of Computation: Information Processing in Single Neurons*, Oxford University Press, New York.
73. Masquelier, T., Guyonneau, R., and Thorpe, S.J. (2008) Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS Comput. Biol.*, **1**, 1–9.
74. Waser, R. and Aono, M. (2007) Nanoionics-based resistive switching memories. *Nat. Mater.*, **6**, 833–840.
75. Cheng, C.H., Tsai, C.Y., Chin, A., and Yeh, F.S. (2010) High performance ultra-low energy RRAM with good retention and endurance. *International Electron Devices Meeting IEDM*, pp. 448–451.
76. Lee, H.Y., Chen, P.S., Wu, T.Y., Chen, Y.S., Wang, C.C., Tzeng, P.J., Lin, C.H., Chen, F., Lien, C.H., and Tsai, M.J. (2008) Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM. *International Electron Devices Meeting IEDM*, pp. 1–4.
77. Govoreanu, B., Kar, G.S., Chen, Y.Y., Paraschiv, V., Kubicek, S., Fantini, A., Radu, I.P., Goux, L., Clima, S., Degraeve, R., Jossart, N., Richard, O., Vandeweyer, T., Seo, K., Hendrickx, P., Pourtois, G., Bender, H., Altimime, L., Wouters, D.J., Kittl, J.A., and Jurczak, M. (2011) 10x10 nm² Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. *International Electron Devices Meeting IEDM*, pp. 31–34.
78. Fantini, A., Goux, L., Degraeve, R., Wouters, D.J., Raghavan, N., Kar, G.,

- Belmonte, A., Chen, Y.Y., Govoreanu, B., and Jurczak, M. (2013) Intrinsic switching variability in HfO₂ RRAM. *International Memory Workshop IMW*, pp. 30–33.
79. Yang, Y., Gao, P., Gaba, S., Chang, T., Pan, X., and Lu, W. (2012) Observation of conducting filament growth in nanoscale resistive memories. *Nat. Commun.*, **3** (732), 1–8.
80. Ambrogio, S., Balatti, S., Cubeta, A., Calderoni, A., Ramaswamy, N., and Ielmini, D. (2014) Statistical fluctuations in HfO_x resistive-switching memory (RRAM): Part II - Random Telegraph Noise. *IEEE Trans. Electron Devices*, **61** (8), 2920–2927.
81. Chen, Y.Y., Degraeve, R., Govoreanu, B., Clima, S., Goux, L., Fantini, A., Kar, G.S., Wouters, D.J., Groeseneken, G., and Jurczak, M. (2013) Postcycling LRS retention analysis in HfO₂/Hf RRAM 1T1R device. *IEEE Electron Device Lett.*, **34** (5), 626–628.
82. Ninomiya, T., Muraoka, S., Wei, Z., Yasuhara, R., Katayama, K., and Takagi, T. (2013) Improvement of data retention during long-term use by suppressing conductive filament expansion in TaO_x bipolar-ReRAM. *IEEE Electron Device Lett.*, **34** (6), 762–764.
83. Yang, J.J., Strukov, D.B., and Stewart, D.R. (2013) Memristive devices for computing. *Nat. Nanotechnol.*, **8**, 13–24.
84. Qiao, N., Mostafa, H., Corradi, F., Osswald, M., Stefanini, F., Sumislawska, D., and Indiveri, G. (2015) A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. *Front. Neurosci.*, **9**, 141.
85. O'Connor, P., Neil, D., Liu, S.C., Delbrück, T., and Pfeiffer, M. (2013) Real-time classification and sensor fusion with a spiking deep belief network. *Front. Neurosci.*, **7**, 178.
86. Abbott, L. and Gerstner, W. (2004) Homeostasis and learning through spike-timing dependent plasticity.
87. Graupner, M. and Brunel, N. (2012) Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 3991–3996, doi: 10.1073/pnas.1109359109.
88. Pickett, M.D., Medeiros-Ribeiro, G., and Williams, R.S. (2013) A scalable neuristor built with Mott memristors. *Nat. Mater.*, **12**, 114–117.
89. Masquelier, T. and Thorpe, S.J. (2007) Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.*, **3** (2), 247–257.
90. Lichtsteiner, P., Posch, C., and Delbrück, T. (2008) An 128x128 120dB 15μs-latency temporal contrast vision sensor. *IEEE J. Solid-State Circuits*, **43** (2), 566–576.
91. Griffiths, T.L. and Tenenbaum, J.B. (2006) Optimal predictions in everyday cognition. *Psychol. Sci.*, **17** (9), 767–773.

Index

Page numbers in *italics* refer to figures and those in bold refer to tables

ab initio simulation 398
 – ECM 422–423
 – VCM
 – – electron transfer 404, 404–405
 – – LRS and HRS state 402, 402–404, 403
 – – migration barriers 406–407
 – – phase transformations and nucleation 405–406, 406
 adiabatic constriction, in 2D 199, 199
 admittance/impedance spectroscopy 341–342
 AF, *see* atomic filament (AF)
 AHR model, *see* anode hydrogen release (AHR) model
 aluminum oxide
 – defects 116–117
 – – acceptor-doped alumina 117–118
 – – donor-doped alumina 118–119
 – electronic structure 76
 – ion transport 153, 153–155
 ambipolar diffusion 136
 amorphous states, electronic structure 75, 75–76
 Anderson insulators 72
 Anderson-Stuart model 486–487, 487
 anion-based memory 289
 anode hole injection (AHI) model 231
 anode hydrogen release (AHR) model, 231, 232, 233–235, 234
 anti-Frenkel defects 99–100
 anti-Schottky defects 100
 AOs, *see* atomic orbitals (AOs)
 application-specific integrated circuits (ASICs) 695
 area-proportional switching 5

area-scaling-type switching, *see* homogeneous interface-type switching
 Arrhenius law 133, 319
 Arrhenius-type behavior 128
 ASICs, *see* application-specific integrated circuits (ASICs)
 atomic filament (AF)
 – barrier electrical conductance 553–554
 – classical resistance 551–552
 – minimal atomic bridge estimation parameters 552
 – nonrectangular barrier 555, 555
 – quantum resistance 552–553, 553
 – transmission through atomic gaps 555–556, 556
 atomic migration mechanism 131, 131
 atomic orbitals (AOs) 49
 atomic-scale electrical switching 516–517, 517
 atomic-size contacts, superconducting subgap structure for Al 203, 203
 atomic switches
 – development 516–517
 – gapless-type switch, *see* gapless-type atomic switches
 – gap-type switch, *see* gap-type atomic switches
 – nanofountain pen with Ag ink 516, 516
 – three-terminal switches, *see* three-terminal atomic switches
 – types and functionalities 518, 541, 541–542
 – working principle 517–519, 518
 atomistic theory, nucleation 267–269, 268, 269

- 500 atom model 501, 501–502
- atom transistors 539

- ballistic conductor 199
- band insulators 71
 - SnO₂: 3d¹⁰ system 175, 175–176, 176
 - TiO₂: 3d⁰ system 176–177, 177
- band structure model 60–62, 61
- band transport, of carriers 168–169
- barrier potential 569–571
 - image charge model 569–570, 570
 - Schottky lowering 571
- bathtub-type connection 592, 592
- Becker-Döring approach 266
- binary Me–O systems 256
- binary oxides
 - CF formation 210
 - Ellingham diagram for 90, 90–91
 - filamentary conduction and quantization effects in 210–218
 - rock-salt structure 50–51, 51
 - RS effect in 216
- bipolar resistive switching (BRS) 343, 343
 - effect 12
 - mode 2, 3
 - bipolar switching behavior
 - electroforming conditions 291–294, 292, 293
 - in VCM case 297
 - bipolar switching TMO RRAM
 - disturb
 - – anomalous behavior 616–617, 617
 - – phenomena 615, 616
 - – understanding and modeling 616, 616
 - endurance
 - – balancing 599, 599–600
 - – degradation 600–601, 601, 601
 - – high-endurance devices 598, 598, 598
 - – memory hierarchy position 598, 599
 - retention
 - – degradation in 603–604
 - – LRS retention 602, 602
 - – oxygen vacancy filamentary 601, 602
 - – trade-off between 604–605
 - RTN
 - – CBRAM 609–610
 - – charge carrier transport-induced 610–611, 611
 - – detection, of signals 609
 - – key mechanisms of 615
 - – non-steady-state fluctuations 609
 - – oxide-based RRAM (OXRAM) 609
 - oxygen vacancy transport-induced 611–615
 - steady-state fluctuations 609
 - voltage conditions 609
 - variability
 - – in bipolar switching devices 607–609, 608
 - – of device characteristics 605
 - – of forming operation 605–606
 - – intrinsic and extrinsic 606
 - – resistive memory 605
 - – in unipolar devices 607
 - bistable resistance switching 9, 10, 10
 - Bloch functions 60
 - block-copolymer self-assembly for, advanced ReRAM 687–688
 - blocking contact 304
 - Blue Brain* project 717
 - Boltzmann distribution function 260
 - Born–Oppenheimer approximation 168, 169
 - Born–von Karman boundary conditions 59, 60
 - bottom-up fabrication methods 661
 - resistive switching devices, framework of 662
 - template-assisted, of NW 663–664
 - VLS 662–663
 - boundary layers
 - enrichment and weak depletion layers 260–261
 - formation of new phases 265–266
 - nanosize effects
 - – due to surface curvature 265
 - – on space-charge regions 263–265, 264
 - space-charge layers 258–260, 262, 264
 - strong depletion layers 261–263
 - BrainScaleS* EU-FET FP7 project 717
 - breakdown, *see also* specific types of breakdown
 - implications of, statistics for ReRAM 237–241, 238, 239, 241
 - and oxide failure statistics 235–237, 236
 - path and inference on filament formation 241–246, 242, 243
 - Brillouin zone (BZ) 61
 - Brouwer diagram, point defects 101–102, 102
 - BRS, *see* bipolar resistive switching (BRS)
 - Butler–Volmer equation 280, 281, 282, 330
 - Butler–Volmer relationship 493–494
 - BZ, *see* Brillouin zone (BZ)

cation based memory cells, operation principle 483, 484

CAS, *see* complementary atomic switch

CB, *see* conduction band

CBM, *see* conduction band minimum (CBM)

CBRAM, *see* conductive bridging random access memory (CBRAM)

cell-based geometrical model 238

cell-based percolation model 236, 237

CF, *see* conductive filament (CF)

chalcogenides 574–575

- materials, ECM memory 484, 486

charge transfer insulators 84

chemical diffusion 135, 135–136

CLBs, *see* configurable logic blocks (CLBs)

CMO, *see* conductive metal oxide (CMO)

CMOS transistors, *see* complementary metal-oxide-semiconductor (CMOS) transistors

compact modeling 399

- ECM 428–431, 430–432
- VCM 417–422, 418–421

complementary atomic switch (CAS)

- atomic switches 706, 707
- based programmable logic cell arrays 709–710, 710, 711
- cell architecture with 707–709, 708, 709
- features of 706
- OFF-state switch 707
- programmable switch 707

complementary metal-oxide-semiconductor (CMOS) transistors 547

complementary resistive switching (CRS) 329, 417

- combinations 644, 645
- mode 3, 4
- read operation of 645
- tantalum oxide-based single stack 647
- vertical integration 645–646

concentration polarization, *see* stoichiometry polarization

conductance

- changes in metallic filaments 207–210, 208, 209
- Landauer-Büttiker theory of 198

conductance quantization 198

- discover in 197, 198
- in ECM cells 204–205
- in metallic nanowires 197–204

conduction band (CB) 55, 61

conduction band minimum (CBM) 55, 61

conductive bridging random access memory (CBRAM) 484

- Butler-Volmer relationship 493–494
- Cu/Ta₂O₅ system 494–495
- electrode position process 492–493
- filament growth and dissolution
 - – Ag-Ge-S film, transmission electron micrograph of 496, 496
 - – bipolar programming 498
 - – Cu-doped SiO₂ device 499, 499–500
 - – limiting factor 497–498
 - – filament morphology
 - – Ag_{0.12}As_{0.35}S_{0.53} glass structure 501, 501
 - – 500 atom model 501, 501–502
 - – on-state resistance 502–503
 - – unconstrained dendritic electrodeposit 500–501, 501
- memory performance 505–506
- 1T-1R
 - – configuration 504–508, 507
 - – memory arrays 506–508, 507
 - – on-state resistance *vs.* programming current 503–505, 504, 505
- switching time *vs.* applied voltage 494, 495
- technological challenges 508–509
- voltage scaling 506

conductive filament (CF) 317

- different shapes of 335
- formation 298–300, 321, 328
 - – binary oxides 210
 - – in ionic conductors 201
 - – microscopic evidence 308–309, 309
 - – point-dynamics-triggering 306
- growth, voltage-driven model 323
- Joule heating 322
- localized conduction channel 327
- resistance 323
 - – estimation 324

conductive metal oxide (CMO) 647

conductivity diffusion experiment 139–141

configurable logic blocks (CLBs) 695

constant voltage stress (CVS) 226–227

continuum modeling 398–399

- ECM 426–428, 427
- VCM 410–417, 411, 413, 414, 416, 417

Coulomb blockade effect 212

Coulomb potential 65

crossbar array

- area efficiency 625
- devices selection
 - – MIEC, *see* mixed-ionic-electronic-conduction (MIEC)
 - – oxide diodes 631–633
 - – oxide tunnel barrier 638–639
 - – Si based 629–631

- – target specifications for 627–629
- – threshold switch 633–638
- – types 629–643
- memory device 623
- neuromorphic systems 719, 719
- NVM-read 627
- NVM-write 625–627, 626
- self-selected resistive memory device 643
- – CRS 644, 644–647, 646
- – hybrid ReRAM 647–649, 648
- – nonlinear ReRAM 649–650, 650
- subarrays 625
- cross-bar structure, gap-type atomic switches 523, 523–524
- cross-point arrays
- tunnel ReRAM device 473–475, 474
- URS 365, 365–366
- CRS, *see* complementary resistive switching (CRS)
- crystal
 - electrons in
 - – band structure model 60–62, 61
 - – computational details 66–67
 - – crystal field splitting 64, 64–65
 - – DOS and partial DOS 62–64, 63
 - – exchange and correlation 65–66, 66
 - – free electron model 58, 58–60, 59
 - – macroscopic properties of 49
 - – structure of metal oxides 50–54, 51–53, 54
- crystal field splitting 64, 64–65
- current-sweep-driven electroforming 291
- current-voltage behaviors, electroforming 290, 290
- current-voltage (I – V) characteristics
 - Nb-doped SrTiO₃ 463–464, 464
 - oxide-based atomic switch 531, 531
 - PCMO cells 459, 459–460
 - Pt/BFO/SRO cell 475–476, 477
 - SRO/LFO/BFO interface 476–477, 478
- Cu₂S-based gap-type atomic switch 521–522, 522
- Cu-SiO₂-based ReRAM device 345, 345
- CVS, *see* constant voltage stress (CVS), cyclability, *see* endurance
- cycling-induced degradation 600
- data retention, ferroelectric resistive switching 477, 479
- DBIE, *see* dielectric-breakdown-induced epitaxy (DBIE)
- DBIM, *see* dielectric-breakdown-induced metal migration (DBIM)
- de-doping effect 264
- defect generation 231–232
- anode hydrogen release model of 233–235, 234
- thermochemical model of 232–233
- defects
 - aluminum oxide, case study 116–117
 - – acceptor-doped alumina 117–118
 - – donor-doped alumina 118–119
 - definition 95
 - hafnium oxide, case study 116
 - one-dimensional/line defects 95–97, 96
 - point, *see* point defects
 - space-charge effects 103–104
 - – Gouy-Chapman situation 105–106
 - – Mott-Schottky situation 104–105, 105
 - strontium titanate, case study 110
 - – acceptor-doped 110–111
 - – donor-doped 111
 - – equations and constants 112
 - – grain boundaries 111–113
 - tantalum oxide, case study 119–121
 - three-dimensional 97
 - titanium oxide (rutile), case study 106, 107
 - – acceptor-doped 108
 - – donar-doped 108–109, 109
 - – nominally pure 107–108
 - – role of dislocations 109
 - two-dimensional 96, 97
 - zero-dimensional 95
 - zirconium oxide, case study 113–115
 - – role of grain boundaries and dislocations 115–116
- defect states, hopping transport via 172–174, 173
- dense film, growth kinetics of 276–278
- density functional theory (DFT) 65
- density of states (DOS), electrons in crystals 62–64, 63
- DFT, *see* density functional theory (DFT)
- dielectric breakdown
 - basics of 226–231
 - oxide 229, 231
 - QPC model for 212–213
 - for SiO₂-based MOS structures 228
 - thin insulator film 225
 - thin oxide films 225
- dielectric-breakdown-induced epitaxy (DBIE) 244
- dielectric-breakdown-induced metal migration (DBIM) 245
- diffusion coefficient
 - defects and ions 131–132
 - on thermodynamic parameters 128–129

- diffusion equation, ion transport 127–128, 128
- diffusion kinetics, in polycrystal 144
- diffusion-limited electrochemical process 282–283
- diffusion mechanism, ion transport 130–131, 131
- diffusion potentials
 - electromotive force 273–274
 - expression for 273
 - formation 273
 - disturb
 - anomalous behavior 616–617, 617
 - phenomena 615, 616
 - understanding and modeling 616, 616
- donor-doped SrTiO₃ cells, *see also* Nb-doped SrTiO₃ (Nb:STO)
- interface-type resistive-switching
 - cation vacancies 460–461
 - *I–V* curves 459, 459–460
 - oxygen-vacancy concentration 461
- DOS, *see* density of states (DOS)
- double injection model 9
- drift-diffusion model 399, 410, 415, 426, 465, 500
- dynamic random access memory (DRAM) 4, 354

- ECM, *see* electrochemical metallization (ECM) memory
- EDL, *see* electrical double layer (EDL)
- EE mechanism, *see* electronic excitation (EE) mechanism
- EELS, *see* electron energy loss spectroscopy (EELS)
- Einstein-Smoluchowski relation 129
- electrical breakdown 615
- electrical double layer (EDL) 260
 - electrostatic potential 264
- electrical transport models
 - band transport of carriers 168–169
 - electronic bandwidth 169
 - hopping transport via defect states 172–174, 173
 - metallic behavior 174
 - polaron formation 169–171, 170
 - polaron transport 171–172
 - thermopower 172
- electrochemical cell, of different half cells 269–270
- electrochemical metallization (ECM) memory 41–43, 42, 43, 44, 363
 - *ab initio* and MD models 422–423
- CBRAM, *see* conductive bridging random access memory (CBRAM)
- cells 318, 320, 329
- pulsed set switching kinetics 331
- compact modeling 428–431, 430–432
- conductance changes in metallic filaments 207–210, 208, 209
- conductance quantization in 204–205
- conductivity vs. Ag concentration 491, 491–492
- continuum modeling 426–428, 427
- current-voltage characteristics and initial device resistance 206–207, 207
- devices 318–319
- effect 12
- ion conducting materials
 - Anderson-Stuart model 486–487, 487
 - chemical dissolution 488–489, 489
 - Cu/SiO₂ interface 490
 - forming/conditioning process 486
 - thermal stability 486
 - types 484–485, 485
- ion transport 490–492
- KMC models 423–426, 424–426
- mechanisms 675–678, 676, 677
- nonlinear switching kinetics 399–401, 401
- switching mechanism 397
- electrochemical model 298
- electrochemical potential 258
- electrochemical process 329
- electrochemical reactions 280, 281
 - charge-transfer process limitations 280–282
 - diffusion-limited 282–283
- electrode materials 575–576
- ReRAM 575–576
- URS
 - anode-mediated reset operation 384–385
 - selection criteria 385–386
- electroforming 605
 - behavior
 - bipolar switching 291–294, 292, 293
 - factors influencing 294–295, 295, 296
 - in bipolar and unipolar switching 295–297
 - current overshoot, problems 310–311
 - forming-free resistive switching 311–312
 - mechanisms 297–298
 - conducting filament formation 298–300, 306, 308–309, 309
 - electrochemical 298
 - point defect 302–303
 - point defect dynamics 304–308, 306
 - purely electronic 298

- – redox reactions and ion/ionic defect migration 300–302
- method 290, 290–291
- phenomenological based 297
- technical issues 310
- voltage distribution 311
- electromotive force (emf)
 - causes
 - – by concentration differences 271–273, 272
 - – by surface curvature effects 270–271
 - diffusion potentials 273–274, 273
 - electrochemical cells of different half cells 269–270
- electron band-band transfer 100
- electron density distribution 49
- electron-electron correlation effects 178
- electron energy loss spectroscopy (EELS) 441, 444, 451–453
- electron gas, two-dimensional 197
- electronic bands 55
- electronic bandwidth, electrical transport model 169
- electronic devices, with memory 38–40
- electronic excitation (EE) mechanism 233
- electronic structure
 - amorphous states 75, 75–76
 - insulators 71, 71–72
 - magnetic insulators 82–84
 - – MnO 85–86, 86
 - – NiO 84–85, 85
 - metals 67–68, 68
 - nontransition metal oxides 76
 - – Al₂O₃ 76
 - – SrO 77, 78
 - – ZnO 77–78, 78
 - point defect states 72–73, 73
 - semiconductors 68–71, 69, 70
 - solids 54–55
 - – from free atoms to solid state 55–58, 56, 57
 - surface states 73–75, 74
 - titanates 79
 - – SrTiO₃ 82, 83
 - – TiO 79, 80
 - – TiO₂ 81–82, 82
 - – Ti₂O₃ 79–80
 - VB metal oxides
 - – metal-insulator transitions: NbO₂, VO₂, and V₂O₃ 86–87, 88
 - – tantalum oxides TaO_x 87–90, 89
 - electron-lattice coupling 71
 - electrons in crystals
 - band structure model 60–62, 61
 - computational details 66–67
 - crystal field splitting 64, 64–65
 - DOS and partial DOS 62–64, 63
 - exchange and correlation 65–66, 66
 - free electron model 58, 58–60, 59
 - electron spectroscopy for chemical analysis (ESCA), *see* X-ray photoelectron spectroscopy (XPS)
 - electron-transfer reaction 320, 330
 - potential energy profile of 281
 - Ellingham diagram, for binary oxides 90, 90–91
 - emf, *see* electromotive force (emf)
 - endurance
 - balancing 599, 599–600
 - degradation 600–601, 601, **601**
 - ferroelectric resistive switching 477–478, 479
 - high-endurance devices 598, **598**, 598
 - memory hierarchy position 598, 599
 - MIEC 640, 643
 - energy-dispersive X-ray spectroscopy (EDX) 441
 - enrichment and weak depletion layers 260–261
 - EPROM, *see* erasable programmable read-only memory (EPROM)
 - equilibrium thermodynamics 254–258, 256–259
 - of point defects 98, 98–99
 - erasable programmable read-only memory (EPROM) 10
 - EXAFS, *see* extended X-ray absorption fine structure (EXAFS)
 - extended X-ray absorption fine structure (EXAFS) 440
 - extrinsic point defects 95
 - impurities and dopants 102–103
 - reactions with environment 100–101
 - extrinsic variability 606
 - Faraday's law 330
 - fast grain-boundary diffusion 143
 - FeRAM 15
 - Fermi-Dirac distribution function 200
 - ferroelectric resistive switching
 - semiconducting oxides
 - – Pt/BFO/SRO cell 475–478, 477, 479
 - – SRO/LFO/BFO interface 476–477, 478
 - tunneling junction 475, 476
 - types 475
 - ferroelectric Schottky diodes 11

- FIB, *see* focused ion beam (FIB)
- Fick's first law 126, 127
- Fick's second law 126
- field programmable gate arrays (FPGAs)
- ASICs 695
 - CAS
 - - atomic switches 706, 707
 - - based programmable logic cell arrays 709–710, 710, 711
 - - cell architecture with 707–709, 708, 709
 - - features of 706
 - - OFF-state switch 707
 - - programmable switch 707
 - core programmable logic in 696
 - energy efficiency, of programmable logic accelerator 710–712
 - monolithically 3D, with BEOL devices 696, 696–698, 697
 - NRE cost 695
 - resistive configuration memory
 - - architecture 698, 699
 - - array 700–705
 - - cell 699–700, 700
- filamentary-based bipolar switching TMO RRAM, *see* bipolar switching TMO RRAM
- filamentary conduction
- and quantization effects in binary oxides 210–218
 - in SiO₂ 210
- filamentary switching 5
- ReRAM 317
 - URS 366–367
- filament formation, breakdown path and inference on 241–246, 242, 243
- filament-growth-controlled three-terminal atomic switches 537–539, 538, 539
- flade potential 279
- flash memory 10
- FN tunneling, *see* Fowler-Nordheim (FN) tunneling
- focused ion beam (FIB)
- milling 243
 - technique 441
- Fourier equation 324
- Fourier's law 127
- Fowler-Nordheim (FN) tunneling 649
- FPGA, *see* field programmable gate arrays (FPGAs)
- free electron model 58, 58–60, 59
- Frenkel defects 99
- anti-Frenkel defects 99–100
- Frenkel disorder 462
- Fuchs-Sondheimer approximation 324
- Fuchs-Sondheimer effects 327
- Fuchs-Sondheimer formula 320, 321
- Galvani potential 258
- gapless-type atomic switches
- moisture effects 533–534, 534
 - oxide-based switch
 - forming process 532, 532–533
 - - *I*–*V* characteristics 531, 531
 - - turn-ON and turn-OFF voltages 533
 - polymer-based switch 536–537, 537
 - quantized conductance 535, 535–536
 - sulfide-based switch 529–530, 530
 - switching time 534–535
 - synaptic behavior 535, 535–536
- gap-type atomic switches
- cross-bar structure 523, 523–524
 - electrochemical process 521–523, 522
 - logic-gate operation 526, 526–527
 - photo-assisted switch 528–529, 529
 - quantized conductance 524, 524–525, 525
 - scanning electron microscope image 522–523, 523
 - switching time 519–521, 520, 521
 - synaptic behavior 527–528, 528
- gate stack, failure defects 244
- GBs, *see* grain boundaries (GBs)
- generalized gradient approximation (GGA) 67
- GGA, *see* generalized gradient approximation (GGA)
- Gibbs-Curie condition 266
- Gibbs free energy 255
- and emf 270
 - for nucleation 267
 - oxide formation 276
- Gibbs phase rule 255
- Gibbs-Thomson equation 265, 267, 270
- Gouy-Chapman distribution 261
- Gouy-Chapman model 261, 262
- Gouy-Chapman type space-charge layers 265
- grain boundaries (GBs) 614–615
- oxygen-ion transport across 152
 - strontium titanate 111–113
- hafnium oxide, defects 116
- half-cell reactions 269–270
- half-filled mott insulators
- correlations and hubbard *U* 177–178
 - α -Fe₂O₃: 3d⁵ system 182–183, 183
 - MnO: 3d⁵ system 179, 180

- NiO: 3d⁸ system 179–182
- Hall effect 169, 182, 197
- hard breakdown (HBD) 226, 230
- hard X-ray photoelectron spectroscopy (HAXPES) 440, 441
- Al/PCMO films 445, 445–446
- Fe-doped SrTiO₃ films 442–443, 443
- Ti/PCMO/SRO films 468, 468–469
- HAXPES, *see* hard X-ray photoelectron spectroscopy (HAXPES)
- HBD, *see* hard breakdown (HBD)
- Hebb-Wagner polarization 115
- Helmholtz double layer 261, 262
- Henderson equation 274
- heterogeneous doping 265
- heterogeneous equilibrium 254–258, 256–259
- HfO_x/AlO_x bilayer ReRAM device 351, 352
- HfO_x based ReRAM device 345, 345
- high-resistance state (HRS) 2, 225, 666
 - degradation 600
 - SET/RESET operation 599
- homogeneous interface-type switching, *see* interface-type resistive-switching
- HRS, *see* high-resistance state (HRS)
- Hubbard model 65, 66, 83
- Hund's rule 85
- hybrid ReRAM 647–649, 648

- ICR, *see* interface controlled resistance (ICR)
- ICT devices, *see* Information and Communication Technologies (ICT) devices
- ideal memristor 34, 36
 - models 35
- incoherent thermal heating mechanism 234
- Information and Communication Technologies (ICT) devices
 - abstract binary element 548
 - essential properties 549
 - information states 547–548
 - READ operation 548–549, 549
 - TALK operation 548
 - WRITE operation 548
- insulating oxides 573–574
- insulator
 - Anderson 72
 - band 71
 - charge transfer 84
 - electronic structure 71, 71–72
 - Mott 72
 - Peierls 71

- insulator film, breakdown 225
- interface controlled resistance (ICR)
 - 556–557
 - contact resistance 560–563
 - material interfaces, electrical property 557–560
- interface-type resistive-switching 5
 - donor-doped SrTiO₃ cells
 - cation vacancies 460–461
 - I–V curves 459, 459–460
 - oxygen-vacancy concentration 461
 - ferroelectric RS effect, *see* ferroelectric resistive switching
 - mixed-valent manganites, *see also* Pr_{1-x}Ca_xMnO₃ (PCMO)
 - crystal structure 461
 - density of states *vs.* energy 461–462, 462
 - oxygen partial pressure *vs.* defect concentrations 462, 463
 - intrinsic point defects 72, 99
 - anti-Frenkel defects 99–100
 - anti-Schottky defects 100
 - electron band-band transfer 100
 - Frenkel defects 99
 - Schottky defects 100
 - intrinsic variability 606
 - inverse effective mass tensor 62
 - ion, diffusion coefficient 131–132
 - ion hopping
 - models 322
 - Mott-Gurney law for 330
 - ionic conductors, CF formation in 201
 - ionic-type ReRAM device, equivalent circuit model of 38, 38
 - ion/ionic defect migration 300–302
 - ionized-defect 322
 - Mott-Gurney model for 322
 - ion transport, in metal oxides
 - accelerated transport along extended defects 143–145
 - case study 145–147
 - alumina 153, 153–155
 - strontium titanate 147–150, 148, 150
 - tantalum pentoxide 155–156, 156
 - yttria-stabilized zirconia 150–153, 151
 - description 125–126
 - diffusion experiment types 134–135
 - chemical diffusion 135, 135–136
 - conductivity 139–141
 - tracer diffusion 137, 137–139, 138
 - hindered transport across extended defects 145, 146
 - macroscopic definition 126–127

- – diffusion coefficient 128–129
- – diffusion equation, solution 127–128, 128
- mass transport and across extended defects 141, 141–143, 143
- microscopic definition 129–130
- – activation barrier for migration 132–133, 133
- – diffusion coefficients of defects and ions 131–132
- – diffusion mechanism 130–131, 131
- irreversible thermodynamics 275
- I–V sweep mode 2, 10

- Jahn-Teller distortion effect 71, 461

- kinetic demixing 283, 284
- kinetic Monte Carlo (KMC) simulation methods 398
 - ECM 423–426, 424–426
 - VCM 407–410, 409
- Kirchhoff's laws 330
- KMC simulation methods, *see* kinetic Monte Carlo (KMC) simulation methods
- Kohn-Sham equation 65
- Kröger-Vink notation 72, 95, 302, 303, 305
- Kubo formalism 39

- Landauer-Büttiker theory 202
 - of conductance 198
- Landauer formulation 212
- lateral channel structures 578, 578–579
- LBL, *see* left bit line (LBL)
- LDA, *see* local density approximation (LDA)
- left bit line (LBL) 701
- line defects, *see* one-dimensional defects
- local density approximation (LDA) 67
- local Joule heating 399
- logic-gate operation, gap-type atomic switches 526, 526–527
- long-term potentiation (LTP) 527–528, 528
- low-resistance state (LRS) 2, 205, 214, 666
 - degradation 600
 - SET/RESET operation 599
- LRS, *see* low-resistance state (LRS)
- LTP, *see* long-term potentiation (LTP)

- Madelung potential 73
- magnetic insulators, electronic structure 82–84
- MnO 85–86, 86
- NiO 84–85, 85
- magnetic RAMs (MRAMs) 13
- MD models, *see* molecular dynamics (MD) simulation models
- memcapacitive system 36
- memcapacitors 36
- meminductive system 36
- meminductors 36
- memory
 - capacitors 36
 - circuit elements 36, 37
 - electronic devices with 38–40
 - inductors 36
 - resistors 36
- memristive devices 718
- memristive element 23
- memristive model 31, 35, 42
- memristive system 33–36, 41
- memristor 34, 542
 - ideal 34, 36
- memristor model, ideal 35
- metal
 - electrodes, oxidation and passivation of 279
 - electronic structure 67–68, 68
 - excess in oxide 101
 - monovalent 202
 - oxidation 101, 275
 - oxidation rate of 278–279
- metal-insulator-metal (MIM) structure 225, 289
 - Nb-doped SrTiO₃ 465
 - ReRAM cells 4–5
 - URS 364, 365
 - VCM switching 440
- metal-insulator transition (MIT) 636–637, 637
 - electronic structure 86–87, 88
 - temperature-induced 191
 - in VO 192
- metallic behavior, electrical transport models 174
- metallic filaments, conductance changes in 207–210, 208
- metallic nanowires, conductance quantization in 197–204
- metal oxide (MeO)
 - crystal structure 50–54, 51–53, 54
 - reaction of formation of 255
- metal oxide (MeO), ion transport in
 - accelerated transport along extended defects 143–145
 - case study 145–147

- – alumina 153, 153–155
- – strontium titanate 147–150, 148, 150
- – tantalum pentoxide 155–156, 156
- – yttria-stabilized zirconia 150–153, 151
- description 125–126
- diffusion experiment types 134–135
- – chemical diffusion 135, 135–136
- – conductivity 139–141
- – tracer diffusion 137, 137–139, 138
- hindered transport across extended defects 145, 146
- macroscopic definition 126–127
- – diffusion coefficient 128–129
- – diffusion equation, solution 127–128, 128
- mass transport and across extended defects 141, 141–143, 143
- microscopic definition 129–130
- – activation barrier for migration 132–133, 133
- – diffusion coefficients of defects and ions 131–132
- – diffusion mechanism 130–131, 131
- metal-oxide-semiconductor (MOS) devices 211
 - SBD $I-V$ characteristics 212, 214
- metal-oxygen octahedron 167, 168
- MIEC, *see* mixed-ionic-electronic-conduction (MIEC)
- MIM structure, *see* metal-insulator-metal (MIM) structure
- minimal nanoionic memory devices
 - atomic filament
 - – barrier electrical conductance 553–554
 - – classical resistance 551–552
 - – minimal atomic bridge estimation parameters 552
 - – nonrectangular barrier 555, 555
 - – quantum resistance 552–553, 553
 - – transmission through atomic gaps 555–556, 556
 - heat dissipation and transfer 567–569
 - ICR 556–557
 - – contact resistance 560–562, 560–563, 563
 - – material interfaces, electrical properties of 557–560
 - – material parameters 557
 - ICT, *see* Information and Communication Technologies (ICT) devices
 - stability 563–565, 564, 565, 566
 - switching mechanisms 549–551, 550
 - switching speed and energy 565–567
- MIR limit condition, *see* Mott-Ioffe-Regel (MIR) limit condition
- mixed-ionic-electronic-conduction (MIEC)
 - array diagnostic monitor 640
 - cumulative distribution functions 640, 641
 - endurance 640, 643
 - integration 640
 - select devices 639–640
 - thickness scaling experiments 640
- mixed-valent manganite resistive-switching, *see also* $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ (PCMO)
 - crystal structure 461
 - density of states *vs.* energy 461–462, 462
 - oxygen partial pressure *vs.* defect concentrations 462, 463
- MnO , electronic structure 85–86, 86
- molecular dynamics (MD) simulation 75, 152
 - molecular dynamics (MD) simulation models 398
 - ECM 422–423
 - VCM
 - – electron transfer, 404, 404–405
 - – LRS and HRS state 402, 402–404, 403
 - – migration barriers 406–407
 - – phase transformations and nucleation 405–406, 406
- molecular orbital (MO) approach 55
- monovalent metals 202
- Moore’s law 728
- MO oxides, cubic rock-salt structure 166
- M_2O_3 oxides, rhombohedral corundum structure 166
- MO_2 oxides, tetragonal rutile structure 166
- MOSFET 10
- Mott-Gurney law, for ion hopping 330
- Mott-Gurney model 322
- Mott-Hubbard gap 66
- Mott insulators 72
- Mott-Ioffe-Regel (MIR) limit condition 174
- Mott-Schottky-type space-charge layer 261, 262, 262
- MRAMs, *see* magnetic RAMs (MRAMs)
- multilevel cell (MLC) operation, of ReRAM 319
- multiplexer (MUX) 697, 697
- MUX, *see* multiplexer (MUX)
- NAND flash 19–20, 473, 582, 588, 589
- nanofountain pen with Ag ink 516, 516
- nanoionic memory devices
 - atomic filament

- – barrier electrical conductance 553–554
- – classical resistance 551–552
- – minimal atomic bridge estimation parameters 552
- – nonrectangular barrier 555, 555
- – quantum resistance 552–553, 553
- – transmission through atomic gaps 555–556, 556
- heat dissipation and transfer 567–569
- ICR 556–557
- contact resistance 560–562, 560–563, 563
- material interfaces, electrical properties of 557–560
- material parameters 557
- ICT, *see* Information and Communication Technologies (ICT) devices
- stability 563–565, 564, 565, 566
- switching mechanisms 549–551, 550
- switching speed and energy 565–567
- nanosecond switching dynamics 353, 353–354, 354
- nanosize effects
 - due to surface curvature 265
 - on space-charge regions 263–265, 264
- nanowires (NWs)
 - conductance quantization in metallic 197–204
 - core-shell 680–681
 - ReRAM
 - concept of 681
 - core-shell metal-oxide 683
 - single all-oxide 664–665
 - structure 661–662
 - template-assisted fabrication methods of 663–664
- nano-XAFS 439, 450
- Nb-doped SrTiO₃ (Nb:STO)
 - counter-eightwise polarity 465
 - drift-diffusion model 465
 - *I*–*V* characteristics 463–464, 464
 - resistance-state retention characteristics 464, 464
 - resistance values, area dependence of 457–458, 458
 - Schottky depletion model 464–465
 - SPM 465
- NbO₂, electronic structure 86–87, 88
- NDR effect, *see* negative differential resistance (NDR) effect
- negative differential resistance (NDR) effect 297–298
- Nernst equation 271, 281
- neurocomputing 716
- neuromorphic systems
 - computing 716–718
 - engineering 715–716
 - ReRAM
 - applications 718–719, 729–730, 730
 - capacitive properties 725, 726
 - conductance update 720, 720
 - crossbar array 719, 719
 - hardware synaptic functions 718
 - multilevel resistance behavior 722–725, 723, 724
 - nonlinear switching kinetics 722
 - nonvolatility, of resistive states 721–722, 722
 - scaling 728–729
 - switching statistics 725–727, 727, 728
 - synaptic array 720, 721
 - tunable resistive state 719
 - volatility, of resistive states 721–722, 722
 - scaling in 728–729
- NiO
 - electronic structure 84–85, 85
 - memory cells, resistance values 457–458, 458
- NiO_x ReRAM device 346, 346–348, 348, 349, 350
- nonlinear ReRAM 649–650, 650
- nonlinear switching kinetics 319, 319, 320
- nonrecurring engineering (NRE) cost 695
- nontransition metal oxides, electronic structure 76
 - Al₂O₃ 76
 - SrO 77, 78
 - ZnO 77–78, 78
- nonvolatile memory (NVM), 16
- nonvolatile resistive-switching behaviors 289
- nonvolatility 1
- nucleation 266
 - atomistic theory 267–269, 268, 269
 - Gibbs free energy for 267
 - macroscopic view 266–267
- nucleation-controlled three-terminal atomic switches 539–541, 540
- nucleation theory 266
- NVM, *see* nonvolatile memory (NVM)
- NVM-read 627
- NVM-write
 - biasing scheme 626–627
 - during operations 626–627
 - voltage drops 626, 626
- NWs, *see* nanowires (NWs)

- OFF-state leakage current devices 628
- OFF states, *see* low-resistance state (LRS)
- Ohm's law 139
- one-dimensional defects 95–97, 96
- 1D1R nanopillar array 686, 686–687
- 1S1R stacked crosspoint architectures 583
 - cumulative line resistance 586–588, 587
 - selector device 583
 - sensing margin 584, 584–586, 585
 - write margin 586, 586
- one transistor-one resistor (1T-1R) memory arrays, CBRAM 504–508, 507
- 1T2R cell 699
- ON-state current density 628
- ON-state resistance modeling 502–503
- ON states, *see* high-resistance state (HRS)
- optical band gap 55
- orbitals and metal-insulator transitions 184–185, 185
- OTS, *see* ovonic threshold switching (OTS)
- ovonic threshold switching (OTS) 634–636, 636
- oxidation
 - metal 101, 275
 - and passivation of metal electrodes 279
 - rate of metals 278–279
- oxide
 - based VCM-type ReRAM 13
 - BD testing 226
 - breakdown 229, 231
 - categories of 165
 - crystal structures of 166–167
 - layers, growth modes 278
 - metal excess in 101
- oxide-based atomic switch
 - forming process 532, 532–533
 - *I*–*V* characteristics 531, 531
 - turn-ON and turn-OFF voltages 533
- oxide diodes
 - metal-oxide Schottky barrier 632–633, 634
 - PN junction 631–632
- oxide film
 - dielectric breakdown 225
 - formation 276, 277, 278
- oxide materials
 - ferroelectric resistive switching
 - Pt/BFO/SRO cell 475–478, 477, 479
 - SRO/LFO/BFO interface 476–477, 478
 - URS
 - conical CF configurations 380, 381
 - RCB model 381–382, 382
 - switching variability and endurance 383, 383–384
 - URS/BRS coexistence 382–383, 383
- oxide tunnel barrier
 - multi-layer oxide-based select device 638–639
 - single layer oxide-based select device 638
- oxygen
 - excorporation reaction 100–101
 - gas formation 300, 302
 - incorporation reaction 101
- oxygen-ion
 - diffusion 152
 - transport 152
- partial density of states
 - electrons in crystals 62–64, 63
 - surface states 74
- passivation potential 279
- PBD, *see* progressive breakdown (PBD)
- PCMs, *see* phase-change memories (PCMs)
- PCRAM 15
- PEEM, *see* photoemission electron microscopy (PEEM)
- Peierls insulator 71
- percolation model
 - cell-based 236, 237
 - development 240
 - multilayer 241
 - to SET transition in ReRAM 237–238, 238
- PF carrier transport, *see* Poole-Frenkel (PF) carrier transport
- phase-change materials 575
- phase-change memories (PCMs) 12
- photo-assisted atomic switch 528–529, 529
- photoemission electron microscopy (PEEM) 441
- pinched hysteresis loop 33, 34
- planar stack structures 574, 576–577
- PMC, *see* programmable metallization cell (PMC)
- point defects 95
 - Brouwer diagram 101–102, 102
 - equilibrium thermodynamics of 98, 98–99
 - extrinsic defects
 - impurities and dopants 102–103
 - reactions with environment 100–101
 - intrinsic defects 72, 99
 - anti-Frenkel defects 99–100
 - anti-Schottky defects 100
 - electron band-band transfer 100

- – Frenkel defects 99
- – Schottky defects 100
- point defect states, electronic structure 72–73, 73
- Poisson-Boltzmann differential equation 104, 260
- polaron
 - formation 169–171, 170
 - transport 171–172
- polycrystal
 - diffusion kinetics in 144
 - mass-transport processes in 141, 141, 143
- polymer-based atomic switch 536–537, 537
- Poole-Frenkel (PF) carrier transport 319, 322
- Poole-Frenkel conduction model 374–375
- Poole-Frenkel effects 327
- Poole-Frenkel emission 226
- potentiodynamic method 290, 290–291, 293, 293
- power law model 129, 229
- programmable metallization cell (PMC) 484
- progressive breakdown (PBD) 230, 231
- $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ (PCMO)
 - cross-sectional TEM images 466, 466
 - $C-V$ curves 467, 467
 - $I-V$ curves 459, 459–460
 - Ti/PCMO/SRO cells 466, 466–467, 467
 - tunneling barriers 458, *see also* tunnel ReRAM device
 - XPS spectra 468, 468–469
- Pt/NiO/Pt URS system
 - model implications 375–379
 - NiO defect chemistry 368–370
 - reset switching modeling 372–373
 - set switching modeling 373–375
 - time-dependent dielectric-breakdown methodology 371–372
- QPC model, *see* quantum point contact (QPC) model
- quantization effects, in binary oxides 210–218
- quantized conductance
 - gapless-type atomic switches 535, 535–536
 - gap-type atomic switches 524, 524–525, 525
- quantum Hall effect 197
- quantum point contact (QPC) model 215, 238
 - for dielectric breakdown 212–213
- quantum resistance, AF 552–553, 553
- quasistatic current-voltage ($I-V$) characteristics
 - apparatus topology 342, 343
 - current compliance and overshoot effects 346–350, 346–350
 - electroforming parameters 344, 344
 - four-wire probe configuration 342–343, 343
 - nonlinearity 344
 - switching parameters 343, 343
 - – *vs.* sweep rates 345, 345–346, 346
 - tunnel ReRAM device 470, 471
 - two-wire configuration 342, 343
- radio frequency (RF) switches 483
- ramp voltage stress (RVS) 227, 238, 606
- random circuit breaker (RCB) model 381–382, 382
- random telegraph noise (RTN)
 - CBRAM 609–610
 - charge carrier transport-induced 610–611, 611
 - detection, of signals 609
 - fluctuations 215
 - key mechanisms of 615
 - non-steady-state fluctuations 609
 - oxide-based RRAM (OXRAM), 609
 - oxygen vacancy transport-induced
 - – bimodal filament configuration 614
 - – dielectric microstructure role 614–615
 - – disturb immunity 614
 - – identification, of vacancy perturbations 611–612, 612
 - – shallow to moderate reset 612–613, 613
 - – very deep reset 613–614, 614
 - – steady-state fluctuations 609
 - – voltage conditions 609
 - RBL resistance, *see* right bit line (RBL) resistance,
- RCB model, *see* random circuit breaker (RCB) model
- redox-based resistive switching memories, *see* resistive random-access memory (ReRAM/RRAM) devices
- redox reactions 300–302
- reductants 575–576
- ReRAM, *see* resistive random-access memory (ReRAM/RRAM) devices
- resistance switching (RS) materials 289
 - chalcogenides 574–575
 - insulating oxides 573–574
 - phase-change materials 575
 - semiconducting oxides 574

- resistive configuration memory
 - architecture 698, 699
 - array
 - - four operation modes 701
 - - measurement 703–705, 704, 705
 - - prototype 702–703, 702–703
 - - reset RL 701
 - cell 699–700, 700
- resistive random-access memory
 - (ReRAM/RRAM) devices 31–32, 37, 197, 253, 297, 310, 311, 363
 - array area efficiency 592–593, 593
 - axial hetero structured NWs 678–680, 679, 680
 - based artificial synapses 23, 24
 - behavior of 319
 - black-box modeling 31, 40, 40
 - bottom-up fabrication methods 661
 - - framework of 662
 - - template-assisted, of NW 663–664
 - - VLS 662–663
 - breakdown statistics for 237–241, 238, 239, 241
 - circuits 21–24
 - core-shell NWs 680–681
 - - advantages of 680–681
 - - Ag electrodes 681–682, 682
 - - Ni electrodes 683–685, 683–686
 - - Ni/NiO core/shell 683–685, 683–686
 - - Si core and a-Si shell 681–682, 682
 - current compliance and overshoot effects 346–350
 - ECM devices 41–43, 42, 43, 44
 - effect in binary oxides 216
 - electrode materials 575–576
 - emerging bottom-up approaches and applications
 - - block-copolymer self-assembly for 687–688
 - - 1D1R nanopillar array 686, 686–687
 - filamentary-switching 317
 - FPGAs, *see* field programmable gate arrays (FPGAs)
 - geometrical location of the switching event in 5, 5
 - gray-box modeling 40, 40
 - history 9–12
 - hybrid 647–649, 648
 - ionic-type 38, 38
 - lateral channel structures 578, 578–579
 - with load resistor 323
 - logic circuits
 - - FPGA 21–22
 - - logic gates 22
 - neuromorphic 22–23
 - material categories 574
 - material implication (IMP) gate implement 22, 23
 - mechanisms of switching 40–41
 - memory applications
 - - device isolation in crossbar arrays 16–19, 17, 19
 - - 3-D technology 19–20, 20
 - - memory hierarchy 20–21, 21
 - - performance requirements and parameter windows 12–15, 16
 - metal-insulator-metal structure 4–5
 - microscopic analysis of 4
 - multilevel cell operation of 319
 - nanometer dimensions of 263
 - neuromorphic systems
 - - applications 718–719, 729–730, 730
 - - capacitive properties 725, 726
 - - computing 716–718
 - - conductance update 720, 720
 - - crossbar array 719, 719
 - - engineering 715–716
 - - hardware synaptic functions 718
 - - multilevel resistance behavior 722–725, 723, 724
 - - nonlinear switching kinetics 722
 - - nonvolatility, of resistive states 721–722, 722
 - - scaling 728–729
 - - switching statistics 725–727, 727, 728
 - - synaptic array 720, 721
 - - tunable resistive state 719
 - - volatility, of resistive states 721–722, 722
 - noise spectra 342
 - nonlinear switching kinetics 319, 319, 320, 649–650, 650
 - NW structure 661–662
 - 1S1R stacked cross point architectures 583
 - - cumulative line resistance 586–588, 587
 - - selector device 583
 - - sensing margin 584, 584–586, 585
 - - write margin 586, 586
 - operation modes of 2, 3, 4
 - parameters of 4
 - planar stack structures 574, 576–577
 - properties and behavior 317–320, 319
 - prospects and challenges 24–25
 - pulsed measurements 350–353
 - - basic setup 350, 351

- – nanosecond timescale 353, 353–354, 354
- – ON and OFF switching dynamics 350, 351, 352
- – subnanosecond time resolution 354–358, 355–358
- – quasistatic measurement, *see* quasistatic current-voltage (*I*–*V*) characteristics
- – reliability aspects
- – characteristics 597
- – different memory applications 597
- – new emerging memory technology 597
- – reliability aspects, bipolar switching TMO RRAM
- – disturb 615–617
- – endurance 598–601
- – retention 601–605
- – RTN, *see* random telegraph noise (RTN)
- – variability 605–609
- – RS materials, *see* resistance switching (RS) materials
- – selection
- – hybrid ReRAM 647–649, 648
- – 1S1R stacked cross point architectures 583
- – oxide diodes 631–633
- – oxide tunnel barrier 638–639
- – Si based 629–631
- – target specifications for 627–629
- – threshold switch 633–638
- – types 629–643
- – selection, mixed-ionic-electronic-conduction
- – array diagnostic monitor 640
- – cumulative distribution functions 640, 641
- – endurance 640, 643
- – integration 640
- – select devices 639–640
- – thickness scaling experiments 640
- – sidewall-conforming stack structures 577–578, 578
- – single all-oxide NWs 664–665
- – ECM mechanisms 675–678, 676, 677
- – electrical polarity characteristics 674, 675
- – identification, of switching spatial location 672, 673
- – latter approach 671, 672
- – NiO 665–668, 665–669
- – planar-type 671, 671
- – pressure dependence 673, 674
- – SEM image of 669, 670
- – stoichiometry polarization 283–285
- – through-multilayer via array
- – bathtub-type connection 592, 592
- – horizontal electrodes 589–592, 590, 591
- – multiple electrode layers 588, 588–589
- – staircase connections 589, 590
- – transistor in parallel architectures 582, 582–583
- – transistor in series architectures 579–581, 579–582
- – transition in bipolar 321
- – as two-terminal elements 40–41
- – universal switching 317–318, 318, 320–321
- – VCM devices 43–46
- – voltage-driven kinetics of switching 322, 323
- – white-box modeling 31, 40, 40
- – write speed 14
- resistor transistor logic (RTL) codes 709–710, 710
- retention 4
- – degradation in 603–604
- – ICR cell 566
- – LRS retention 602, 602
- – oxygen vacancy filamentary 601, 602
- – trade-off between 604–605
- rhenum trioxide structure 53, 53
- right bit line (RBL) resistance 13, 701–702
- RRAM, *see* resistive random-access memory (ReRAM/RRAM) devices
- RS materials, *see* resistance switching (RS) materials
- RTN, *see* random telegraph noise (RTN)

- SBD, *see* soft breakdown (SBD)
- scanning probe microscopy (SPM) technique, Nb-doped SrTiO₃ 465
- scanning transmission X-ray microscopy ((S)TXM) 440
- Pt/TiO₂/Pt resistive-switching device 448–449, 449, 450
- scattering-degraded resistivity 324
- Schottky defects 100
- Schottky depletion model 464–465
- Schottky-like barrier, ferroelectric diode 475, 476
- Schottky-like metal/conducting oxide interfaces 459, 459–460
- Schottky lowering 571
- Schrödinger-like equation 65
- SCL, *see* space-charge layer (SCL)

- self-doping reaction 302
- self-select feature 469
- semiconducting oxides 574
- semiconductor parameter analyzers (SPAs) 341, 342, 344, 350
- semiconductors
 - electronic structure 68–71, 69, 70
 - wide-band-gap 71
- sensing margin, 1S1R stacked cross point architectures 584, 584–586, 585
- SET operation 2–3
- short-term plasticity (STP) 527–528, 528
- Si-based select devices
 - bidirectional switches 630, 631
 - diodes 630
 - MSM diode 631
 - PN diodes 630
 - transistors 629
- sidewall-conforming stack structures 577–578, 578
- Simmons equation 330
- single-electron tunneling (SET) mechanism 212
- soft breakdown (SBD) 225, 226, 230
- solid electrolyte 258, 271–272
- solid, electronic structure 54–55
 - from free atoms to solid state 55–58, 56, 57
- solid-state reactions 275–279
- Sommerfeld model, *see* free electron model
- source measurement units (SMUs) 342, 343
- space-charge effects, defects 103–104
 - Gouy-Chapman situation 105–106
 - Mott-Schottky situation 104–105, 105
- space-charge layer (SCL) 103, 262, 264
 - driving force for formation of 258–260
- SPAs, *see* Semiconductor parameter analyzers (SPAs)
- SPICE modeling approach 41
- spike-timing-dependent plasticity (STDP) 23, 719
- spinel oxides 167
- SpiNNaker 717
- spin-polarized approach 67
- SRAM, *see* static random access memory (SRAM)
- SrO, electronic structure 77, 78
- SrTiO₃, electronic structure 82, 83
- staircase connections, through-multilayer via array 589, 590
- static random access memory (SRAM) 15, 354
- STDP, *see* spike-timing-dependent plasticity (STDP)
- Stern layer 261
- stoichiometry polarization 283–285
- Storage class memory 624
- strong depletion layers 261–263
- strontium titanate
 - defects 110
 - – acceptor-doped 110–111
 - – donor-doped 111
 - – equations and constants 112
 - – grain boundaries 111–113
 - – ion transport, in metal oxides 147–150, 148, 150
- STTRAM 15
- subnanosecond switching dynamics 354–358, 355–358
- sulfide-based atomic switch 529–530, 530
- surface curvature
 - effects, emf cause 270–271
 - nanosize effects due to 265
- surface states, electronic structure 73–75, 74
- switching time
 - CBRAM 494, 495
 - gapless-type atomic switches 534–535
 - gap-type atomic switches 519–521, 520, 521
- synaptic behavior
 - gapless-type atomic switches 535, 535–536
 - gap-type atomic switches 527–528, 528
- tantalum oxide
 - defects 119–121
 - TaO_x, electronic structure 87–90, 89
- tantalum pentoxide, ion transport 155–156, 156
- TaO_x-based ReRAM device 345–346, 346, 350, 351, 357, 357–358, 358
- TCM, *see* thermochemical memory (TCM)
- temperature-dependent conductivity 173
- temperature-induced metal-insulator transitions in oxides
 - Fe₃O₄: mixed-valent system 190, 190–191
 - limitation 191–192
 - orbitals and metal-insulator transitions 184–185, 185
 - Ti₂O₃: 3d¹ system 187–188
 - VO₂: 3d¹ system 186–187, 189
 - V₂O₃: 3d² system 189–190
- template-assisted fabrication methods, of NW 663–664
- thermal resistance 324
- thermochemical mechanism cells 41

- thermochemical memory (TCM) 289
 - cells 41
 - devices 318
 - mechanism 363
- thermochemical model, of defect generation 232–233
- thermodynamics
 - basic laws 254
 - classical 254, 274
 - equilibrium 254–258, 256–259
 - irreversible 275
 - parameter, diffusion coefficient 128–129
 - of point defects 98, 98–99
- thermopower, electrical transport models 172
- thin insulator film, dielectric breakdown 225
- three-dimensional defects 97
- 3D monolithic FPGA
 - conventional 696
 - ReRAM-based configuration memory 698
 - 8-to-1 routing MUX 697, 697
- 3D multilayer stacking 628
- 3D technology, ReRAM application in 19–20, 20
- three-terminal atomic switches
 - filament-growth 537–539, 538, 539
 - nucleation 539–541, 540
- threshold switch 633
 - MIT 636–637, 637
 - OTS 634–636, 636
 - TVS 637–638
- threshold vacuum switch (TVS) 637–638
- through-multilayer via array
 - bathtub-type connection 592, 592
 - horizontal electrodes 589–592, 590, 591
 - multiple electrode layers 588, 588–589
 - staircase connections 589, 590
- TiN/NiO/Ni cells 373, 385, 386
- TiO₂
 - electronic structure 81–82, 82
 - material, URS 380
 - ReRAM device 354, 354
- TiO, electronic structure 79, 80
- Ti₂O₃, electronic structure 79–80
- TiO_x-based ReRAM device 344, 344, 348, 349
- titanates, electronic structure 79
 - SrTiO₃ 82, 83
 - TiO 79, 80
 - TiO₂ 81–82, 82
 - Ti₂O₃ 79–80
- titanium oxide (rutile), defects 106, 107
 - acceptor-doped 108
 - donor-doped 108–109, 109
- nominally pure 107–108
- role of dislocations 109
- TMOs, *see* transition metal oxides (TMOs)
- tracer diffusion 137, 137–139, 138
- transistor
 - in parallel architectures 582, 582–583
 - in series architectures 579–581, 579–582
- transition metal oxides (TMOs) 289, 291, 367–368
 - electronic-carrier types of 304
 - hypostoichiometric 303
 - layer, oxygen vacancies from 304–305
 - structure
 - bonding and electronic structure 167–168
 - crystal structures of oxides 166–167
- transmission X-ray microscopy (TXM) 440
 - Pt/TiO₂/Pt resistive-switching device 448–449, 449
- transport process, and chemical reactions 274–275
- TrueNorth, neural network chip 717
- tunnel ReRAM device
 - advantage 469
 - challenges 469
 - deposition time *vs.* current density 470–471, 472
 - device characteristics 471, 472
 - mechanism and modeling 472–473, 473
 - passive cross-point arrays 473–475, 474
 - quasistatic *I*–*V* characteristics 470, 471
 - read currents and read current densities 470, 471
 - structure and TEM cross section 469–470, 470
- two-dimensional defects 96, 97
- two-dimensional electron gas 197
- two-terminal device models 32
 - dynamical systems approach 33
 - memcapacitive system 36
 - meminductive system 36
 - memristive system 33–36
 - memristor 34
 - ideal circuit element approach 32
 - lumped elements 32
- two-terminal element, ReRAM devices as 40–41
- two-terminal memristive elements 1–6
- two-terminal resistive switching material systems 11
- TXM, *see* transmission X-ray microscopy (TXM)

- unipolar resistive-switching (URS) 343, 343
 - basic operation 364
 - cross-point architecture 365, 365–366
 - electrode materials
 - – anode-mediated reset operation 384–385
 - – selection criteria 385–386
 - filamentary switching 366–367
 - metal-insulator-metal structures 364, 365
 - mode 2, 3, 4
 - oxide materials
 - – conical CF configurations 380, 381
 - – RCB model 381–382, 382
 - – switching variability and endurance 383, 383–384
 - – URS/BRS coexistence 382–383, 383
 - Pt/NiO/Pt system
 - – model implications 375–379
 - – NiO defect chemistry 368–370
 - – reset switching modeling 372–373
 - – set switching modeling 373–375
 - – time-dependent dielectric-breakdown methodology 371–372
 - TiO₂ material 380
 - TMO materials 367–368
- URS, *see* unipolar resistive-switching (URS)

- valence band maximum (VBM) 55, 61
- valence change mechanism 43–46, 210
- valence change memory (VCM) 289, 363
 - *ab initio* and MD models
 - – electron transfer 404, 404–405
 - – LRS and HRS state 402, 402–404, 403
 - – migration barriers 406–407
 - – phase transformations and nucleation 405–406, 406
 - bipolar switching behavior 297
 - cells 318, 319
 - compact modeling 417–422, 418–421
 - continuum modeling 410–417, 411, 413, 414, 416, 417
 - effect 11, 12
 - KMC methods 407–410, 409
 - materials 289–290
 - nonlinear switching kinetics 399–401, 401
 - spectroscopy techniques
 - – metal/oxide interface 442–446
 - – schematic overview 439
 - – TMOs, localized switching effects in 446–453
 - switching mechanism 395–396
 - voltage/current sweep 290, 290

- valence charge density distribution, metal 68, 69, 70
- vapor-liquid-solid method (VLS) 662–663
- variability
 - in bipolar switching devices
 - – ion-movement model 609
 - – microscopic origin 607–608, 608
 - – modeling and physical description 608
 - – particular oxide 607
 - – resistance window 607
 - of device characteristics 605
 - of forming operation 605–606
 - intrinsic and extrinsic 606
 - resistive memory 605
 - in unipolar devices 607
- variable-diameter model, universal switching characteristics 321–329, 334
- variable-gap model, universal switching characteristics 329–334
- variable-gap/variable-diameter states, coexistence of 334, 334–337, 336
- VBM, *see* valence band maximum (VBM)
- VB metal oxides, electronic structure
- metal-insulator transitions: NbO₂, VO₂, and V₂O₃ 86–87, 88
- tantalum oxides TaO_x 87–90, 89
- VCM, *see* valence change memory (VCM)
- VEs, *see* vibrational excitations
- vibrational excitations (VEs) 234
- VLS, *see* vapor-liquid-solid method (VLS)
- VO₂, electronic structure 86–87
- V₂O₃, electronic structure 86–87
- voltage compatibility, with memory element 628
- voltage-driven kinetics of switching 322, 323

- Wadsley defects 257, 258
- Wagner's parabolic law 278
- Weibull distribution 227, 236
- Wentzel-Kramers-Brillouin (WKB)
 - approximation 200
- wide-band-gap semiconductors 71
- wireless sensor nodes (WSNs) 710, 711
- WKB approximation, *see* Wentzel-Kramers-Brillouin (WKB)
 - approximation
- write margin, 1S1R stacked cross point architectures 586, 586
- WSNs, *see* wireless sensor nodes (WSNs)

- XAFS, *see* X-ray absorption fine structure spectroscopy (XAFS)

- XANES, *see* X-ray absorption near-edge spectroscopy (XANES)
XAS, *see* X-ray absorption spectroscopy (XAS)
XPS, *see* X-ray photoelectron spectroscopy (XPS)
X-ray absorption fine structure spectroscopy (XAFS) 439
X-ray absorption near-edge spectroscopy (XANES) 439
– Cr-doped SrTiO₃ 446–447, 447
X-ray absorption spectroscopy (XAS) 439, 441
– Ta₂O₅ resistive-switching devices 450–451, 451
X-ray photoelectron spectroscopy (XPS) 440–442, 444, 445
- Al/PCMO films 443–444, 444
– spectra, Ti/PCMO/SRO films 468, 468–469
– Ti/PCMO films 445
- YSZ, ion transport, *see* yttria-stabilized zirconia (YSZ), ion transport
yttria-stabilized zirconia (YSZ), ion transport 150–153, 151
- Zandt-Honig-Goodenough model 81
zero-dimensional defects 95
zirconium oxide, defects 113–116
ZnO, electronic structure 77–78, 78