Contents

Acknowledgments xi

Foreword 1 xiii
Pierre FICHEUX

Foreword 2 xv
Maryline CHETTO

Part 1. Introduction 1

Chapter 1. General Introduction 3

1.1. The outburst of digital data 3
1.2. Performance and power consumption of storage systems 4
1.3. Memory hierarchy and storage technology 6
1.4. Introduction to flash memory integration 8
1.5. Scope of the book 11
1.6. Target audience 11
1.7. Outline of the book 12
1.8. How to read this book 13

Chapter 2. Flash Memories: Structure and Constraints 15

2.1. General presentation of flash memory 16
2.1.1. The different types of flash memory 16
2.1.2. Operating physical principles 16

2.1.3. Simplified hierarchical architecture of a NAND flash memory chip 18
2.1.4. Operations on flash memory 21
2.2. Constraints and limitations 24
2.2.1. Erase-before-write constraint 24
2.2.2. Wear constraint 25
2.2.3. Reliability limitation 26
2.3. Flash memory constraint management systems: general concepts 26
2.3.1. Management of erase-before-write constraint 27
2.3.2. Wear leveling 31
2.3.3. Reliability management 32
2.3.4. Constraint management systems 32
2.4. Conclusion 33

Chapter 3. Evaluation of Performance and Power Consumption of Storage Systems 35

3.1. Benchmarking storage systems based on flash memory 36
3.1.1. Micro-benchmarks 37
3.1.2. Macro-benchmarks 39
3.1.3. I/O traces 41
3.2. Performance and power consumption metrics of storage systems 42
3.2.1. Performance metrics 42
3.2.2. Power consumption metrics 44
3.3. Performance and power consumption measurements for flash memory based storage systems 44
3.3.1. Performance exploration through measurements 45
3.3.2. Exploration of the power consumption of storage systems based on NAND flash memory through measurements 47
3.4. Evaluation of performance and power consumption through simulation 48
3.4.1. General concepts 48
3.4.2. FlashSim 50
3.4.3. NandFlashSim 52
3.4.4. Other simulators and general comparison 52
3.5. Conclusion 52

Part 2. Embedded Domain and File Systems for Flash Memory: Flash File Systems 55

Chapter 4. Flash File Systems 57

4.1. General presentation of FFSs 57
4.1.1. Storage management and flash memory constraints 58
4.1.2. Embedded constraints and scalability of FFSs 59
4.2. Integration of FFS storage systems in computer systems: the Linux example 60
4.2.1. Linux Virtual File System *VFS* 61
4.2.2. NAND driver *MTD* 62
4.3. Presentation of the most popular FFSs: JFFS2, YAFFS2 and UBIFS 63
4.3.1. JFFS2 64
4.3.2. YAFFS2 66
4.3.3. UBI and UBIFS 69
4.4. Other state-of-the-art FFSs 71
4.5. Conclusion 73

Chapter 5. Methodology for Performance and Power Consumption Exploration of Flash File Systems 75

5.1. General presentation of exploration methodology 76
5.1.1. Methods and tools 76
5.1.2. Hardware platform 77
5.2. A toolset for performance exploration of FFS-based systems on Linux 78
5.2.1. Flashmon 80
5.2.2. VFSMon and FuncMon 84
5.3. Exploration of power consumption: Open-PEOPLE platform . 87
5.4. Conclusion 91

Chapter 6. Performance and Power Consumption of Dedicated File Systems: Experimental Results 93

6.1. Hardware and driver levels 93
6.1.1. Performance and power consumption of basic flash operations 94
6.1.2. Impact of the MTD read buffer 96
6.1.3. Conclusion 99

6.2. FFS level – focus on JFFS2 99
6.2.1. Read at the FFS level 99
6.2.2. Write at the FFS level 108
6.2.3. Conclusion 116
6.3. VFS level 117
6.3.1. Page cache 117
6.3.2. Read-ahead mechanism 120
6.3.3. The write-back mechanism 122
6.3.4. Conclusion 123
6.4. Conclusion 123

Part 3 . Flash Translation Layers 127

Chapter 7. Flash Translation Layer 129

7.1. Introduction 129
7.2. Basic mapping schemes 130
7.2.1. Page-level mapping scheme 131
7.2.2. Block-level mapping scheme 131
7.2.3. Hybrid mapping scheme 132
7.3. Complex mapping schemes 133
7.3.1. Log block FTL 134
7.3.2. Page-level mapping FTL 138
7.3.3. FTL with partitioned flash memory 139
7.4. Wear leveling 140
7.4.1. Wear leveling based on the number of erase operations 141
7.4.2. Wear leveling based on the number of write cycles 142
7.5. Garbage collection algorithms 142
7.6. Cache mechanisms for flash memory 144
7.7. Conclusion 146

Chapter 8. Methodology for the Evaluation of SSD Performance and Power Consumption 149

8.1. Introduction 150
8.1.1. Method and tools 150
8.1.2. Hardware platform 151
8.2. I/O software stack in Linux 151
8.2.1. The generic block layer 153

8.2.2. The I/O scheduler 154
8.2.3. The device driver 156
8.3. Context: the Cloud 156
8.4. I/O monitoring tools for performance exploration 157
8.4.1. Level 1: the hypervisor 158
8.4.2. Level 2: host VFS 160
8.4.3. Level 3: file system 162
8.4.4. Level 4: block layer 162
8.5. Performance and energy consumption analysis 166
8.5.1. Measurement of the overall system (PDU) 166
8.5.2. Measurement of the storage system (sensors) 168
8.6. Conclusion 169

Chapter 9. Performance and Power Consumption of SSD Based Systems: Experimental Results 171

9.1. Introduction 171
9.2. Impact of I/Os on performance and energy consumption 172
9.3. A macroscopic view of performance and power consumption of storage systems 174
9.3.1. Hardware/software configuration for experiments 175
9.3.2. Measurement results 176
9.4. A microscopic view of performance and power consumption of storage systems 185
9.4.1. The use of micro-benchmarks 185
9.4.2. Video application case study 190
9.5. Conclusions 198

Part 4. Emerging Non-volatile Memories 201

Chapter 10. Emerging Non-volatile Memories 203

10.1. Introduction 203
10.2. NVM integration 204
10.2.1. Integration as a storage system 204
10.2.2. Integration as a main memory 206
10.2.3. Integration in CPU caches 206
10.3. PCM or phase-change memory 207
10.3.1. Basic concepts 207

10.3.2. PCM integration 208
10.3.3. PCM as a main memory 208
10.3.4. PCM as a storage system 209
10.3.5. Open questions 210
10.4. MRAM or magneto-resistive memory 211
10.4.1. Basic concepts 212
10.4.2. MRAM integration 213
10.4.3. Open questions 215
10.5. FeRAM or ferroelectric memory 215
10.5.1. Basic concepts 216
10.5.2. FeRAM integration 217
10.6. ReRAM or resistive memory 218
10.6.1. Basic concepts 219
10.6.2. ReRAM integration 221
10.6.3. Open questions 222
10.7. Conclusion 223

Conclusion 225

Bibliography 229

Index 249