Contents

List of Figures **ix**

List of Tables **xi**

Abbreviations and Acronyms **xiii**

Abstract **xv**

Resumen **xvii**

Resum **xx**

1 Introduction **1**
1.1 Problem Description . . . 2
1.2 Dealing with Scalability in Future CMPs . . . 4
1.3 Objectives of the Thesis . . . 5
1.4 Contributions of the Thesis . . . 6
1.5 Thesis Outline . . . 7

2 Background and Related Work **9**
2.1 Background . . . 10
2.1.1 Cache Hierarchy . . . 10
2.1.2 Non Uniform Cache Access (NUCA) . . . 11
2.1.3 Coherence Protocols . . . 12
2.1.3.1 MOESI Protocol . . . 12
2.1.3.2 Update and Invalidation Protocols . . . 14
2.1.3.3 Directory-based and Snoopy Protocols . . . 15
2.1.3.4 Type of Misses . . . 17
2.2 Baseline Architecture . . . 18
2.3 Related Work . . . 18
2.3.1 Directory Caches . . . 19
2.3.1.1 Duplicate-tag and directories . . . 19
2.3.1.2 Sparse directories . . . 20
2.3.2 Processor Caches . . . 21
2.3.2.1 Energy-efficient cache designs . . . 22
2.3.2.2 Private-shared optimizations . . . 23

3 Experimental Framework **25**
3.1 Simulation tools . . . 26

3.1.1 Simics-GEMS . . . 26
3.1.2 CACTI . . . 27
3.2 Benchmarks . . . 27
3.2.1 Barnes . . . 28
3.2.2 Cholesky . . . 29
3.2.3 FFT . . . 29
3.2.4 FMM . . . 30
3.2.5 LU . . . 30
3.2.6 Ocean . . . 30
3.2.7 Radiosity . . . 31
3.2.8 Radix . . . 31
3.2.9 Raytrace . . . 32
3.2.10 Volrend . . . 32
3.2.11 Water-Nsq . . . 32
3.2.12 Blackscholes . . . 33
3.2.13 Swaptions . . . 33
3.2.14 FaceRec . . . 34
3.2.15 MPGdec . . . 34
3.2.16 MPGenc . . . 35
3.2.17 SpeechRec . . . 35
3.2.18 Tomcatv . . . 36
3.2.19 Unstructured . . . 36
3.2.20 Apache . . . 37
3.2.21 SPEC-JBB . . . 37
3.3 Metrics and Methodology . . . 37
3.4 Common System Parameters . . . 39

4 Directory Scalability 41
4.1 PS-Directory . . . 42
4.1.1 Analyzing the Behavior of Private and Shared Blocks from the Directory Point of View . . . 42
4.1.2 PS-Directory Organization and Basic Behavior . . . 45
4.1.3 Experimental Evaluation . . . 49
4.1.3.1 Impact of PS-Directory on Performance . . . 50
4.1.3.2 Impact of PS-Directory on Area and Energy . . . 54
4.1.3.3 Directory Coverage Ratio Analysis . . . 56
4.2 DWP-Directory . . . 59
4.2.1 Application Characterization . . . 60
4.2.2 DWP-Directory Architecture . . . 63
4.2.3 Basic DWP-Directory Working Behavior . . . 64
4.2.4 Repartitioning Approach . . . 65
4.2.5 Experimental Evaluation . . . 67
4.2.5.1 Way Adaptation Analysis . . . 68
4.2.5.2 Impact of DWP-Directory on Performance . . . 70
4.2.5.3 Impact of the DWP-Directory on Energy Consumption . 72
4.2.5.4 Impact on Area Requirements . . . 76
4.3 Summary . . . 76

5 Filtering Techniques **79**
5.1 Analyzing the Cache Hierarchy Access . . . 80
5.2 PS-Cache . . . 81
5.2.1 The PS Page Classification Mechanism . . . 82
5.2.2 The PS-Cache Architecture . . . 84
5.2.3 Experimental Evaluation . . . 86
5.2.3.1 Private-Shared Blocks Behavior Analysis . . . 87
5.2.3.2 Impact of PS-Cache on Energy Consumption . . . 90
5.3 Tag-Filter Architecture . . . 92
5.3.1 Last Tag Bits Distribution . . . 92
5.3.2 TF-Architecture Scheme . . . 93
5.3.3 Experimental Evaluation . . . 97
5.3.3.1 Compared Schemes . . . 97
5.3.3.2 TF Architecture in Processor Caches . . . 98
5.3.3.3 TF Architecture in Directory Caches . . . 101
5.4 Summary . . . 103

6 Conclusions **105**
6.1 Contributions . . . 106
6.2 Future Work . . . 108
6.3 Publications . . . 109

References **113**