

ENERGY-EFFICIENT DISTRIBUTED COMPUTING SYSTEMS

**WILEY SERIES ON PARALLEL
AND DISTRIBUTED COMPUTING**

Editor: Albert Y. Zomaya

A complete list of titles in this series appears at the end of this volume.

ENERGY-EFFICIENT DISTRIBUTED COMPUTING SYSTEMS

Edited by

Albert Y. Zomaya
Young Choon Lee

IEEE
 **computer**
society

 **WILEY**

A JOHN WILEY & SONS, INC., PUBLICATION

Cover Image: Baris Simsek/iStockphoto

Copyright © 2012 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Zomaya, Albert Y.

Energy-efficient distributed computing systems / Albert Y. Zomaya, Young Choon Lee.

p. cm.

ISBN 978-0-470-90875-4 (hardback)

1. Computer networks—Energy efficiency. 2. Electronic data processing—Distributed processing—Energy conservation. 3. Green technology. I. Lee, Young Choon, 1973— II. Title.

TK5105.5.Z66 2012

004'.36—dc23

2011042246

Printed in the United States of America

ISBN: 9780470908754

10 9 8 7 6 5 4 3 2 1

To our families for their help, support, and patience.

CONTENTS

PREFACE	xxix
ACKNOWLEDGMENTS	xxxix
CONTRIBUTORS	xxxiii
1 POWER ALLOCATION AND TASK SCHEDULING ON MULTIPROCESSOR COMPUTERS WITH ENERGY AND TIME CONSTRAINTS	1
<i>Keqin Li</i>	
1.1 Introduction	1
1.1.1 Energy Consumption	1
1.1.2 Power Reduction	2
1.1.3 Dynamic Power Management	3
1.1.4 Task Scheduling with Energy and Time Constraints	4
1.1.5 Chapter Outline	5
1.2 Preliminaries	5
1.2.1 Power Consumption Model	5
1.2.2 Problem Definitions	6
1.2.3 Task Models	7
1.2.4 Processor Models	8
1.2.5 Scheduling Models	9
1.2.6 Problem Decomposition	9
	vii

1.2.7	Types of Algorithms	10
1.3	Problem Analysis	10
1.3.1	Schedule Length Minimization	10
1.3.1.1	<i>Uniprocessor computers</i>	10
1.3.1.2	<i>Multiprocessor computers</i>	11
1.3.2	Energy Consumption Minimization	12
1.3.2.1	<i>Uniprocessor computers</i>	12
1.3.2.2	<i>Multiprocessor computers</i>	13
1.3.3	Strong NP-Hardness	14
1.3.4	Lower Bounds	14
1.3.5	Energy-Delay Trade-off	15
1.4	Pre-Power-Determination Algorithms	16
1.4.1	Overview	16
1.4.2	Performance Measures	17
1.4.3	Equal-Time Algorithms and Analysis	18
1.4.3.1	<i>Schedule length minimization</i>	18
1.4.3.2	<i>Energy consumption minimization</i>	19
1.4.4	Equal-Energy Algorithms and Analysis	19
1.4.4.1	<i>Schedule length minimization</i>	19
1.4.4.2	<i>Energy consumption minimization</i>	21
1.4.5	Equal-Speed Algorithms and Analysis	22
1.4.5.1	<i>Schedule length minimization</i>	22
1.4.5.2	<i>Energy consumption minimization</i>	23
1.4.6	Numerical Data	24
1.4.7	Simulation Results	25
1.5	Post-Power-Determination Algorithms	28
1.5.1	Overview	28
1.5.2	Analysis of List Scheduling Algorithms	29
1.5.2.1	<i>Analysis of algorithm LS</i>	29
1.5.2.2	<i>Analysis of algorithm LRF</i>	30
1.5.3	Application to Schedule Length Minimization	30
1.5.4	Application to Energy Consumption Minimization	31
1.5.5	Numerical Data	32
1.5.6	Simulation Results	32
1.6	Summary and Further Research	33
	References	34

2 POWER-AWARE HIGH PERFORMANCE COMPUTING 39

Rong Ge and Kirk W. Cameron

2.1	Introduction	39
2.2	Background	41
2.2.1	Current Hardware Technology and Power Consumption	41
2.2.1.1	<i>Processor power</i>	41
2.2.1.2	<i>Memory subsystem power</i>	42
2.2.2	Performance	43
2.2.3	Energy Efficiency	44
2.3	Related Work	45
2.3.1	Power Profiling	45
2.3.1.1	<i>Simulator-based power estimation</i>	45
2.3.1.2	<i>Direct measurements</i>	46
2.3.1.3	<i>Event-based estimation</i>	46
2.3.2	Performance Scalability on Power-Aware Systems	46
2.3.3	Adaptive Power Allocation for Energy-Efficient Computing	47
2.4	PowerPack: Fine-Grain Energy Profiling of HPC Applications	48
2.4.1	Design and Implementation of PowerPack	48
2.4.1.1	<i>Overview</i>	48
2.4.1.2	<i>Fine-grain systematic power measurement</i>	50
2.4.1.3	<i>Automatic power profiling and code synchronization</i>	51
2.4.2	Power Profiles of HPC Applications and Systems	53
2.4.2.1	<i>Power distribution over components</i>	53
2.4.2.2	<i>Power dynamics of applications</i>	54
2.4.2.3	<i>Power bounds on HPC systems</i>	55
2.4.2.4	<i>Power versus dynamic voltage and frequency scaling</i>	57
2.5	Power-Aware Speedup Model	59
2.5.1	Power-Aware Speedup	59
2.5.1.1	<i>Sequential execution time for a single workload $T_1(w, f)$</i>	60

2.5.1.2	<i>Sequential execution time for an ON-chip/OFF-chip workload</i>	60
2.5.1.3	<i>Parallel execution time on N processors for an ON-/OFF-chip workload with $DOP = i$</i>	61
2.5.1.4	<i>Power-aware speedup for DOP and ON-/OFF-chip workloads</i>	62
2.5.2	Model Parametrization and Validation	63
2.5.2.1	<i>Coarse-grain parametrization and validation</i>	64
2.5.2.2	<i>Fine-grain parametrization and validation</i>	66
2.6	Model Usages	69
2.6.1	Identification of Optimal System Configurations	70
2.6.2	PAS-Directed Energy-Driven Runtime Frequency Scaling	71
2.7	Conclusion	73
	References	75

3 ENERGY EFFICIENCY IN HPC SYSTEMS **81**

Ivan Rodero and Manish Parashar

3.1	Introduction	81
3.2	Background and Related Work	83
3.2.1	CPU Power Management	83
3.2.1.1	<i>OS-level CPU power management</i>	83
3.2.1.2	<i>Workload-level CPU power management</i>	84
3.2.1.3	<i>Cluster-level CPU power management</i>	84
3.2.2	Component-Based Power Management	85
3.2.2.1	<i>Memory subsystem</i>	85
3.2.2.2	<i>Storage subsystem</i>	86
3.2.3	Thermal-Conscious Power Management	87
3.2.4	Power Management in Virtualized Datacenters	87
3.3	Proactive, Component-Based Power Management	88
3.3.1	Job Allocation Policies	88
3.3.2	Workload Profiling	90
3.4	Quantifying Energy Saving Possibilities	91
3.4.1	Methodology	92
3.4.2	Component-Level Power Requirements	92
3.4.3	Energy Savings	94
3.5	Evaluation of the Proposed Strategies	95
3.5.1	Methodology	96

3.5.2	Workloads	96
3.5.3	Metrics	97
3.6	Results	97
3.7	Concluding Remarks	102
3.8	Summary	103
	References	104
4	A STOCHASTIC FRAMEWORK FOR HIERARCHICAL SYSTEM-LEVEL POWER MANAGEMENT	109
	<i>Peng Rong and Massoud Pedram</i>	
4.1	Introduction	109
4.2	Related Work	111
4.3	A Hierarchical DPM Architecture	113
4.4	Modeling	114
4.4.1	Model of the Application Pool	114
4.4.2	Model of the Service Flow Control	118
4.4.3	Model of the Simulated Service Provider	119
4.4.4	Modeling Dependencies between SPs	120
4.5	Policy Optimization	122
4.5.1	Mathematical Formulation	122
4.5.2	Optimal Time-Out Policy for Local Power Manager	123
4.6	Experimental Results	125
4.7	Conclusion	130
	References	130
5	ENERGY-EFFICIENT RESERVATION INFRASTRUCTURE FOR GRIDS, CLOUDS, AND NETWORKS	133
	<i>Anne-Cécile Orgerie and Laurent Lefèvre</i>	
5.1	Introduction	133
5.2	Related Works	134
5.2.1	Server and Data Center Power Management	135
5.2.2	Node Optimizations	135
5.2.3	Virtualization to Improve Energy Efficiency	136
5.2.4	Energy Awareness in Wired Networking Equipment	136
5.2.5	Synthesis	137
5.3	ERIDIS: Energy-Efficient Reservation Infrastructure for Large-Scale Distributed Systems	138
5.3.1	ERIDIS Architecture	138

5.3.2	Management of the Resource Reservations	141
5.3.3	Resource Management and On/Off Algorithms	145
5.3.4	Energy-Consumption Estimates	146
5.3.5	Prediction Algorithms	146
5.4	EARI: Energy-Aware Reservation Infrastructure for Data Centers and Grids	147
5.4.1	EARI's Architecture	147
5.4.2	Validation of EARI on Experimental Grid Traces	147
5.5	GOC: Green Open Cloud	149
5.5.1	GOC's Resource Manager Architecture	150
5.5.2	Validation of the GOC Framework	152
5.6	HERMES: High Level Energy-Aware Model for Bandwidth Reservation in End-To-End Networks	152
5.6.1	HERMES' Architecture	154
5.6.2	The Reservation Process of HERMES	155
5.6.3	Discussion	157
5.7	Summary	158
	References	158

6 ENERGY-EFFICIENT JOB PLACEMENT ON CLUSTERS, GRIDS, AND CLOUDS **163**

Damien Borgetto, Henri Casanova, Georges Da Costa, and Jean-Marc Pierson

6.1	Problem and Motivation	163
6.1.1	Context	163
6.1.2	Chapter Roadmap	164
6.2	Energy-Aware Infrastructures	164
6.2.1	Buildings	165
6.2.2	Context-Aware Buildings	165
6.2.3	Cooling	166
6.3	Current Resource Management Practices	167
6.3.1	Widely Used Resource Management Systems	167
6.3.2	Job Requirement Description	169
6.4	Scientific and Technical Challenges	170
6.4.1	Theoretical Difficulties	170
6.4.2	Technical Difficulties	170
6.4.3	Controlling and Tuning Jobs	171
6.5	Energy-Aware Job Placement Algorithms	172

6.5.1	State of the Art	172
6.5.2	Detailing One Approach	174
6.6	Discussion	180
6.6.1	Open Issues and Opportunities	180
6.6.2	Obstacles for Adoption in Production	182
6.7	Conclusion	183
	References	184
7	COMPARISON AND ANALYSIS OF GREEDY ENERGY-EFFICIENT SCHEDULING ALGORITHMS FOR COMPUTATIONAL GRIDS	189
	<i>Peder Lindberg, James Leingang, Daniel Lysaker, Kashif Bilal, Samee Ullah Khan, Pascal Bouvry, Nasir Ghani, Nasro Min-Allah, and Juan Li</i>	
7.1	Introduction	189
7.2	Problem Formulation	191
7.2.1	The System Model	191
7.2.1.1	PEs	191
7.2.1.2	DVS	191
7.2.1.3	Tasks	192
7.2.1.4	Preliminaries	192
7.2.2	Formulating the Energy-Makespan Minimization Problem	192
7.3	Proposed Algorithms	193
7.3.1	Greedy Heuristics	194
7.3.1.1	<i>Greedy heuristic scheduling algorithm</i>	196
7.3.1.2	<i>Greedy-min</i>	197
7.3.1.3	<i>Greedy-deadline</i>	198
7.3.1.4	<i>Greedy-max</i>	198
7.3.1.5	<i>MaxMin</i>	199
7.3.1.6	<i>ObFun</i>	199
7.3.1.7	<i>MinMin StdDev</i>	202
7.3.1.8	<i>MinMax StdDev</i>	202
7.4	Simulations, Results, and Discussion	203
7.4.1	Workload	203
7.4.2	Comparative Results	204
7.4.2.1	<i>Small-size problems</i>	204
7.4.2.2	<i>Large-size problems</i>	206
7.5	Related Works	211

7.6 Conclusion	211
References	212

8 TOWARD ENERGY-AWARE SCHEDULING USING MACHINE LEARNING 215

Josep LL. Berral, Iñigo Goiri, Ramon Nou, Ferran Julià, Josep O. Fitó, Jordi Guitart, Ricard Gavalda, and Jordi Torres

8.1 Introduction	215
8.1.1 Energetic Impact of the Cloud	216
8.1.2 An Intelligent Way to Manage Data Centers	216
8.1.3 Current Autonomic Computing Techniques	217
8.1.4 Power-Aware Autonomic Computing	217
8.1.5 State of the Art and Case Study	218
8.2 Intelligent Self-Management	218
8.2.1 Classical AI Approaches	219
8.2.1.1 <i>Heuristic algorithms</i>	219
8.2.1.2 <i>AI planning</i>	219
8.2.1.3 <i>Semantic techniques</i>	219
8.2.1.4 <i>Expert systems and genetic algorithms</i>	220
8.2.2 Machine Learning Approaches	220
8.2.2.1 <i>Instance-based learning</i>	221
8.2.2.2 <i>Reinforcement learning</i>	222
8.2.2.3 <i>Feature and example selection</i>	225
8.3 Introducing Power-Aware Approaches	225
8.3.1 Use of Virtualization	226
8.3.2 Turning On and Off Machines	228
8.3.3 Dynamic Voltage and Frequency Scaling	229
8.3.4 Hybrid Nodes and Data Centers	230
8.4 Experiences of Applying ML on Power-Aware Self-Management	230
8.4.1 Case Study Approach	231
8.4.2 Scheduling and Power Trade-Off	231
8.4.3 Experimenting with Power-Aware Techniques	233
8.4.4 Applying Machine Learning	236
8.4.5 Conclusions from the Experiments	238
8.5 Conclusions on Intelligent Power-Aware Self-Management	238
References	240

9 ENERGY EFFICIENCY METRICS FOR DATA CENTERS 245

Javid Taheri and Albert Y. Zomaya

9.1	Introduction	245
9.1.1	Background	245
9.1.2	Data Center Energy Use	246
9.1.3	Data Center Characteristics	246
	9.1.3.1 <i>Electric power</i>	247
	9.1.3.2 <i>Heat removal</i>	249
9.1.4	Energy Efficiency	250
9.2	Fundamentals of Metrics	250
9.2.1	Demand and Constraints on Data Center Operators	250
9.2.2	Metrics	251
	9.2.2.1 <i>Criteria for good metrics</i>	251
	9.2.2.2 <i>Methodology</i>	252
	9.2.2.3 <i>Stability of metrics</i>	252
9.3	Data Center Energy Efficiency	252
9.3.1	Holistic IT Efficiency Metrics	252
	9.3.1.1 <i>Fixed versus proportional overheads</i>	254
	9.3.1.2 <i>Power versus energy</i>	254
	9.3.1.3 <i>Performance versus productivity</i>	255
9.3.2	Code of Conduct	256
	9.3.2.1 <i>Environmental statement</i>	256
	9.3.2.2 <i>Problem statement</i>	256
	9.3.2.3 <i>Scope of the CoC</i>	257
	9.3.2.4 <i>Aims and objectives of CoC</i>	258
9.3.3	Power Use in Data Centers	259
	9.3.3.1 <i>Data center IT power to utility power relationship</i>	259
	9.3.3.2 <i>Chiller efficiency and external temperature</i>	260
9.4	Available Metrics	260
9.4.1	The Green Grid	261
	9.4.1.1 <i>Power usage effectiveness (PUE)</i>	261
	9.4.1.2 <i>Data center efficiency (DCE)</i>	262
	9.4.1.3 <i>Data center infrastructure efficiency (DCiE)</i>	262
	9.4.1.4 <i>Data center productivity (DCP)</i>	263

9.4.2	McKinsey	263
9.4.3	Uptime Institute	264
9.4.3.1	<i>Site infrastructure power overhead multiplier (SI-POM)</i>	265
9.4.3.2	<i>IT hardware power overhead multiplier (H-POM)</i>	266
9.4.3.3	<i>DC hardware compute load per unit of computing work done</i>	266
9.4.3.4	<i>Deployed hardware utilization ratio (DH-UR)</i>	266
9.4.3.5	<i>Deployed hardware utilization efficiency (DH-UE)</i>	267
9.5	Harmonizing Global Metrics for Data Center Energy Efficiency	267
	References	268
10	AUTONOMIC GREEN COMPUTING IN LARGE-SCALE DATA CENTERS	271
	<i>Haoting Luo, Bithika Khargharia, Salim Hariri, and Youssif Al-Nashif</i>	
10.1	Introduction	271
10.2	Related Technologies and Techniques	272
10.2.1	Power Optimization Techniques in Data Centers	272
10.2.2	Design Model	273
10.2.3	Networks	274
10.2.4	Data Center Power Distribution	275
10.2.5	Data Center Power-Efficient Metrics	276
10.2.6	Modeling Prototype and Testbed	277
10.2.7	Green Computing	278
10.2.8	Energy Proportional Computing	280
10.2.9	Hardware Virtualization Technology	281
10.2.10	Autonomic Computing	282
10.3	Autonomic Green Computing: A Case Study	283
10.3.1	Autonomic Management Platform	285
10.3.1.1	<i>Platform architecture</i>	285
10.3.1.2	<i>DEVS-based modeling and simulation platform</i>	285
10.3.1.3	<i>Workload generator</i>	287
10.3.2	Model Parameter Evaluation	288

10.3.2.1	<i>State transitioning overhead</i>	288
10.3.2.2	<i>VM template evaluation</i>	289
10.3.2.3	<i>Scalability analysis</i>	291
10.3.3	Autonomic Power Efficiency Management Algorithm (Performance Per Watt)	291
10.3.4	Simulation Results and Evaluation	293
10.3.4.1	<i>Analysis of energy and performance trade-offs</i>	296
10.4	Conclusion and Future Directions	297
	References	298

11 ENERGY AND THERMAL AWARE SCHEDULING IN DATA CENTERS 301

Gaurav Dhiman, Raid Ayoub, and Tajana S. Rosing

11.1	Introduction	301
11.2	Related Work	302
11.3	Intermachine Scheduling	305
11.3.1	Performance and Power Profile of VMs	305
11.3.2	Architecture	309
11.3.2.1	<i>vgnode</i>	309
11.3.2.2	<i>vgxen</i>	310
11.3.2.3	<i>vgdom</i>	312
11.3.2.4	<i>vgserv</i>	312
11.4	Intramachine Scheduling	315
11.4.1	Air-Forced Thermal Modeling and Cost	316
11.4.2	Cooling Aware Dynamic Workload Scheduling	317
11.4.3	Scheduling Mechanism	318
11.4.4	Cooling Costs Predictor	319
11.5	Evaluation	321
11.5.1	Intermachine Scheduler (vGreen)	321
11.5.2	Heterogeneous Workloads	323
11.5.2.1	<i>Comparison with DVFS policies</i>	325
11.5.2.2	<i>Homogeneous workloads</i>	328
11.5.3	Intramachine Scheduler (Cool and Save)	328
11.5.3.1	<i>Results</i>	331
11.5.3.2	<i>Overhead of CAS</i>	333
11.6	Conclusion	333
	References	334

12	QOS-AWARE POWER MANAGEMENT IN DATA CENTERS	339
	<i>Jiayu Gong and Cheng-Zhong Xu</i>	
12.1	Introduction	339
12.2	Problem Classification	340
12.2.1	Objective and Constraint	340
12.2.2	Scope and Time Granularities	340
12.2.3	Methodology	341
12.2.4	Power Management Mechanism	342
12.3	Energy Efficiency	344
12.3.1	Energy-Efficiency Metrics	344
12.3.2	Improving Energy Efficiency	346
12.3.2.1	<i>Energy minimization with performance guarantee</i>	346
12.3.2.2	<i>Performance maximization under power budget</i>	348
12.3.2.3	<i>Trade-off between power and performance</i>	348
12.3.3	Energy-Proportional Computing	350
12.4	Power Capping	351
12.5	Conclusion	353
	References	356
13	ENERGY-EFFICIENT STORAGE SYSTEMS FOR DATA CENTERS	361
	<i>Sudhanva Gurumurthi and Anand Sivasubramaniam</i>	
13.1	Introduction	361
13.2	Disk Drive Operation and Disk Power	362
13.2.1	An Overview of Disk Drives	362
13.2.2	Sources of Disk Power Consumption	363
13.2.3	Disk Activity and Power Consumption	365
13.3	Disk and Storage Power Reduction Techniques	366
13.3.1	Exploiting the STANDBY State	368
13.3.2	Reducing Seek Activity	369
13.3.3	Achieving Energy Proportionality	369
13.3.3.1	<i>Hardware approaches</i>	369
13.3.3.2	<i>Software approaches</i>	370
13.4	Using Nonvolatile Memory and Solid-State Disks	371
13.5	Conclusions	372
	References	373

14	AUTONOMIC ENERGY/PERFORMANCE OPTIMIZATIONS FOR MEMORY IN SERVERS	377
	<i>Bithika Khargharia and Mazin Yousif</i>	
14.1	Introduction	378
14.2	Classifications of Dynamic Power Management Techniques	380
14.2.1	Heuristic and Predictive Techniques	380
14.2.2	QoS and Energy Trade-Offs	381
14.3	Applications of Dynamic Power Management (DPM)	382
14.3.1	Power Management of System Components in Isolation	382
14.3.2	Joint Power Management of System Components	383
14.3.3	Holistic System-Level Power Management	383
14.4	Autonomic Power and Performance Optimization of Memory Subsystems in Server Platforms	384
14.4.1	Adaptive Memory Interleaving Technique for Power and Performance Management	384
14.4.1.1	<i>Formulating the optimization problem</i>	386
14.4.1.2	<i>Memory appflow</i>	389
14.4.2	Industry Techniques	389
14.4.2.1	<i>Enhancements in memory hardware design</i>	390
14.4.2.2	<i>Adding more operating states</i>	390
14.4.2.3	<i>Faster transition to and from low power states</i>	390
14.4.2.4	<i>Memory consolidation</i>	390
14.5	Conclusion	391
	References	391
15	ROD: A PRACTICAL APPROACH TO IMPROVING RELIABILITY OF ENERGY-EFFICIENT PARALLEL DISK SYSTEMS	395
	<i>Shu Yin, Xiaojun Ruan, Adam Manzanares, and Xiao Qin</i>	
15.1	Introduction	395
15.2	Modeling Reliability of Energy-Efficient Parallel Disks	396
15.2.1	The MINT Model	396
15.2.1.1	<i>Disk utilization</i>	398
15.2.1.2	<i>Temperature</i>	398
15.2.1.3	<i>Power-state transition frequency</i>	399
15.2.1.4	<i>Single disk reliability model</i>	399
15.2.2	MAID, Massive Arrays of Idle Disks	400
15.3	Improving Reliability of MAID via Disk Swapping	401

15.3.1	Improving Reliability of Cache Disks in MAID	401
15.3.2	Swapping Disks Multiple Times	404
15.4	Experimental Results and Evaluation	405
15.4.1	Experimental Setup	405
15.4.2	Disk Utilization	406
15.4.3	The Single Disk Swapping Strategy	406
15.4.4	The Multiple Disk Swapping Strategy	409
15.5	Related Work	411
15.6	Conclusions	412
	References	413
16	EMBRACING THE MEMORY AND I/O WALLS FOR ENERGY-EFFICIENT SCIENTIFIC COMPUTING	417
	<i>Chung-Hsing Hsu and Wu-Chun Feng</i>	
16.1	Introduction	417
16.2	Background and Related Work	420
16.2.1	DVFS-Enabled Processors	420
16.2.2	DVFS Scheduling Algorithms	421
16.2.3	Memory-Aware, Interval-Based Algorithms	422
16.3	β -Adaptation: A New DVFS Algorithm	423
16.3.1	The Compute-Boundedness Metric, β	423
16.3.2	The Frequency Calculating Formula, f^*	424
16.3.3	The Online β Estimation	425
16.3.4	Putting It All Together	427
16.4	Algorithm Effectiveness	429
16.4.1	A Comparison to Other DVFS Algorithms	429
16.4.2	Frequency Emulation	432
16.4.3	The Minimum Dependence to the PMU	436
16.5	Conclusions and Future Work	438
	References	439
17	MULTIPLE FREQUENCY SELECTION IN DVFS-ENABLED PROCESSORS TO MINIMIZE ENERGY CONSUMPTION	443
	<i>Nikzad Babaii Rizvandi, Albert Y. Zomaya, Young Choon Lee, Ali Javadzadeh Boloori, and Javid Taheri</i>	
17.1	Introduction	443
17.2	Energy Efficiency in HPC Systems	444
17.3	Exploitation of Dynamic Voltage–Frequency Scaling	446
17.3.1	Independent Slack Reclamation	446

17.3.2	Integrated Schedule Generation	447
17.4	Preliminaries	448
17.4.1	System and Application Models	448
17.4.2	Energy Model	448
17.5	Energy-Aware Scheduling via DVFS	450
17.5.1	Optimum Continuous Frequency	450
17.5.2	Reference Dynamic Voltage–Frequency Scaling (RDVFS)	451
17.5.3	Maximum-Minimum-Frequency for Dynamic Voltage–Frequency Scaling (MMF-DVFS)	452
17.5.4	Multiple Frequency Selection for Dynamic Voltage–Frequency Scaling (MFS-DVFS)	453
17.5.4.1	<i>Task eligibility</i>	454
17.6	Experimental Results	456
17.6.1	Simulation Settings	456
17.6.2	Results	458
17.7	Conclusion	461
	References	461
18	THE PARAMOUNTCY OF RECONFIGURABLE COMPUTING	465
	<i>Reiner Hartenstein</i>	
18.1	Introduction	465
18.2	Why Computers are Important	466
18.2.1	Computing for a Sustainable Environment	470
18.3	Performance Progress Stalled	472
18.3.1	Unaffordable Energy Consumption of Computing	473
18.3.2	Crashing into the Programming Wall	475
18.4	The Tail is Wagging the Dog (Accelerators)	488
18.4.1	Hardwired Accelerators	489
18.4.2	Programmable Accelerators	490
18.5	Reconfigurable Computing	494
18.5.1	Speedup Factors by FPGAs	498
18.5.2	The Reconfigurable Computing Paradox	501
18.5.3	Saving Energy by Reconfigurable Computing	505
18.5.3.1	<i>Traditional green computing</i>	506
18.5.3.2	<i>The role of graphics processors</i>	507
18.5.3.3	<i>Wintel versus ARM</i>	508
18.5.4	Reconfigurable Computing is the Silver Bullet	511

18.5.4.1	<i>A new world model of computing</i>	511
18.5.5	The Twin-Paradigm Approach to Tear Down the Wall	514
18.5.6	A Mass Movement Needed as Soon as Possible	517
18.5.6.1	<i>Legacy software from the mainframe age</i>	518
18.5.7	How to Reinvent Computing	519
18.6	Conclusions	526
	References	529
19	WORKLOAD CLUSTERING FOR INCREASING ENERGY SAVINGS ON EMBEDDED MPSOCS	549
	<i>Ozcan Ozturk, Mahmut Kandemir, and Sri Hari Krishna Narayanan</i>	
19.1	Introduction	549
19.2	Embedded MPSoC Architecture, Execution Model, and Related Work	550
19.3	Our Approach	551
19.3.1	Overview	551
19.3.2	Technical Details and Problem Formulation	553
19.3.2.1	<i>System and job model</i>	553
19.3.2.2	<i>Mathematical programing model</i>	554
19.3.2.3	<i>Example</i>	557
19.4	Experimental Evaluation	560
19.5	Conclusions	564
	References	565
20	ENERGY-EFFICIENT INTERNET INFRASTRUCTURE	567
	<i>Weirong Jiang and Viktor K. Prasanna</i>	
20.1	Introduction	567
20.1.1	Performance Challenges	568
20.1.2	Existing Packet Forwarding Approaches	570
20.1.2.1	<i>Software approaches</i>	570
20.1.2.2	<i>Hardware approaches</i>	571
20.2	SRAM-Based Pipelined IP Lookup Architectures: Alternative to TCAMs	571
20.3	Data Structure Optimization for Power Efficiency	573
20.3.1	Problem Formulation	574

20.3.1.1	<i>Non-pipelined and pipelined engines</i>	574
20.3.1.2	<i>Power function of SRAM</i>	575
20.3.2	Special Case: Uniform Stride	576
20.3.3	Dynamic Programming	576
20.3.4	Performance Evaluation	577
20.3.4.1	<i>Results for non-pipelined architecture</i>	578
20.3.4.2	<i>Results for pipelined architecture</i>	578
20.4	Architectural Optimization to Reduce Dynamic Power Dissipation	580
20.4.1	Analysis and Motivation	581
20.4.1.1	<i>Traffic locality</i>	582
20.4.1.2	<i>Traffic rate variation</i>	582
20.4.1.3	<i>Access frequency on different stages</i>	583
20.4.2	Architecture-Specific Techniques	583
20.4.2.1	<i>Inherent caching</i>	584
20.4.2.2	<i>Local clocking</i>	584
20.4.2.3	<i>Fine-grained memory enabling</i>	585
20.4.3	Performance Evaluation	585
20.5	Related Work	588
20.6	Summary	589
	References	589

21 DEMAND RESPONSE IN THE SMART GRID: A DISTRIBUTED COMPUTING PERSPECTIVE **593**

Chen Wang and Martin De Groot

21.1	Introduction	593
21.2	Demand Response	595
21.2.1	Existing Demand Response Programs	595
21.2.2	Demand Response Supported by the Smart Grid	597
21.3	Demand Response as a Distributed System	600
21.3.1	An Overlay Network for Demand Response	600
21.3.2	Event Driven Demand Response	602
21.3.3	Cost Driven Demand Response	604
21.3.4	A Decentralized Demand Response Framework	609
21.3.5	Accountability of Coordination Decision Making	610
21.4	Summary	611
	References	611

22 RESOURCE MANAGEMENT FOR DISTRIBUTED MOBILE COMPUTING	615
<i>Jong-Kook Kim</i>	
22.1 Introduction	615
22.2 Single-Hop Energy-Constrained Environment	617
22.2.1 System Model	617
22.2.2 Related Work	620
22.2.3 Heuristic Descriptions	621
22.2.3.1 <i>Mapping event</i>	621
22.2.3.2 <i>Scheduling communications</i>	621
22.2.3.3 <i>Opportunistic load balancing and minimum energy greedy heuristics</i>	622
22.2.3.4 <i>ME-MC heuristic</i>	622
22.2.3.5 <i>ME-ME heuristic</i>	624
22.2.3.6 <i>CRME heuristic</i>	625
22.2.3.7 <i>Originator and random</i>	626
22.2.3.8 <i>Upper bound</i>	626
22.2.4 Simulation Model	628
22.2.5 Results	630
22.2.6 Summary	634
22.3 Multihop Distributed Mobile Computing Environment	635
22.3.1 The Multihop System Model	635
22.3.2 Energy-Aware Routing Protocol	636
22.3.2.1 <i>Overview</i>	636
22.3.2.2 <i>DSDV</i>	637
22.3.2.3 <i>DSDV remaining energy</i>	637
22.3.2.4 <i>DSDV-energy consumption per remaining energy</i>	637
22.3.3 Heuristic Description	638
22.3.3.1 <i>Random</i>	638
22.3.3.2 <i>Estimated minimum total energy (EMTE)</i>	638
22.3.3.3 <i>K-percent-speed (KPS) and K-percent-energy (KPE)</i>	639
22.3.3.4 <i>Energy ratio and distance (ERD)</i>	639
22.3.3.5 <i>ETC and distance (ETCD)</i>	640
22.3.3.6 <i>Minimum execution time (MET)</i>	640

22.3.3.7	<i>Minimum completion time (MCT) and minimum completion time with DVS (MCT-DVS)</i>	640
22.3.3.8	<i>Switching algorithm (SA)</i>	640
22.3.4	Simulation Model	641
22.3.5	Results	643
22.3.5.1	<i>Distributed resource management</i>	643
22.3.5.2	<i>Energy-aware protocol</i>	644
22.3.6	Summary	644
22.4	Future Work	647
	References	647
23	AN ENERGY-AWARE FRAMEWORK FOR MOBILE DATA MINING	653
	<i>Carmela Comito, Domenico Talia, and Paolo Trunfio</i>	
23.1	Introduction	653
23.2	System Architecture	654
23.3	Mobile Device Components	657
23.4	Energy Model	659
23.5	Clustering Scheme	664
23.5.1	Clustering the M2M Architecture	666
23.6	Conclusion	670
	References	670
24	ENERGY AWARENESS AND EFFICIENCY IN WIRELESS SENSOR NETWORKS: FROM PHYSICAL DEVICES TO THE COMMUNICATION LINK	673
	<i>Flávia C. Delicato and Paulo F. Pires</i>	
24.1	Introduction	673
24.2	WSN and Power Dissipation Models	676
24.2.1	Network and Node Architecture	676
24.2.2	Sources of Power Dissipation in WSNs	679
24.3	Strategies for Energy Optimization	683
24.3.1	Intranode Level	684
24.3.1.1	<i>Duty cycling</i>	685
24.3.1.2	<i>Adaptive sensing</i>	691
24.3.1.3	<i>Dynamic voltage scale (DVS)</i>	693
24.3.1.4	<i>OS task scheduling</i>	694

24.3.2	Internode Level	695
24.3.2.1	<i>Transmission power control</i>	695
24.3.2.2	<i>Dynamic modulation scaling</i>	696
24.3.2.3	<i>Link layer optimizations</i>	698
24.4	Final Remarks	701
	References	702
25	NETWORK-WIDE STRATEGIES FOR ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORKS	709
	<i>Flávia C. Delicato and Paulo F. Pires</i>	
25.1	Introduction	709
25.2	Data Link Layer	711
25.2.1	Topology Control Protocols	712
25.2.2	Energy-Efficient MAC Protocols	714
25.2.2.1	<i>Scheduled MAC protocols in WSNs</i>	716
25.2.2.2	<i>Contention-based MAC protocols</i>	717
25.3	Network Layer	719
25.3.1	Flat and Hierarchical Protocols	722
25.4	Transport Layer	725
25.5	Application Layer	729
25.5.1	Task Scheduling	729
25.5.2	Data Aggregation and Data Fusion in WSNs	733
25.5.2.1	<i>Approaches of data fusion for energy efficiency</i>	735
25.5.2.2	<i>Data aggregation strategies</i>	736
25.6	Final Remarks	740
	References	741
26	ENERGY MANAGEMENT IN HETEROGENEOUS WIRELESS HEALTH CARE NETWORKS	751
	<i>Nima Nikzad, Priti Aghera, Piero Zappi, and Tajana S. Rosing</i>	
26.1	Introduction	751
26.2	System Model	753
26.2.1	Health Monitoring Task Model	753
26.3	Collaborative Distributed Environmental Sensing	755
26.3.1	Node Neighborhood and Localization Rate	757
26.3.2	Energy Ratio and Sensing Rate	758
26.3.3	Duty Cycling and Prediction	759
26.4	Task Assignment in a Body Area Network	760

26.4.1	Optimal Task Assignment	760
26.4.2	Dynamic Task Assignment	762
26.4.2.1	<i>DynAGreen algorithm</i>	763
26.4.2.2	<i>DynAGreenLife algorithm</i>	768
26.5	Results	771
26.5.1	Collaborative Sensing	771
26.5.1.1	<i>Results</i>	772
26.5.2	Dynamic Task Assignment	776
26.5.2.1	<i>Performance in static conditions</i>	777
26.5.2.2	<i>Dynamic adaptability</i>	780
26.6	Conclusion	784
	References	785

INDEX	787
-------	-----

PREFACE

The scope of energy-efficient computing is not limited to main computing components (e.g., processors, storage devices, and visualization facilities), but it can expand to a much larger range of resources associated with computing facilities, including auxiliary equipment, water used for cooling, and even physical and floor space that these resources occupy. Energy consumption in computing facilities raises various monetary, environmental, and system performance concerns.

Recent advances in hardware technologies have improved the energy consumption issue to a certain degree. However, it still remains a serious concern for energy-efficient computing because the amount of energy consumed by computing and auxiliary hardware resources is affected substantially by their usage patterns. In other words, resource underutilization or overloading incurs a higher volume of energy consumption when compared with efficiently utilized resources. This calls for the development of various software energy-saving techniques and new algorithms that are more energy efficient.

This book, *Energy-Efficient Distributed Computing Systems*, seeks to provide an opportunity for researchers to explore different energy consumption issues and their impact on the design of new computing systems. The book is quite timely since the field of distributed computing as a whole is undergoing many changes. Vast literature exists today on such energy consumption paradigms and frameworks and their implications for a wide range of distributed platforms.

The book is intended to be a virtual roundtable of several outstanding researchers, which one might invite to attend a conference on energy-efficient computing systems. Of course, the list of topics that is explored here is by no means exhaustive, but most of the conclusions provided here should be extended to other computing platforms that are not covered here. There was a decision to limit the number of chapters while providing more pages for contributing

authors to express their ideas, so that the book remains manageable within a single volume.

We also hope that the topics covered in this book will get the readers to think of the implications of such new ideas on the developments in their own fields. The book endeavors to strike a balance between theoretical and practical coverage of innovative problem-solving techniques for a range of distributed platforms. The book is intended to be a repository of paradigms, technologies, and applications that target the different facets of energy consumption in computing systems.

The 26 chapters were carefully selected to provide a wide scope with minimal overlap between the chapters to reduce duplications. Each contributor was asked that his/her chapter should cover review material as well as current developments. In addition, the choice of authors was made so as to select authors who are leaders in their respective disciplines.

ALBERT Y. ZOMAYA
YOUNG CHOON LEE

ACKNOWLEDGMENTS

First and foremost, we would like to thank and acknowledge the contributors to this volume for their support and patience, and the reviewers for their useful comments and suggestions that helped in improving the earlier outline of the book and presentation of the material. Also, I should extend my deepest thanks to Simone Taylor and Diana Gialo from Wiley (USA) for their collaboration, guidance, and most importantly, patience in finalizing this handbook. Finally, I would like to acknowledge the efforts of the team from Wiley's production department for their extensive efforts during the many phases of this project and the timely manner in which the book was produced.

ALBERT Y. ZOMAYA
YOUNG CHOON LEE

CONTRIBUTORS

PRITI, AGHERA, University of California, San Diego, CA, USA.

AL-NASHIF, YOUSSEF, NSF Center for Autonomic Computing, The University of Arizona, USA.

AYOUB, RAID, University of California, San Diego, CA, USA.

BERRAL, JOSEP LL., Computer Architecture Dept. and Department of Software, UPC-Barcelona Tech., Catalonia, Spain.

BILAL, KASHIF, Department of Computer Science, North Dakota State University, Fargo, ND, USA.

BOLOORI, ALI JAVADZADEH, Centre for Distributed and High Performance Computing, School of Information Technologies, University of Sydney, NSW, Australia.

BORGETTO, DAMIEN, University Paul Sabatier, Toulouse, France.

BOUVRY, PASCAL, Faculty of Sciences, Technology, and Communications, University of Luxembourg, Luxembourg.

CAMERON, KIRK W., Virginia Tech, VA, USA.

CASANOVA, HENRI, University of Hawai'i at Manoa, Hawai'i, USA.

COMITO, CARMELA, DEIS, University of Calabria, Rende (CS), Italy.

DA COSTA, GEORGES, University Paul Sabatier, Toulouse, France.

DELICATO, FLAVIA C., Computer Science Department, Federal University of Rio de Janeiro—RN, Brazil.

DHIMAN, GAURAV, University of California, San Diego, CA, USA.

FENG, WU-CHUN, Virginia Tech, Blacksburg, Virginia, USA.

JOSEPH. O. FITO, Computer Architecture Dept. and Barcelona Supercomputing Center, UPC-Barcelona Tech., Catalonia, Spain.

GAVALDA, RICARD, Department of Software, UPC-Barcelona Tech., Catalonia, Spain.

GE, RONG, The Department of Mathematics, Statistics, and Computer Science, Marquette University, WI, USA.

GHANI, NASIR, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA.

GOIRI, INIGO, Computer Architecture Dept. and Barcelona Supercomputing Center, UPC-Barcelona Tech., Catalonia, Spain.

GONG, JIAYU, Department of Electrical and Computer Engineering, Wayne State University, MI, USA.

DE GROOT, MARTIN, CSIRO ICT Center, Epping, NSW, Australia.

GUITART, JORDI, Computer Architecture Dept. and Barcelona Supercomputing Center, UPC-Barcelona Tech., Catalonia, Spain.

GURUMURTHI, SUDHANVA, Dept. of Computer Science, University of Virginia, Charlottesville, VA, USA.

HARIRI, SALIM, NSF Center for Autonomic Computing, The University of Arizona, USA.

HARTENSTEIN, REINER, Department of Computer Science, Kaiserslautern University of Technology, Kaiserslautern, Germany.

HSU, CHUNG-HSING, Oak Ridge National Laboratory, Oak Ridge, TN, USA.

JIANG, WEIRONG, Juniper Networks, Inc., Sunnyvale, CA, USA.

JULIA, FERRAN, Computer Architecture Dept., UPC-Barcelona Tech., Catalonia, Spain.

KANDEMIR, MAHMUT, Pennsylvania State University, PA, USA.

KHAN, SAMEE ULLAH, Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA.

KHARGHARIA, BITHIKA, Cisco Systems, Inc., Durham, NC, USA.

KIM, JONG-KOOK, School of Electrical Engineering, Korea University, Korea.

LEE, YOUNG CHOON, Centre for Distributed and High Performance Computing, School of Information Technologies, University of Sydney, NSW, Australia.

LEFEVRE, LAURENT, INRIA, Ecole Normale Supérieure de Lyon, University of Lyon, France.

LEINGANG, JAMES, Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA.

LI, JUAN, Department of Computer Science, North Dakota State University, Fargo, ND, USA.

LI, KEQIN, State University of New York, New Paltz, NY, USA.

LINDBERG, PEDER, Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA.

LUO, HAOTING, NSF Center for Autonomic Computing, The University of Arizona, AZ, USA.

LYSAKER, DANIEL, Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA.

MANZANARES, ADAM, Los Alamos National Laboratory, Los Alamos, NM, USA.

MIN-ALLAH, NASRO, Department of Computer Science, COMSATS Institute of Information Technology, Pakistan.

NARAYANAN, SRI HARI KRISHNA, Argonne National Laboratory, IL, USA.

NIKZAD, NIMA, University of California, San Diego, CA, USA.

NOU, RAMON, Computer Architecture Dept. and Barcelona Supercomputing Center, UPC-Barcelona Tech., Catalonia, Spain.

ORGERIE, ANNE-CECILE, Ecole Normale Supérieure de Lyon, Lyon, France.

OZTURK, OZCAN, Bilkent University, Turkey.

PARASHAR, MANISH, NSF Cloud and Autonomic Computing Center and Rutgers Discovery Informatics Institute, Rutgers University, NJ, USA.

PEDRAM, MASSOUD, University of Southern California, Los Angeles, CA, USA.

PIERSON, JEAN-MARC, University Paul Sabatier, Toulouse, France.

PIRES, PAULO F., Computer Science Department, Federal University of Rio de Janeiro - RN, Brazil.

PRASANNA, VIKTOR K., University of Southern California, Los Angeles, CA, USA.

QIN, XIAO, Auburn University, Auburn, AL, USA.

RIZVANDI, NIKZAD BABAIL, Centre for Distributed and High Performance Computing, School of Information Technologies, University of Sydney, NSW, Australia.

RODERO, IVAN, NSF Cloud and Autonomic Computing Center and Rutgers Discovery Informatics Institute, Rutgers University, NJ, USA.

RONG, PENG, Brocade Communications Systems, San Jose, CA, USA.

ROSING, TAJANA SIMUNIC, University of California, San Diego, CA, USA.

RUAN, XIAOJUN, Auburn University, Auburn, AL, USA.

SIVASUBRAMANIAM, ANAND, Dept. of Computer Science and Engineering, The Pennsylvania State University, PA, USA.

TAHERI, JAVID, Centre for Distributed and High Performance Computing, School of Information Technologies, University of Sydney, NSW, Australia.

TALIA, DOMENICO, DEIS, University of Calabria, Rende (CS), Italy.

TORRES, JORDI, Computer Architecture Dept. and Barcelona Supercomputing Center, UPC-Barcelona Tech., Catalonia, Spain.

TRUNFIO, PAOLO, DEIS, University of Calabria, Rende (CS), Italy.

WANG, CHEN, CSIRO ICT Center, Epping, NSW, Australia.

XU, CHENG-ZHONG, Department of Electrical and Computer Engineering, Wayne State University, MI, USA.

YIN, SHU, Auburn University, Auburn, AL, USA.

YOUSIF, MAZIN, T-Systems International, Inc., Portland, OR, USA.

ZAPPI, PIERO, University of California, San Diego, CA, USA.

ZOMAYA, ALBERT Y., Centre for Distributed and High Performance Computing, School of Information Technologies, University of Sydney, NSW, Australia.