



Master Informatics Eng.

2016/17

A.J.Proença

Concepts from undergrad Computer Systems (1)

(most slides are borrowed, mod's in green)

AJProença, Advanced Architectures, MiEI, UMinho, 2016/17

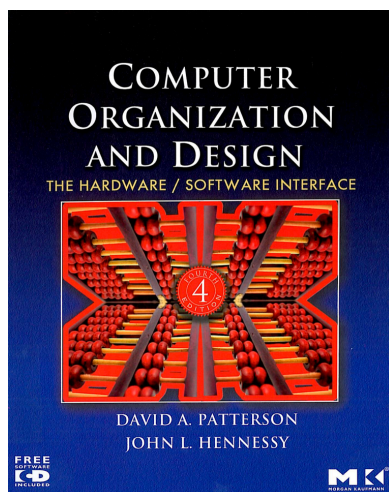
1

Advanced Architectures



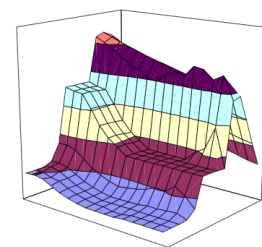
Concepts from undergrad Computer Systems

– most slides are borrowed from



and some from

Computer Systems
A Programmer's Perspective¹
(Beta Draft)



Randal E. Bryant
David R. O'Hallaron

August 1, 2001

more details at
<http://gec.di.uminho.pt/lei/sc/>

AJProença, Advanced Architectures, MiEI, UMinho, 2016/17

2

Background for Advanced Architectures



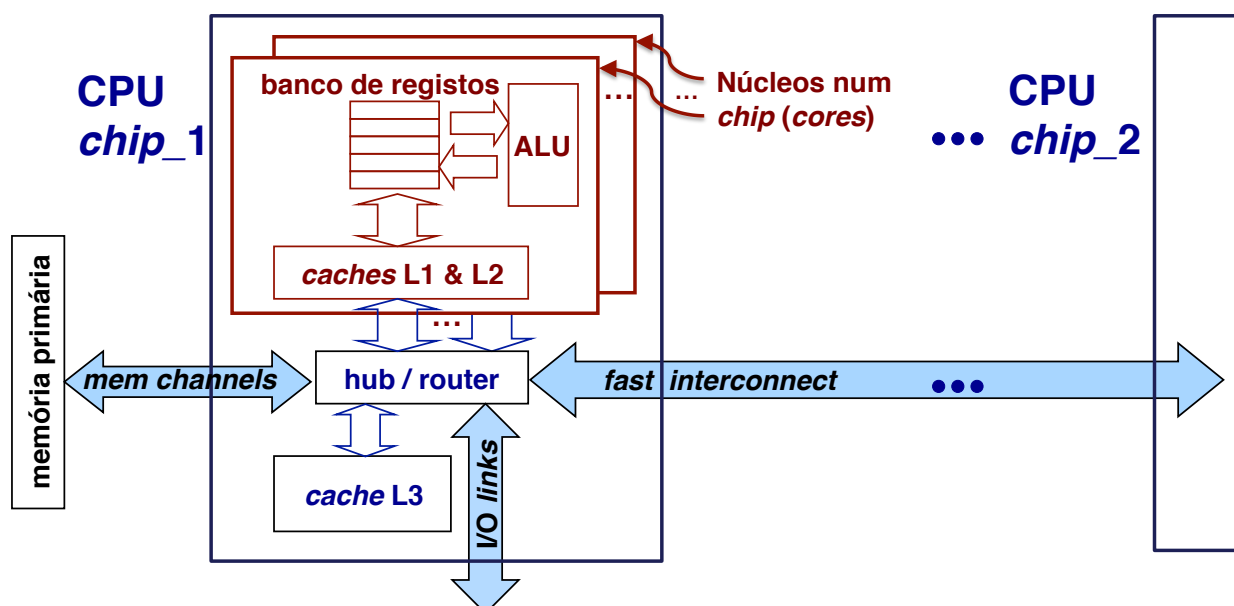
Key concepts to revise:

- *numerical data representation (for error analysis)*
- *ISA (Instruction Set Architecture)*
- *how C compilers generate code (a look into assembly code)*
 - *how scalar and structured data are allocated*
 - *how control structures are implemented*
 - *how to call/return from function/procedures*
 - *what architecture features impact performance*
- *Improvements to enhance performance in a single CPU*
 - *ILP: pipeline, multiple issue, ...*
 - *data parallelism: SIMD/vector processing, ...*
 - *memory hierarchy: cache levels, ...*
 - *thread-level parallelism*

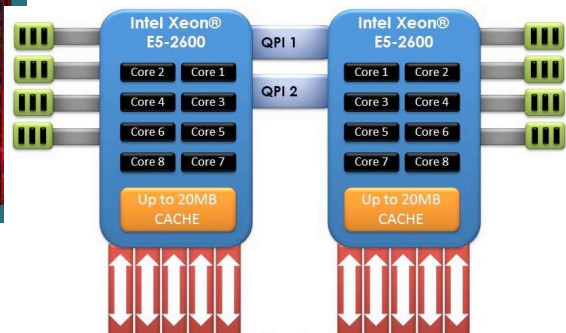
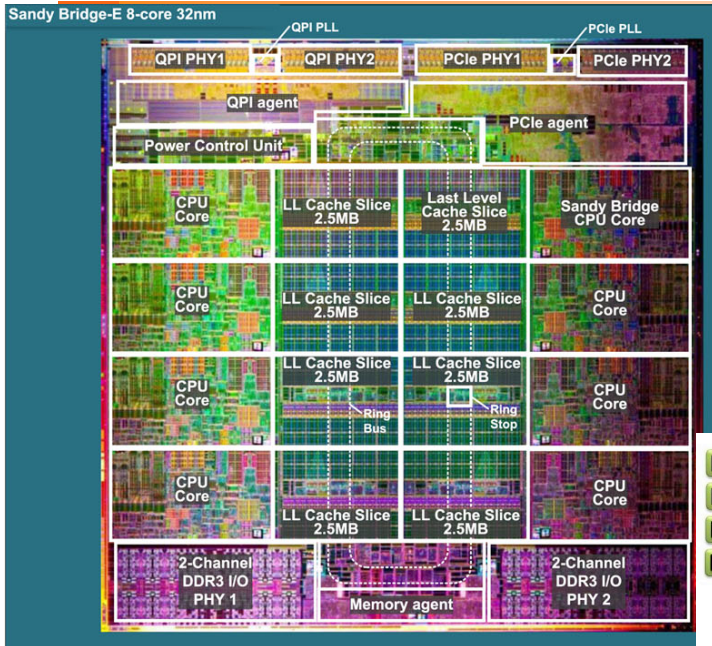
A hierarquia de cache em arquiteturas multicore



As arquiteturas *multicore* mais recentes:



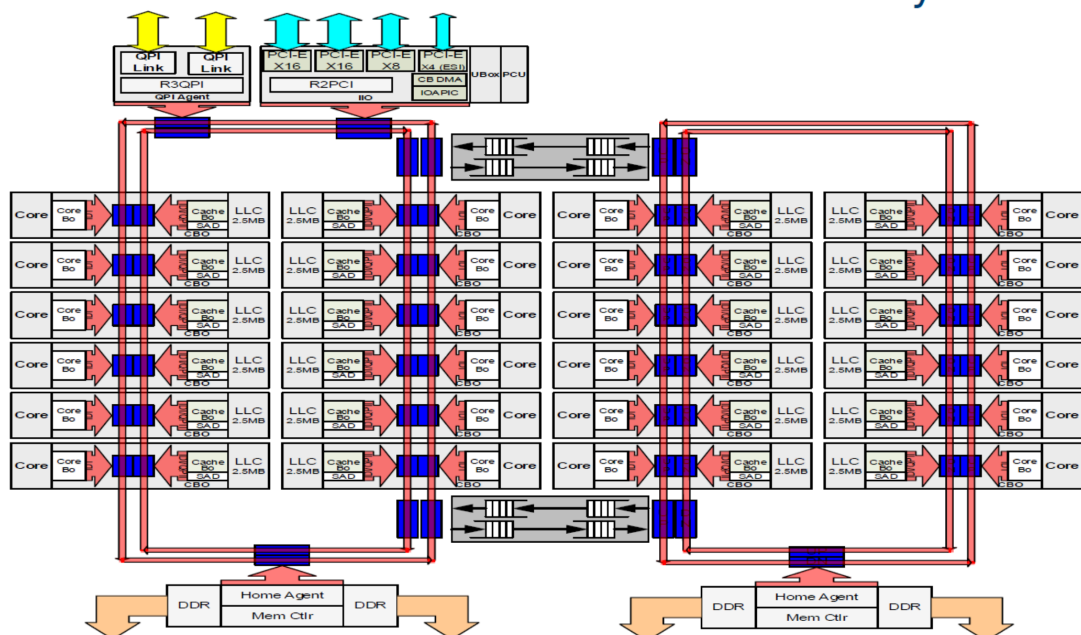
Lançamento da Intel em 2012: Sandy/Ivy Bridge (8-core)



AJProença, Sistemas de Computação, UMinho, 2013/14

Intel in 2016: Broadwell-EP Xeon (22-core)

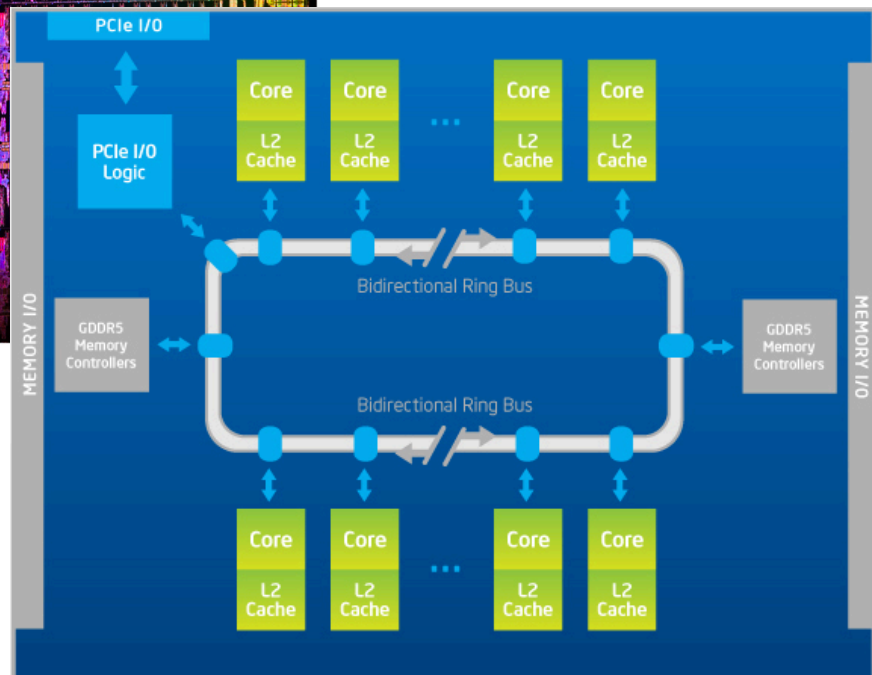
Intel® Xeon® Processor E5 v4 Product Family HCC



AJProença, Advanced Architectures, MiEI, UMinho, 2016/17



Chips da Intel em 2012/13: Xeon Phi com 60 cores



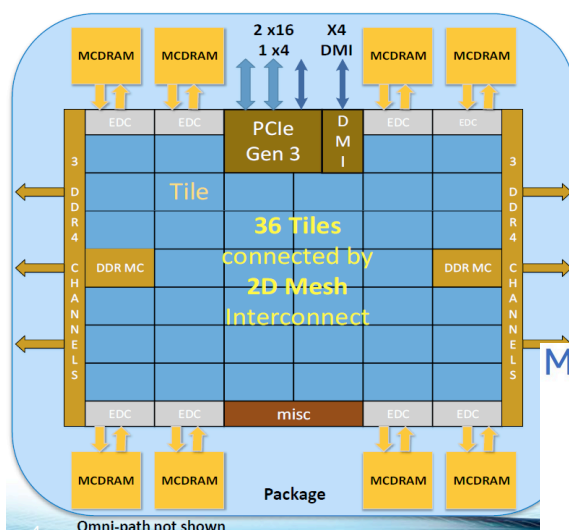
AJProença, Sistemas de Computação, UMinho, 2013/14

7

Intel new Phi in 2016: KNL with 72 cores



Knights Landing Overview



TILE	2 VPU	CHA	2 VPU
	Core	1MB L2	Core

Chip: 36 Tiles interconnected by 2D Mesh

Tile: 2 Cores + 2 VPU/core + 1 MB L2

Memory: MCDRAM: 16 GB on-package; High BW

DDR4: 6 channels @ 2400 up to 384GB

IO: 36 lanes PCIe Gen3. 4 lanes of DMI for chipset

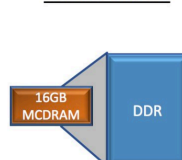
Node: 1-Socket only

Fabric: Omni-Path on-package (not shown)

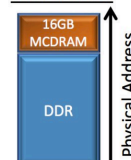
Memory Modes

Three Modes. Selected at boot

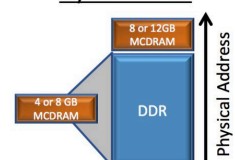
Cache Mode



Flat Mode

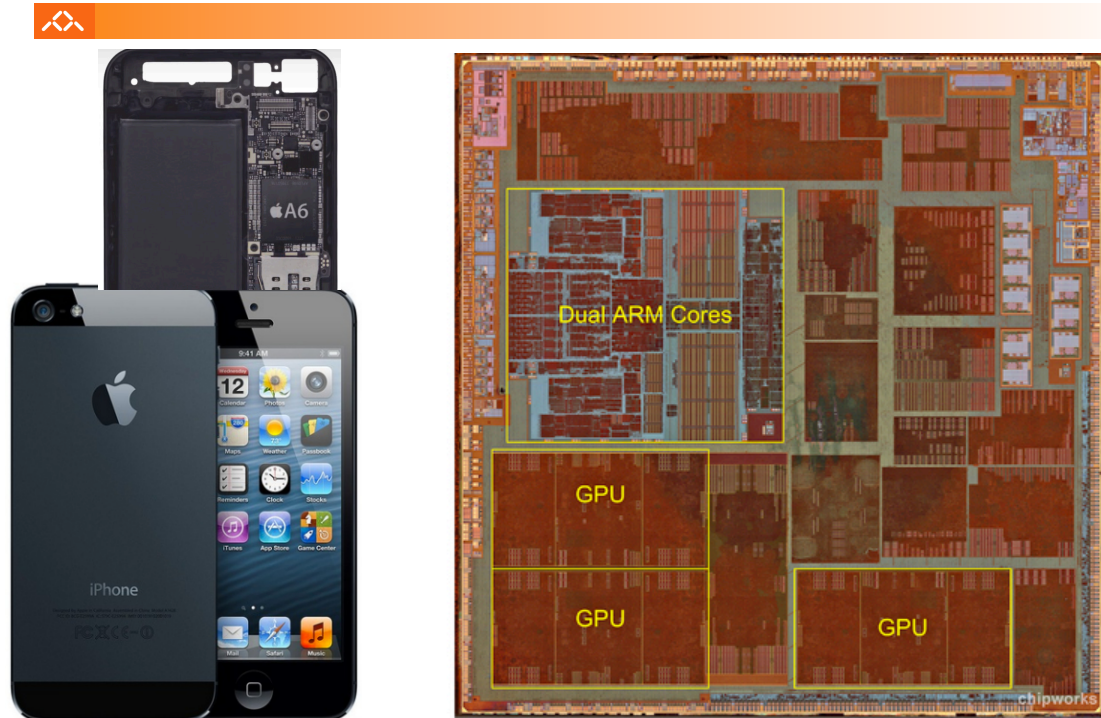


Hybrid Mode



AJProença, Advanced Architectures, MiEI, UMinho,

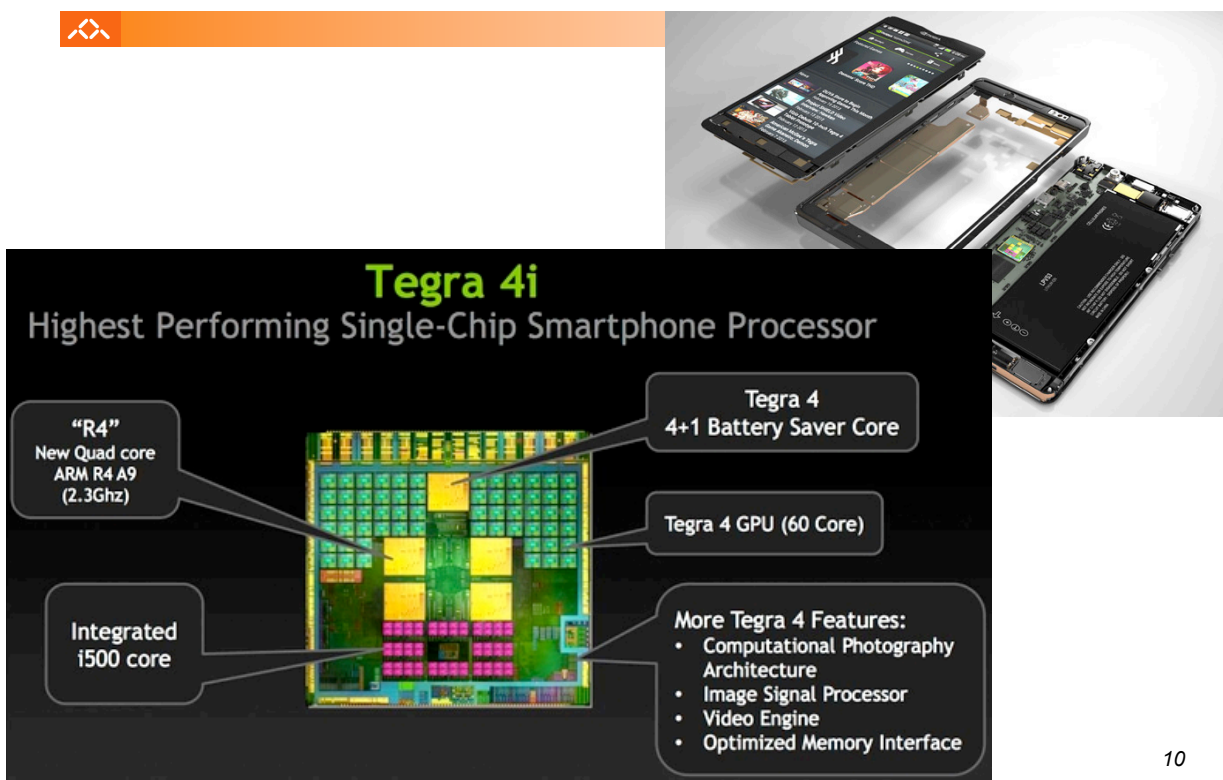
Exemplo de chip com processadores RISC:
2x ARM's no A6 do iPhone 5



AJProença, Sistemas de Computação, UMinho, 2013/14

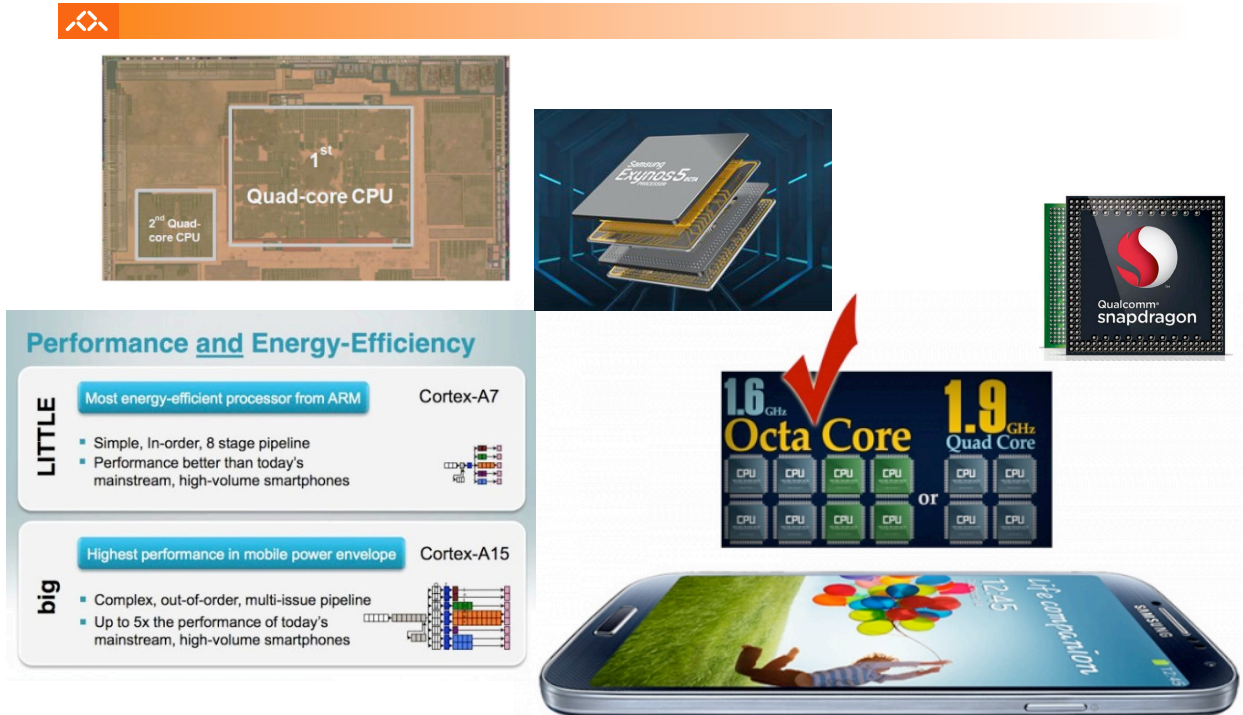
9

Exemplo de chip com processadores RISC:
4+1 ARM's no Tegra 4i da NVidia



10

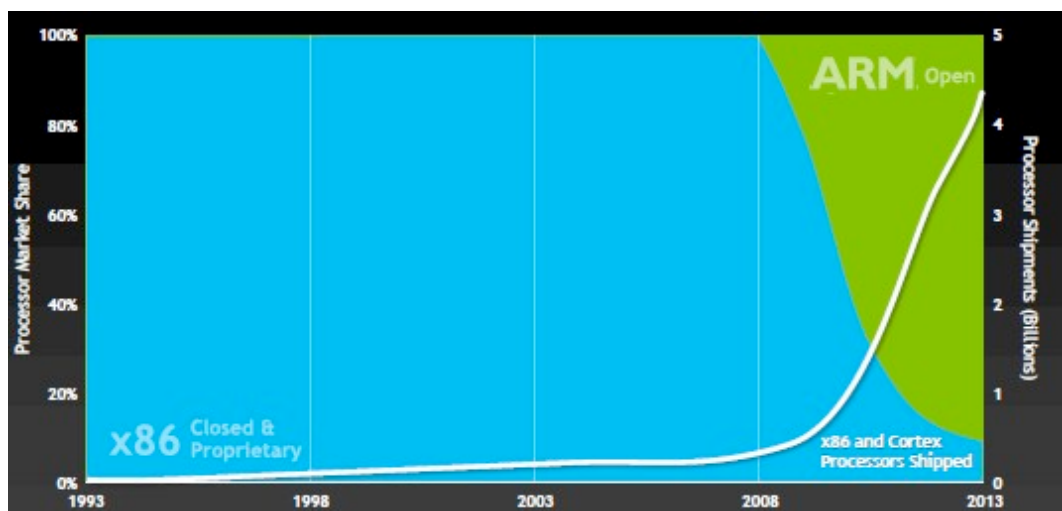
Exemplo de chip com processadores RISC: 4+4 ARM's no Exynos 5 Octa, Galaxy S 4



AJProença, Sistemas de Computação, UMinho, 2013/14

11

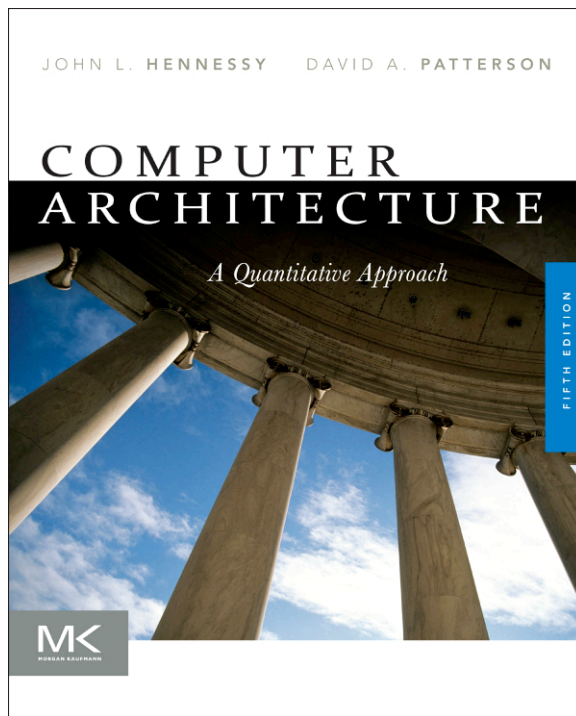
Processadores Intel x86 versus ARM



AJProença, Sistemas de Computação, UMinho, 2013/14

12

Key textbook for AA



Computer Architecture, 5th Edition

Hennessy & Patterson

Table of Contents

Printed Text

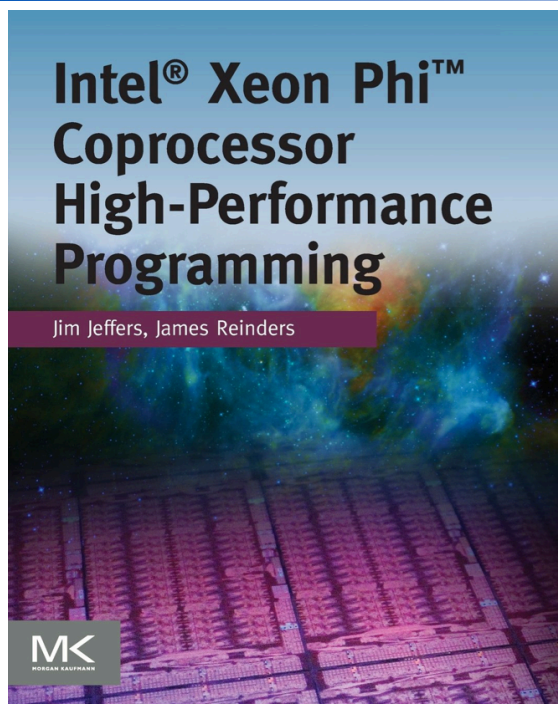
Chap 1: Fundamentals of Quantitative Design and Analysis
Chap 2: Memory Hierarchy Design
Chap 3: Instruction-Level Parallelism and Its Exploitation
Chap 4: Data-Level Parallelism in Vector, SIMD, and GPU Architectures
Chap 5: Multiprocessors and Thread-Level Parallelism
Chap 6: The Warehouse-Scale Computer
App A: Instruction Set Principles
App B: Review of Memory Hierarchy
App C: Pipelining: Basic and Intermediate Concepts

Online

App D: Storage Systems
App E: Embedded Systems
App F: Interconnection Networks
App G: Vector Processors
App H: Hardware and Software for VLIW and EPIC
App I: Large-Scale Multiprocessors and Scientific Applications
App J: Computer Arithmetic
App K: Survey of Instruction Set Architectures
App L: Historical Perspectives



Recommended textbook (1)

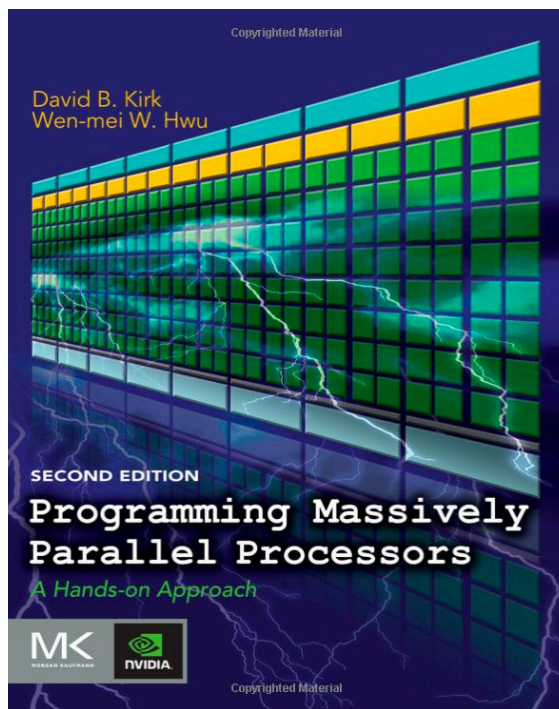


Contents

1. Introduction
2. High Performance examples
3. Benchmarking Apps
4. Real-world Situations
5. Lots of Data (Vectors)
6. Lots of Tasks (not Threads)
7. Processing Parallelism
8. Coprocessor Architecture
9. Coprocessor System Software
10. Linux on the Coprocessor
11. Math Library
12. MPI
13. Profiling
14. Summary



Recommended textbook (2)



Contents

- 1 Introduction
- 2 History of GPU Computing
- 3 Introduction to Data Parallelism and CUDA C
- 4 Data-Parallel Execution Model
- 5 CUDA Memories
- 6 Performance Considerations
- 7 Floating-Point Considerations
- 8 Parallel Patterns: Convolution
- 9 Parallel Patterns: Prefix Sum
- 10 Parallel Patterns: Sparse Matrix-Vector Multiplication
- 11 Application Case Study: Advanced MRI Reconstruction
- 12 Application Case Study: Molecular Visualization and Analysis
- 13 Parallel Programming and Computational Thinking
- 14 An Introduction to OpenCL
- 15 Parallel Programming with OpenACC
- 16 Thrust: A Productivity-Oriented Library for CUDA
- 17 CUDA FORTRAN
- 18 An Introduction to C11 AMP
- 19 Programming a Heterogeneous Computing Cluster
- 20 CUDA Dynamic Parallelism
- 21 Conclusion and Future Outlook

