

Contents

- 1 Introduction 1**
 - 1.1 Challenges in Heterogeneous Computing 3
 - 1.2 The Need for an Open-Source Research Platform . . . 6
 - 1.3 Challenges of an Open-Source Research Platform . . . 7
 - 1.4 Contributions and Publications 8
 - 1.5 Outline 12
- 2 Architecture and Components of a Heterogeneous Research Platform 15**
 - 2.1 Hardware Architecture 15
 - 2.1.1 PMCA ISA and Architecture 16
 - 2.1.2 Heterogeneous Architecture 17
 - 2.1.3 PMCA Configurability 18
 - 2.2 Software Architecture 19
 - 2.2.1 Heterogeneous Cross Compilation 21
 - 2.2.2 Heterogeneous OpenMP Runtime 22
 - 2.2.3 Compiler Support for Hybrid-MMU-Based SVM 22
 - 2.2.4 Host Runtime Library and Linux Driver 23
 - 2.2.5 PMCA Virtual Memory Management Library . 23
 - 2.3 Prototype Platform and Evaluation 24
 - 2.3.1 Carrier Platforms 24
 - 2.3.2 Case Study: Parallel Speedup Analysis 26
 - 2.3.3 Case Study: Virtual Memory Performance Analysis 27
 - 2.3.4 Case Study: Heterogeneous Speedup Analysis . 29
 - 2.4 Related Work 33
 - 2.5 Summary and Limitations 36

3	High-Performance Non-Coherent On-Chip Communication	39
3.1	Architecture	41
3.1.1	Elementary Components: Network (De)Muxes .	43
3.1.2	Network Junctions: Crossbars and Crosspoints	48
3.1.3	Concurrent Transactions: ID Width Converters	51
3.1.4	Data Width Converters	54
3.1.5	Clock Domain Crossing	57
3.1.6	Data Movement: DMA Engine	57
3.1.7	On-Chip Memory Controllers	59
3.1.8	Last Level Cache	61
3.2	Implementation Results	64
3.2.1	Elementary Components: Network (De)Muxes .	64
3.2.2	Network Junctions: Crossbars and Crosspoints	67
3.2.3	Concurrent Transactions: ID Width Converters	69
3.2.4	Data Width Converters	72
3.2.5	Clock Domain Crossing	74
3.2.6	Data Streaming: DMA Engine	74
3.2.7	On-Chip Memory Controllers	75
3.2.8	Last Level Cache	77
3.2.9	Complexity Overview and Summary	79
3.3	Full-System Case Study	79
3.3.1	Network Design	81
3.3.2	Network Microarchitecture and Implementation	82
3.3.3	Network Performance	85
3.4	Related Work	87
3.5	Summary	91
4	Modular and Scalable Support for Atomic Operations in a Shared Memory Multiprocessor	93
4.1	Background	95
4.1.1	RISC-V ISA	95
4.1.2	Modern On-Chip Communication and AXI . .	96
4.2	Design and Architecture	97
4.2.1	Design Overview	98
4.2.2	LR/SC Stage	98
4.2.3	AMO Stage	100
4.2.4	Liveness Guarantees	102

4.3	Evaluation	103
4.3.1	Evaluated Architecture	103
4.3.2	Terminology: Atomic Locality	104
4.3.3	Throughput and Contention	104
4.3.4	22nm FDSOI Hardware Complexity	109
4.4	Related Work	109
4.5	Summary	111
5	Scalable and Efficient Virtual Memory Sharing with TLB Prefetching and MMU-Aware DMA Engine	113
5.1	Related Work	115
5.2	Target Architecture Template	118
5.3	Implementation	121
5.3.1	Helper Thread Prefetching	121
5.3.2	Multi-Threaded TLB Miss Handling	125
5.3.3	Hybrid-IOMMU-Capable DMA Engine	126
5.4	Results	129
5.4.1	Evaluation Platform	129
5.4.2	Benchmark Description	130
5.4.3	Benchmark Results	131
5.4.4	Hardware Requirements of Hybrid-IOMMU-Capable DMA	136
5.5	Summary	136
6	Mixed-Data-Model Heterogeneous Compilation and OpenMP Offloading	137
6.1	Background	139
6.1.1	Target Architecture	139
6.1.2	OpenMP and Offloading	141
6.1.3	Data Models	142
6.1.4	Accelerator Address Space Restricted Offloading	142
6.1.5	Heterogeneous Compilation: State of the Art .	143
6.2	Mixed-Data-Model Offloading	144
6.2.1	Mixed-Data-Model <i>OpenMP</i> Offloading	145
6.3	Mixed-Data-Model Compilation	145
6.3.1	Feasibility in GCC and LLVM	145
6.3.2	Front-end or Optimizer?	146
6.3.3	Our Mixed-Data-Model Compiler in LLVM . .	146

6.4	HW Support for Extended Addressing	153
6.4.1	Additional, Wider Load & Store Instructions .	153
6.4.2	Additional Control and Status Registers (CSRs)	154
6.4.3	Memory-Mapped External Register	155
6.5	Evaluation	156
6.5.1	Methodology	156
6.5.2	Benchmark Results	158
6.6	Related Work	160
6.7	Summary	162
7	A State-of-the-Art Open-Source Heterogeneous Research Platform	163
7.1	Platform	165
7.1.1	Hardware Architecture	166
7.1.2	Toolchain and Compilers	169
7.1.3	Runtime Libraries and Operating System Support	173
7.1.4	Application Programming Interface	176
7.2	Evaluation	178
7.2.1	Application-Level Case Study	180
7.2.2	Runtime-Level and Toolchain Case Study . . .	184
7.2.3	System Architecture-Level Case Study	187
7.2.4	Accelerator ISA-Level Case Study	190
7.3	Related Work	192
7.4	Summary	197
8	Conclusions	199
8.1	Main Results	199
8.2	Outlook	204
A	Acronyms	207
B	Bibliography	213
C	Curriculum Vitae	261