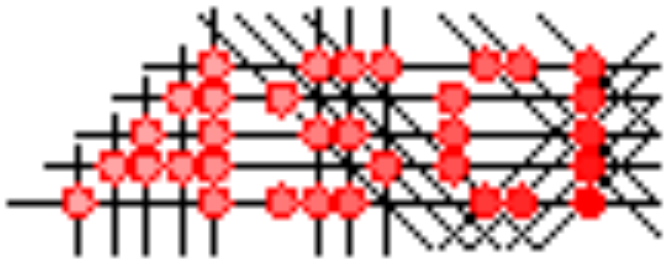
ASCI Spring School on Heterogeneous Computing Systems

Main organizers

Henk Corporaal
TU Eindhoven

Gerard Smit
University of Twente



Spring School Topics

- Heterogeneous Computing Architectures (HCA)
  - Introduction and Motivation, Concepts, Examples
- Accelerator concepts and design
  - GPU / SIMD, CGRA based, Accelerator embedding/comm.
  - Heterogeneous computing with accelerators (e.g. CPUs + GPUs)
- Compilers and Application Mapping to HCA
  - Data Reuse, Polyhedral models, Halide (DSL), OpenMP4, Parallelization and Vectorization
  - OpenCL to FPGA
- Modelling (time, area, energy) and Design Space Exploration of HCAs
- Future architecture directions, like
  - Near Memory Computing
  - Memristor based computing
  - Quantum accelerators
  - Approximate computing
- Heterogeneous Data Centers

Lectures and Special slots

- Introduction + 12 Lecture slots (2 hours each)
- Tuesday afternoon
  - 13:00 - … Social event
- Thursday
  - 16:00 - …presentations by students

Slot # (1-8)	Speakers	Tentative Topic
Monday May 29: Introduction and examples of heterogeneous computing		
2	Gerard Smit (UTwente)	Welcome and introduction
3	Henk Corporaal (TU/e) +	Introduction to Heterogeneous Computing Systems
4	Andre Kokkeler (UTwente)	
5	Stephan Wong (TUDelft)	VLIW / Liquid computing
6	Jan Kuper (QBayLogic)	Clash: From Functional Description to Accelerator
7	Mark Wijtvliet (TU/e)	CGRA (Coarse Grain Reconfigurable Architectures)
8	Mark Wijtvliet, Stephan Wong, Ana Varbanescu	Lab introductions + Handout Questions
Tuesday May 30: Code generation and advanced optimizations		
1	Roel Jordans (Radboud Univ)	Introduction to Code generation, Vectorization
2		
3	Sven Verdoolaege (Leuven)	Advanced Code optimizations, like Loop tiling, Parallelization, and
4		Loop fusion
Afternoon: Social event, visit of National Military Museum, Soesterberg, at the former NATO airbase		
Wednesday May 31: GPU, New languages, and Accelerators, Future systems		
1	Ana Varbanescu (UvA)	CPU+GPU: architecture, language, code partitioning
2		
3	Kees van Berkel (TUE)	Exascale Computing for Radio Astronomy: GPU or FPGA?
4	Rosilde Corvino (Intel)	Halide: the holey grail to computing?
5	Maurice Peemen (FEI)	Deep Learning: Deep Neural Networks, how do they operate?
6		Boosting Deep Learning efficiency
7		Recent developments
8	Said Hamdioui (TUDelft)	Future architectures: Computing in Memory
Thursday June 1: DSE, Accelerators, Future Systems, and Student pitch		
1	Marco Bekooij (NXP)	Modeling and DSE for heterogeneous systems
2		
3	Carmina Almudever (TUDelft)	Future architectures: Quantum Computing
4	Leandro Florin (IBM Research)	Accelerators for the SKA (Square Kilometer Array) Telescope
5	Akash Kumar (TUDresden)	Approximate Computing, going ultra low power
6		
7	Students	Pitch presentations on selected topics
8		Q&A

Daily schedule:

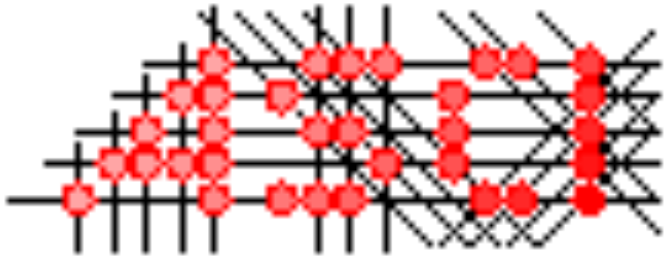
- 9.00 – 10.45: slot 1,2
- break
- 11.00 – 12.45: slot 3,4
- lunch
- 14.00 – 15.45: slot 5,6
- break
- 16.00 – 17.45: slot 7,8
- 18.30 dinner

Student labs (40 h of work, including report writing)

- **CGRA lab:** Mark Wijtvliet
- **Liquid lab:** Stephan Wong
- **GPU lab:** Ana Varbanescu

ASCI credit points

- Students who want to get 4 ASCI credit points
  - Attend the lectures at the spring school
  - Participate in the labs / home assignment
  - Prepare a small paper before June 18
  - Give a short presentation on selected paper topic on Thursday



Why heterogeneous?

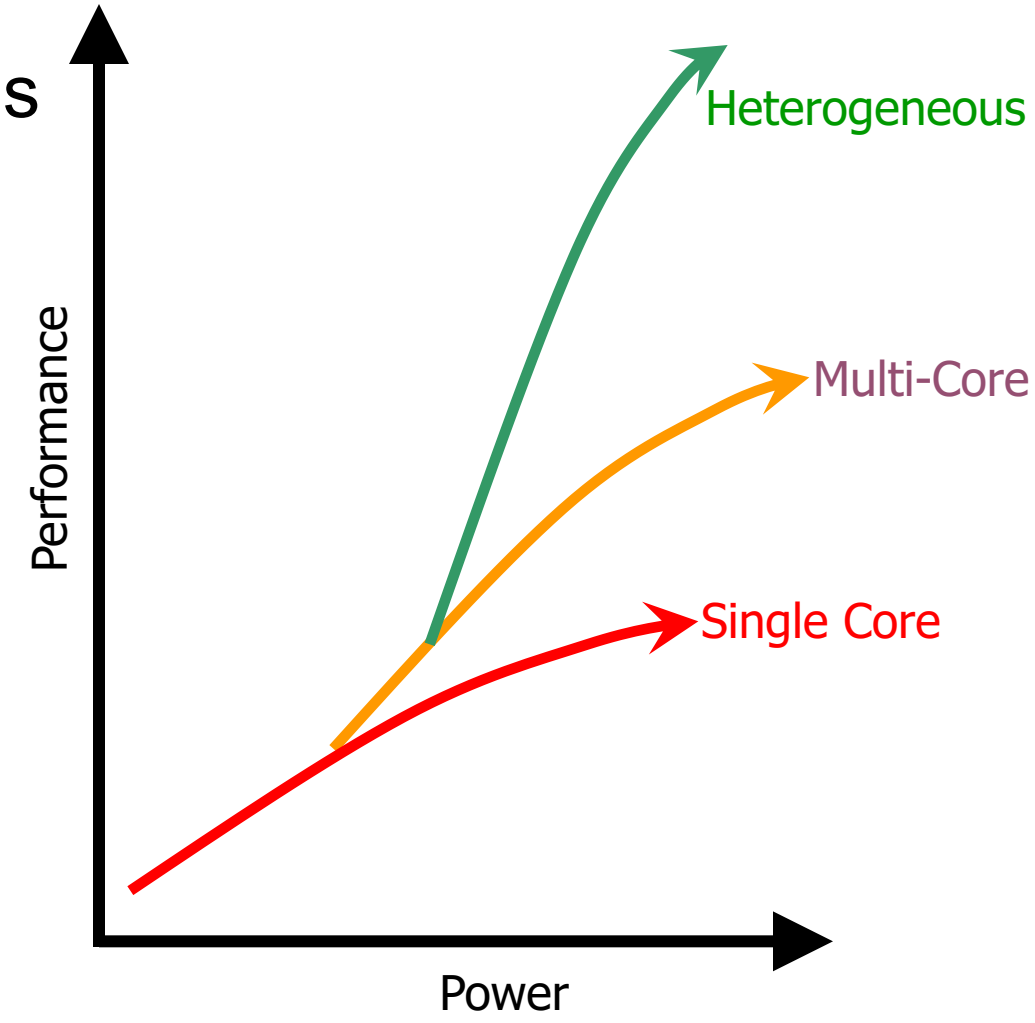
- David Patterson UC Berkeley in 2008

” Today, we have hit the wall as to the practical limit to the amount of power that a microprocessor chip can dissipate; in the past each generation of chip used more power while getting more performance. We need to invent a new way to get more performance without more power ”

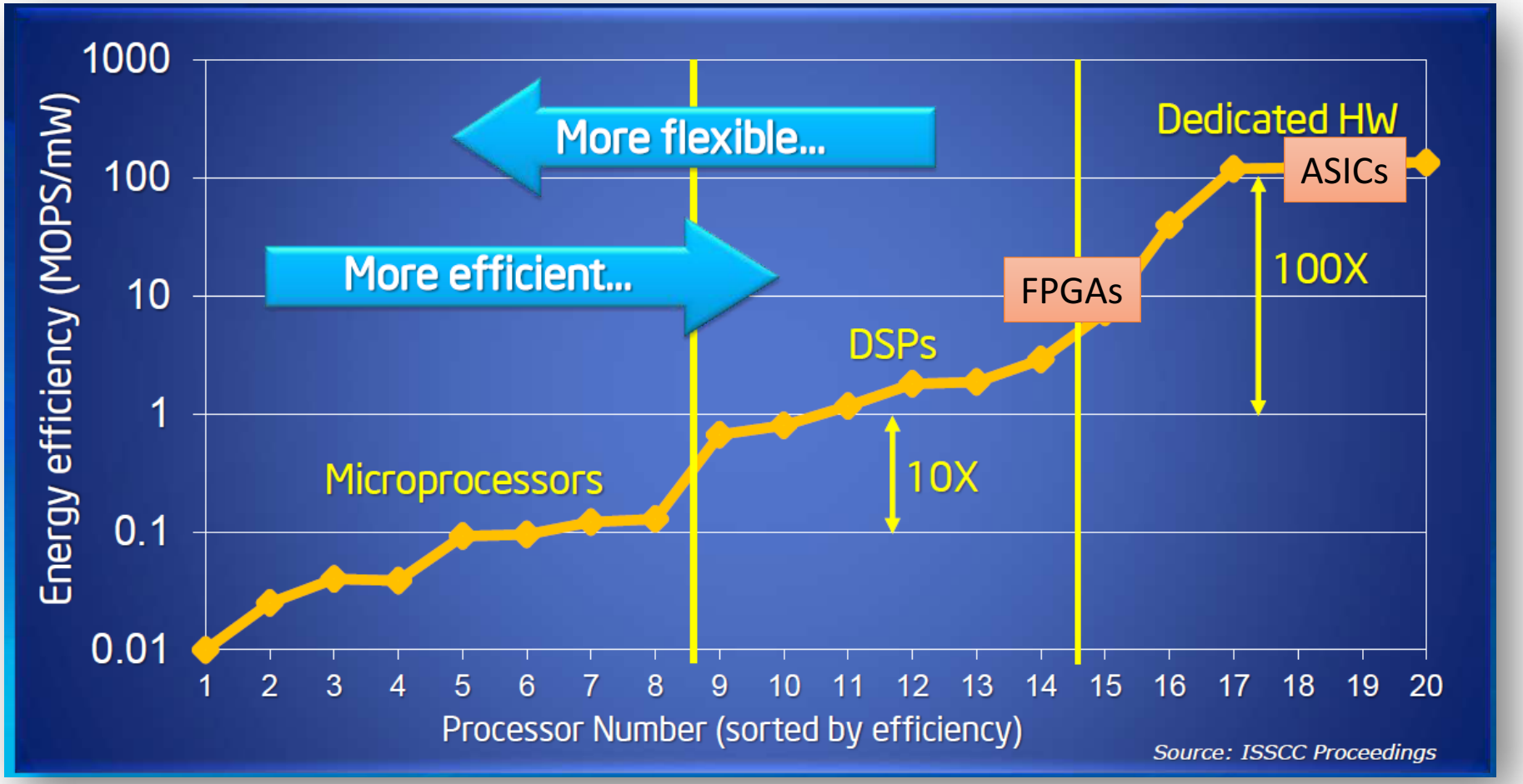
- Power has become a critical design parameter: [FLOPS/Watt]

Why heterogeneous?

- Heterogeneous extends performance and efficiency
- Multi-core throughput performance extended
- Single core performance flat & power limited



Efficiency of processors vs dedicated HW

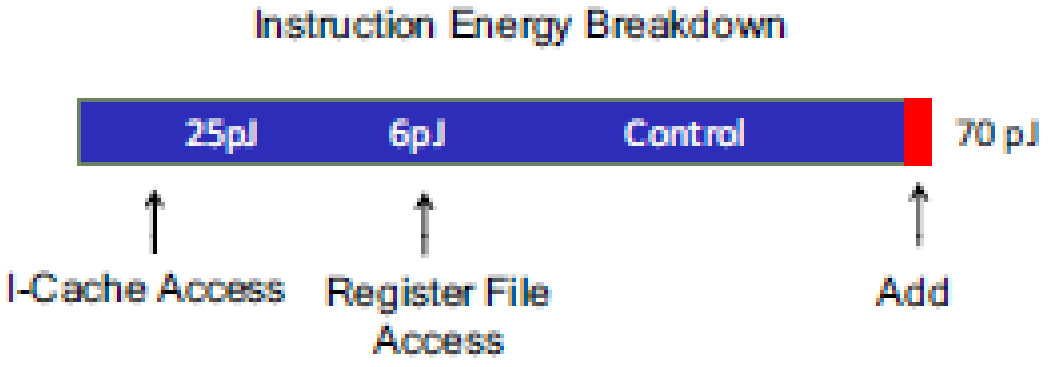


Energy costs of an instruction of a CPU in 45 nm technology

Integer	
Add	
8 bit	0.03pJ
32 bit	0.1pJ
Mult	
8 bit	0.2pJ
32 bit	3.1pJ

FP	
FAdd	
16 bit	0.4pJ
32 bit	0.9pJ
FMult	
16 bit	1.1pJ
32 bit	3.7pJ

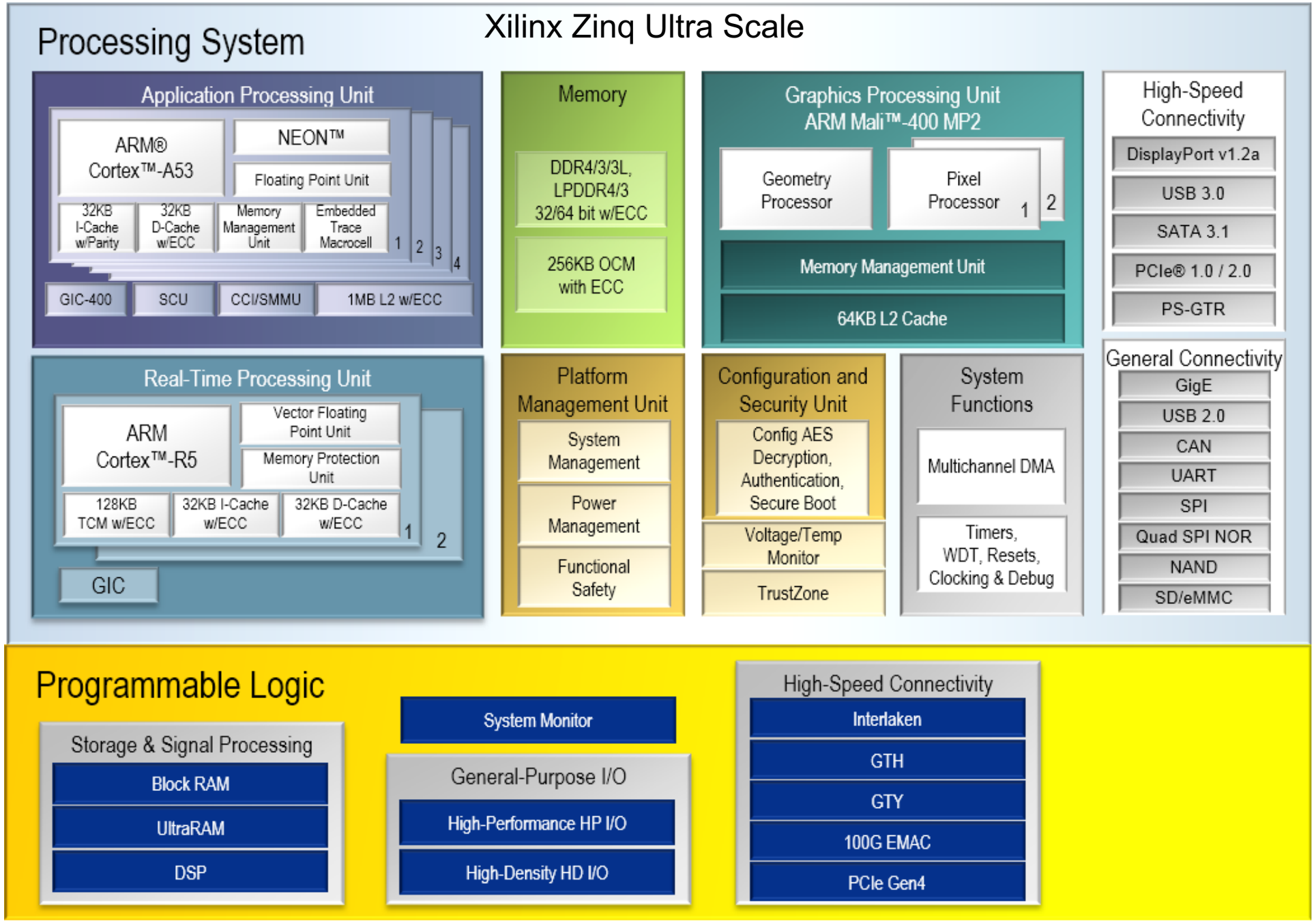
Memory	
Cache	(64bit)
8KB	10pJ
32KB	20pJ
1MB	100pJ
DRAM	1.3-2.6nJ



The costs of data transport

Operation (8-bit operand)	Energy/Op (45 nm)	Cost (vs. ALU)
ALU operation	0.05 pJ	1X
Move 10 mm on-chip	2.4 pJ	50X
Load from on-chip SRAM	2.5 pJ	50X
Send to off-chip DRAM	**320 pJ**	**6,400X**

* Data from J. Brunhaver, W. Dally, M. Horowitz, Stanford University


Xilinx Zinq Ultra Scale
Processing System
Application Processing Unit
ARM® Cortex™-A53
NEON™
Floating Point Unit
32KB I-Cache w/Parity
32KB D-Cache w/ECC
Memory Management Unit
Embedded Trace Macrocell
1
2
3
4
GIC-400
SCU
CCI/SMMU
1MB L2 w/ECC
Memory
DDR4/3/3L, LPDDR4/3 32/64 bit w/ECC
256KB OCM with ECC
Graphics Processing Unit ARM Mali™-400 MP2
Geometry Processor
Pixel Processor
1
2
Memory Management Unit
64KB L2 Cache
High-Speed Connectivity
DisplayPort v1.2a
USB 3.0
SATA 3.1
PCIe® 1.0 / 2.0
PS-GTR
Real-Time Processing Unit
ARM Cortex™-R5
Vector Floating Point Unit
Memory Protection Unit
128KB TCM w/ECC
32KB I-Cache w/ECC
32KB D-Cache w/ECC
1
2
GIC
Platform Management Unit
System Management
Power Management
Functional Safety
Configuration and Security Unit
Config AES Decryption, Authentication, Secure Boot
Voltage/Temp Monitor
TrustZone
System Functions
Multichannel DMA
Timers, WDT, Resets, Clocking & Debug
General Connectivity
GigE
USB 2.0
CAN
UART
SPI
Quad SPI NOR
NAND
SD/eMMC
Programmable Logic
Storage & Signal Processing
Block RAM
UltraRAM
DSP
System Monitor
General-Purpose I/O
High-Performance HP I/O
High-Density HD I/O
High-Speed Connectivity
Interlaken
GTH
GTY
100G EMAC
PCIe Gen4

Challenges of heterogeneous

- Architectures of heterogeneous: efficiency / programmability
  - Ultra-low power: approximate computing
  - Coarse-grain accelerators
  - Computation in memory

- Modelling applications and DSE for heterogeneous
- Code generation / compilers for heterogeneous
- Programming large-scale heterogeneous systems

- Combination CPU + GPU

ASCI spring school

- Lectures of experts on certain topics
  - Ask questions during or after the lectures
- Lab sessions to get hands-on experience
- Discuss your topics with your fellow students; there is plenty of time during coffee/lunch breaks and evenings

Spring school is setup to stimulate interaction and promote lively discussions between participants

Questions?

Have a fruitful 4 days

www.asci.tudelft.nl