Master Informatics Eng.

2019/20

A.J.Proença

Data Parallelism 5 (*other PUs, ...*)

(most slides are borrowed)

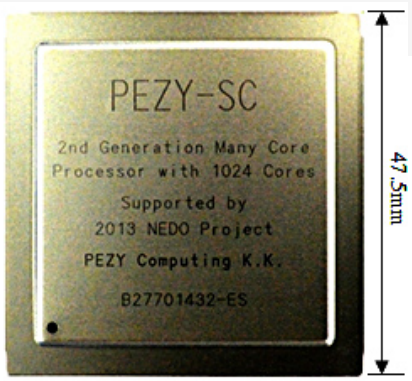
Beyond Vector/SIMD architectures

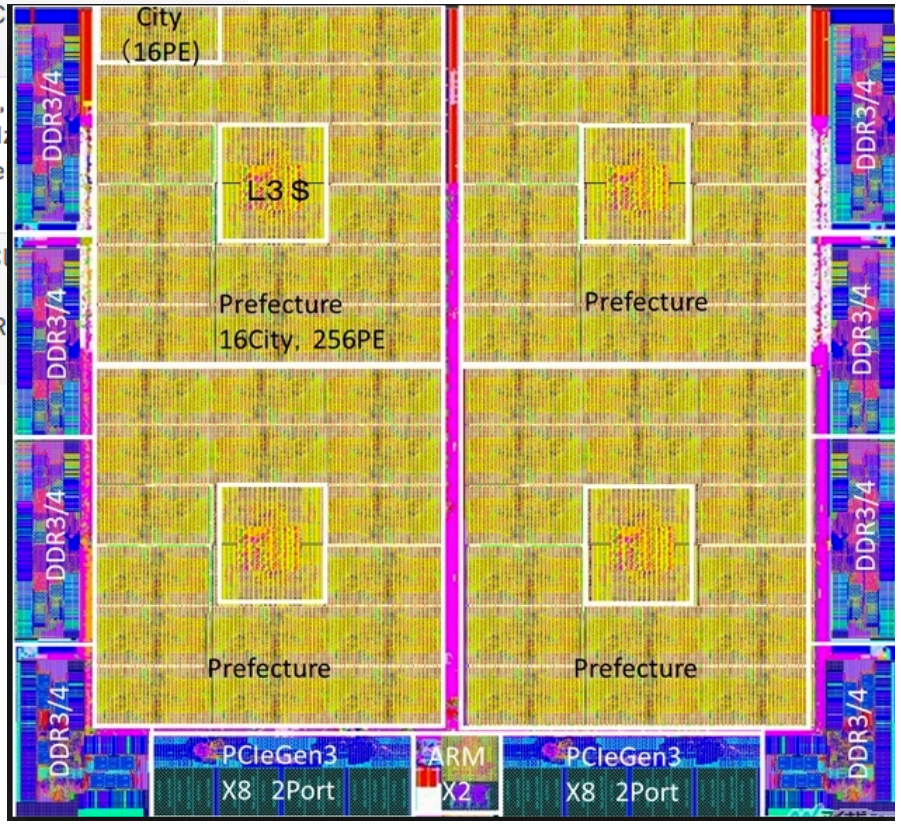
- Vector/SIMD-extended architectures are hybrid approaches
  - mix (super)**scalar + vector** op capabilities on a single device
  - **highly pipelined** approach to reduce memory access penalty
  - **tightly-closed access to shared memory**: lower latency
- Evolution of Vector/SIMD-extended architectures
  - **PU (Processing Unit) cores with wider vector units**
    - x86 many-core: **Intel MIC / Xeon KNL**
    - others: **...**
  - **coprocessors (require a host scalar processor): accelerator devices**
    - on disjoint physical memories (e.g., **Xeon KNC** with PCI-Expr, **PEZY-SC**)
    - ISA-free architectures: **...**
    - focus on SIMT/SIMD to hide memory latency: **GPU**-type approach
    - ...
  - **heterogeneous PUs in a SoC: multicore PUs with GPU-cores**
    - ...

PEZY-SC: Peta_Exa_Zetta_Yotta-SuperComputer:
a 1024-core many-core processor chip, each core 8-way SMT

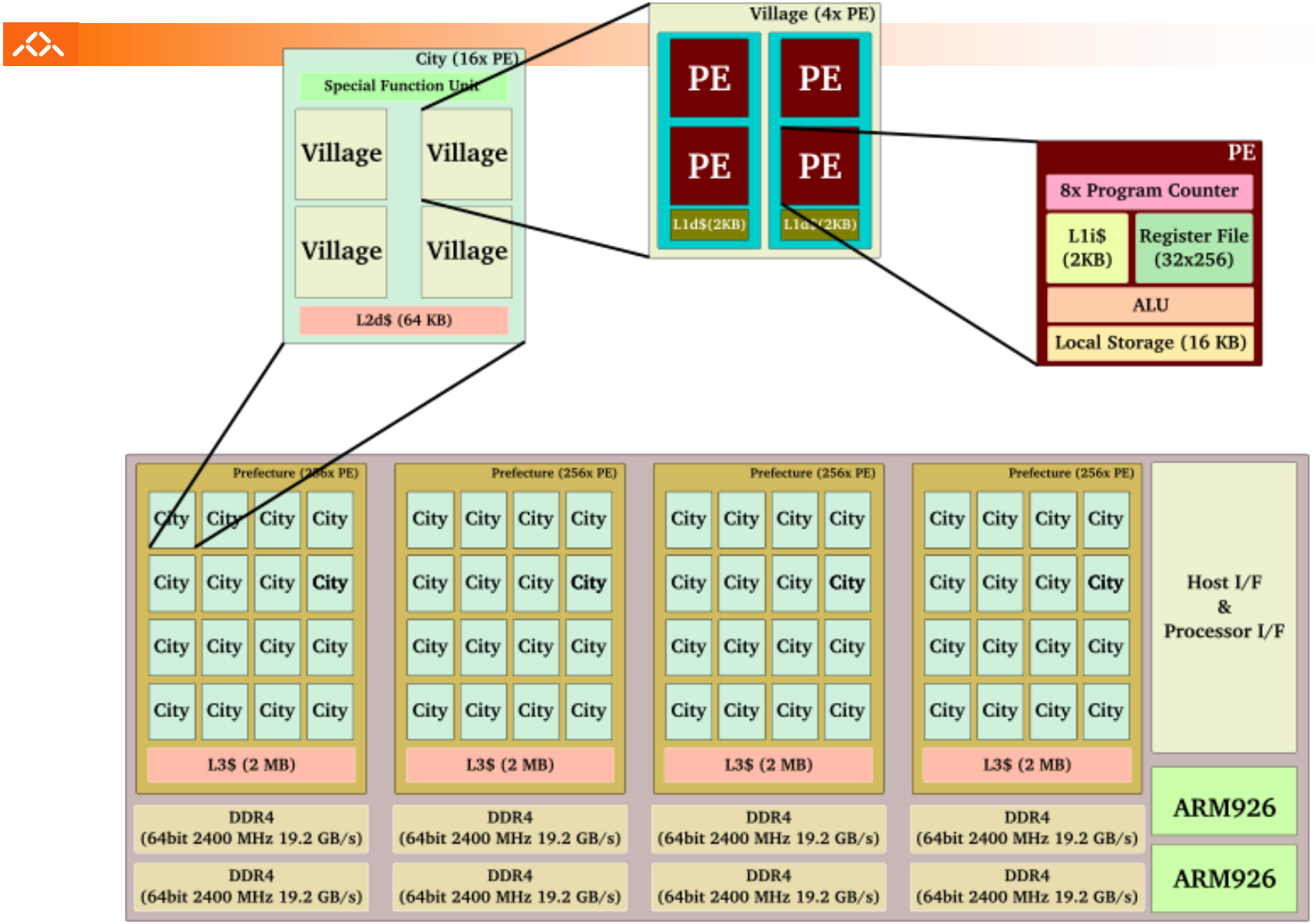
Green500 Rank	TOP500 Rank	MFLOPS/W	Site	System	Total Power(kW)
1	160	7031.4	RIKEN	ExaScaler-1.4 80Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SC	50.3
2	392	6841.3	High Energy Accelerator Research Organization /KEK	ExaScaler-1.4 16Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband, PEZY-SC	28.3
3	366	6217.9	High Energy Accelerator Research Organization /KEK	ExaScaler 32U256SC Cluster, Intel Xeon E5-2660v2 10C Infiniband FDR, PEZY-SC	32.6
4	215	5272.1	GSI Helmholtz Center	ASUS ESC4000 FDR/G2S, Xeon E5-2690v2 10C 3GH Infiniband FDR, AMD Fire S9150	
5	469	4258.1	GSIC Center, Tokyo Institute of Technology	LX 1U-4GPU/104Re-1G C Intel Xeon E5-2620v2 6C 2.100GHz, Infiniband FDR K20x	

Green500 list
June'2015

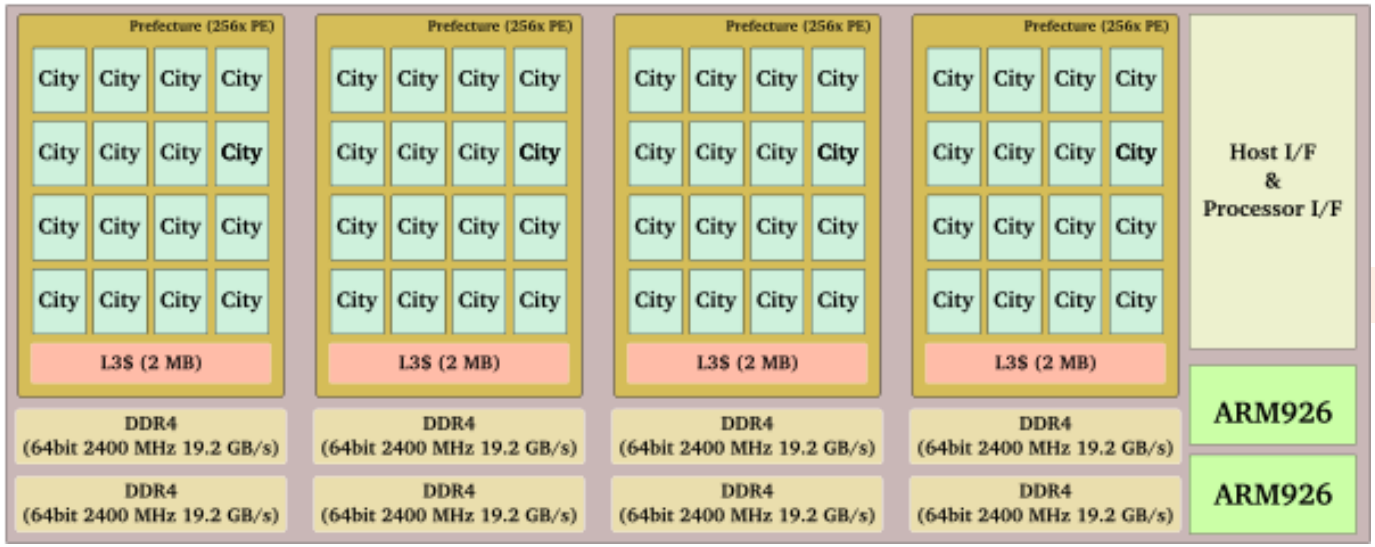




PEZY-SC: Peta_Exa_Zetta_Yotta-SuperComputer:
a 1024-core many-core processor chip

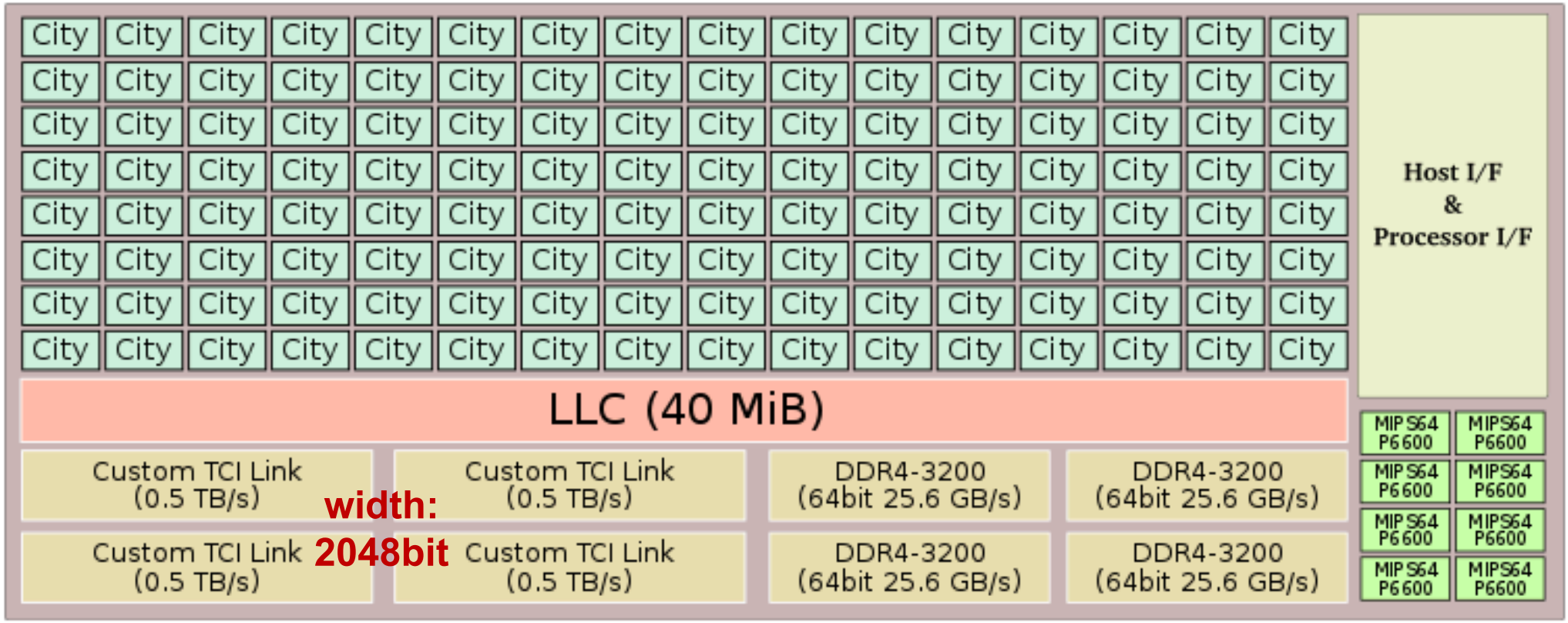


Evolution: the PEZY-SC2



PEZY-SC with 2x 32-bit ARM cores (2015)

PEZY-SC2 with 8x 64-bit MIPS cores sharing 40 MiB LLC (2017)



PEZY-SC2 in Green500

Green500 List for November 2017

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	259	**Shoubu system B** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	794,400	842.0	50	17.009
2	307	**Suiren2** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. High Energy Accelerator Research Organization /KEK Japan	762,624	788.2	47	16.759
3	276	**Sakura** - ZettaScaler-2.2, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	794,400	824.7	50	16.657
4	149	**DGX SaturnV Volta** - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100 , Nvidia NVIDIA Corporation United States	22,440	1,070.0	97	15.113
5	4	**Gyoukou** - ZettaScaler-2.2 HPC system, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , ExaScaler Japan Agency for Marine-Earth Science and Technology Japan	19,860,000	19,135.8	1,350	14.173

None of these systems are in Nov'19 list, only a PEZY-SC2 in 2nd position

The PEZY-SCx Road Map

PEZY-SCx Processor Roadmap

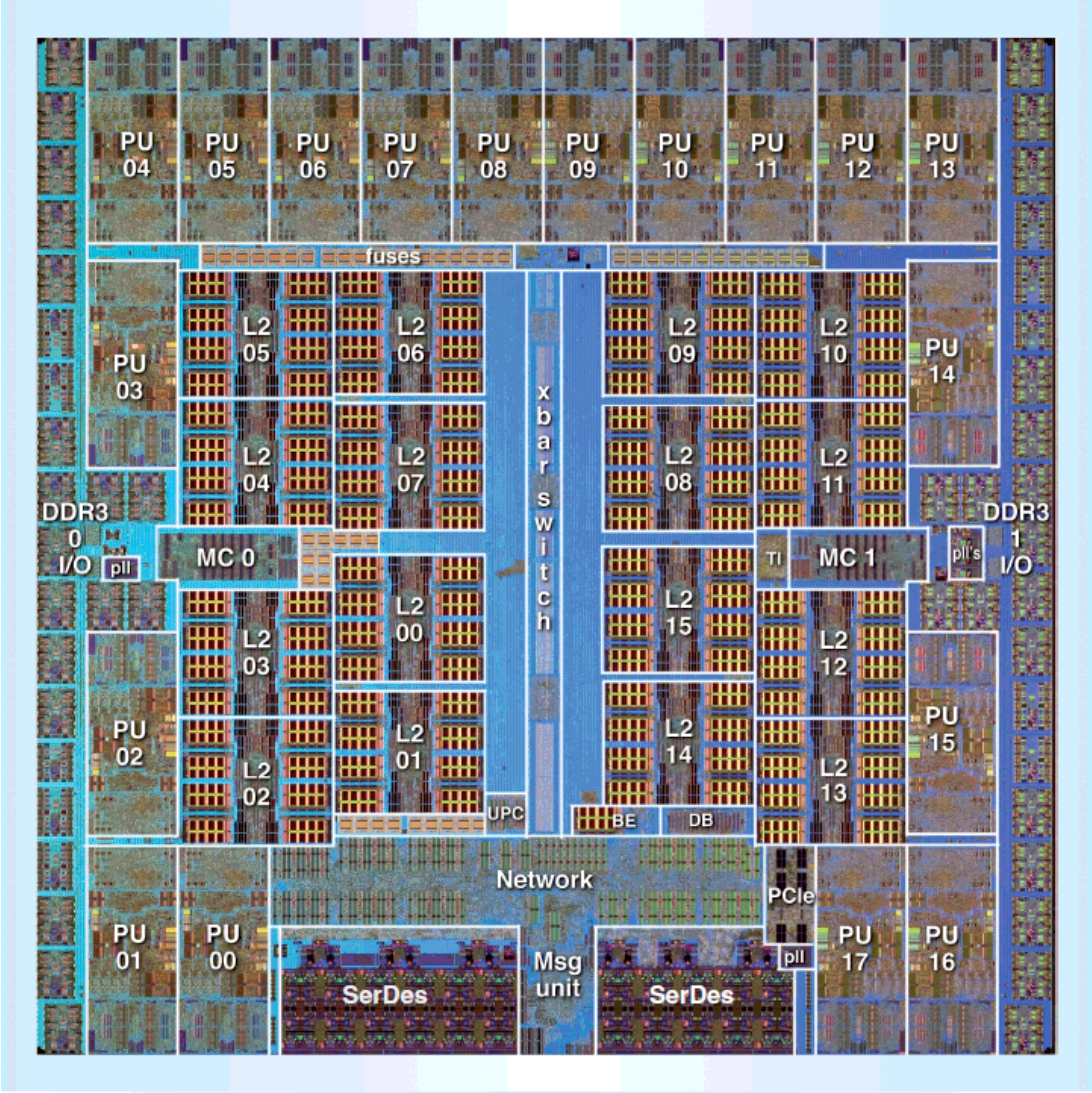
	PEZY-SC	PEZY-SC2	PEZY-SC3	PEZY-SC4
Process	28nm	16nm	7nm	5nm
Die Size	412mm2	620mm2	700mm2	740mm2
Number of Cores	**1,024**	**2,048**	**8,192**	**16,384**
Core Voltage	**0.9V**	**0.8V**	**0.65V**	**0.55V**
Core Clock	733MHz	1GHz	1.33GHz	1.6GHz
DRAM-IO	DDR4	DDR4	DDR4/5	DDR5
DDR Clock	2,133MHz	2,666MHz	3.6GHz	4GHz
Port	8	4	4	4
Wide-IO Clock	-	2GHz DDR	3 GHz DDR	3GHz DDR
Wide-IO Width		1,024bit	2,048bit	4,096bit
Wide-IO Ports		4	8	8
Memory Bandwidth	**153.6GB/s**	**2.1TB/s**	**12.2TB/s**	**24.4TB/s**
Peripheral IO	PCI3e Gen3	PCIe Gen4	Custom Optical	Custom Optical
Peripheral IO lane	24	32	128	512
Peripheral IO Bandwidth	**32GB/s**	**64GB/s**	**256GB/s**	**1TB/s**
DP Performance	**1.5TFLOPS**	**4.1TFLOPS**	**21.8TFLOPS**	**52.5TFLOPS**
SP Performance	3.0TFLOPS	8.2TFLOPS	43.6TFLOPS	105TFLOPS
HP Performance	-	16.4TFLOPS	87.2TFLOPS	210TFLOPS
Power Consumption	100W	200W	400W	640W
Power Efficiency	15GFLOPS/w	20.5GFLOPS/w	54.5GFLOPS/w	82.0GFLOPS/w
System Efficiency	**6.7GFLOPS/w**	**15GFLOPS/w**	**40GFLOPS/w**	**60GFLOPS/w**

PEZY-SC3 expected in end 2019

Beyond Vector/SIMD architectures

- Vector/SIMD-extended architectures are hybrid approaches
  - mix (super)**scalar + vector** op capabilities on a single device
  - **highly pipelined** approach to reduce memory access penalty
  - **tightly-closed access to shared memory**: lower latency
- Evolution of Vector/SIMD-extended architectures
  - **PU (Processing Unit) cores with wider vector units**
    - x86 many-core: **Intel MIC / Xeon KNL**
    - others: **IBM** Power BlueGene/Q, **ShenWay** 260, **Matrix-2000, A64FX Arm**
  - **coprocessors (require a host scalar processor): accelerator devices**
    - on disjoint physical memories (e.g., **Xeon KNC** with PCI-Expr, **PEZY-SC**)
    - ISA-free architectures: **...**
    - ...
  - **heterogeneous PUs in a SoC: multicore PUs with GPU-cores**
    - ...

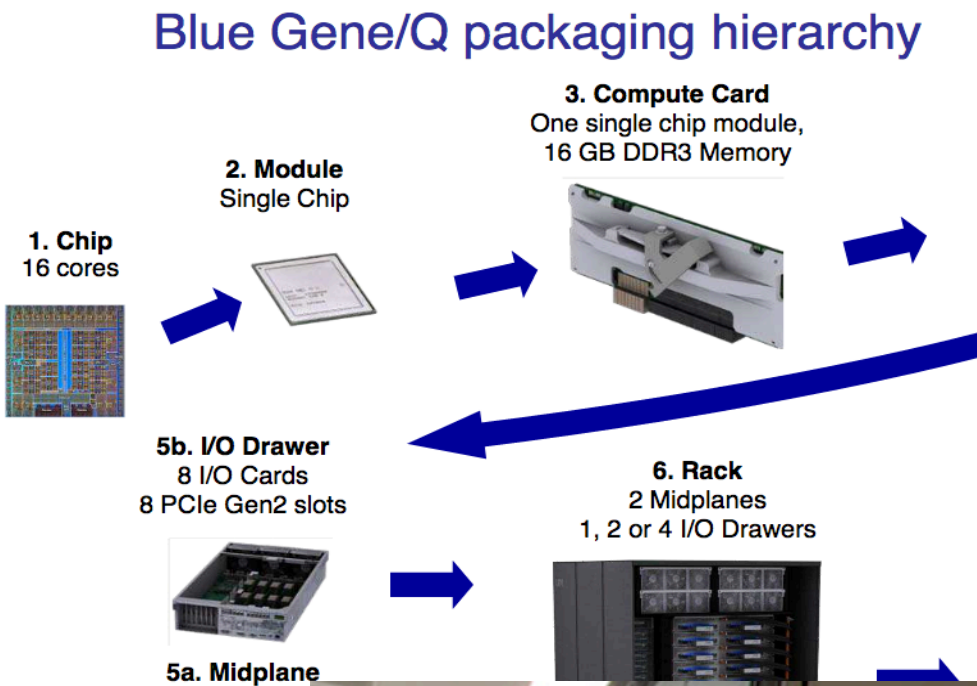
IBM Power BlueGene/Q Compute (chip)



Features:

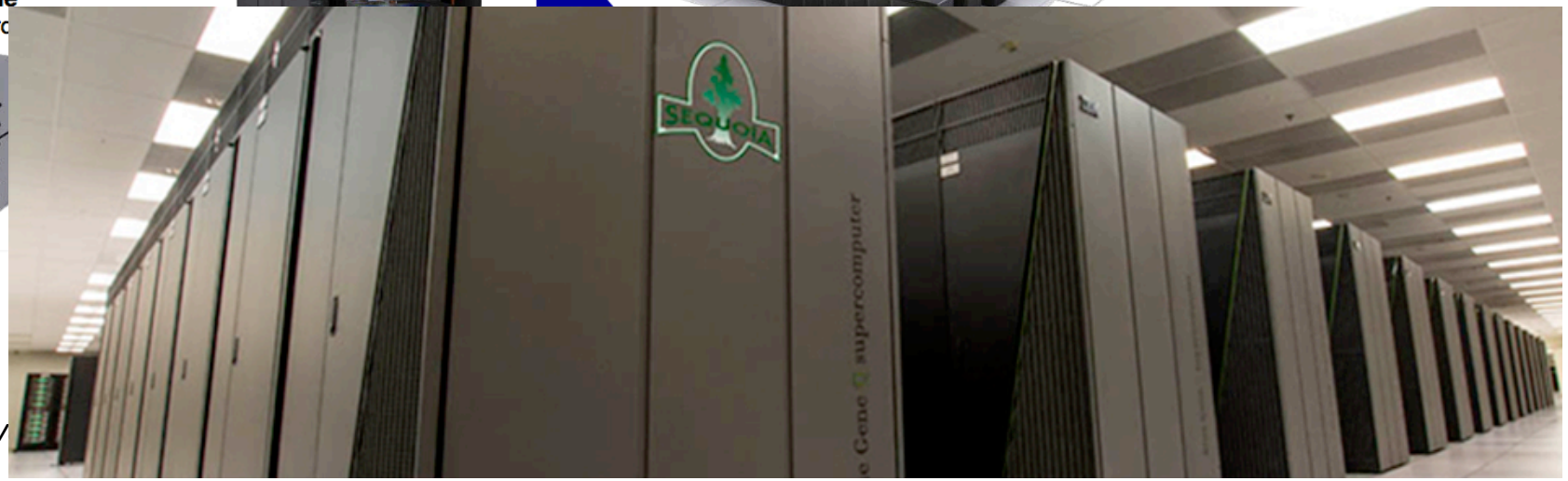
- launched in 2010/11
  (TOP500: #1 in Jun12, #4 in Jun16)
- 18-cores
  - 16 compute,
    1 OS support, 1 redundant
  - 64 bits PowerISA
  - 1.6 GHz
  - L1 I/D cache => 16 kiB / 16 kiB
  - **each core: quad-FPU**
    (4-wide double precision SIMD)
  - **each core: 4-way SMT**
- shared L2 cache: 32 MiB
- dual memory controller
- IBM ended development of BlueGene project in 2015...

IBM Power BlueGene/Q Compute (Sequoia)



Ref: SC2010

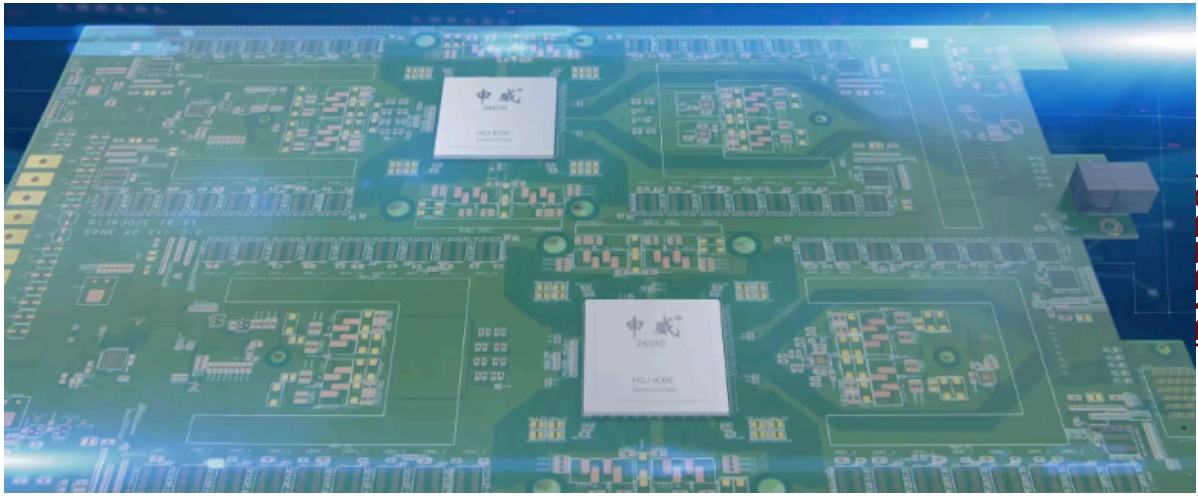
Jun'12: #1
Nov'12: #2
Jun'13: #3
Nov'13: #3
Jun'14: #3
Nov'14: #3
Jun'15: #3
Nov'15: #3
Jun'16: #4
Nov'16: #4
Jun'17: #5
Nov'17: #6
Jun'18: #8
Nov'18: #10
Jun'19: #13
Nov'19: #12



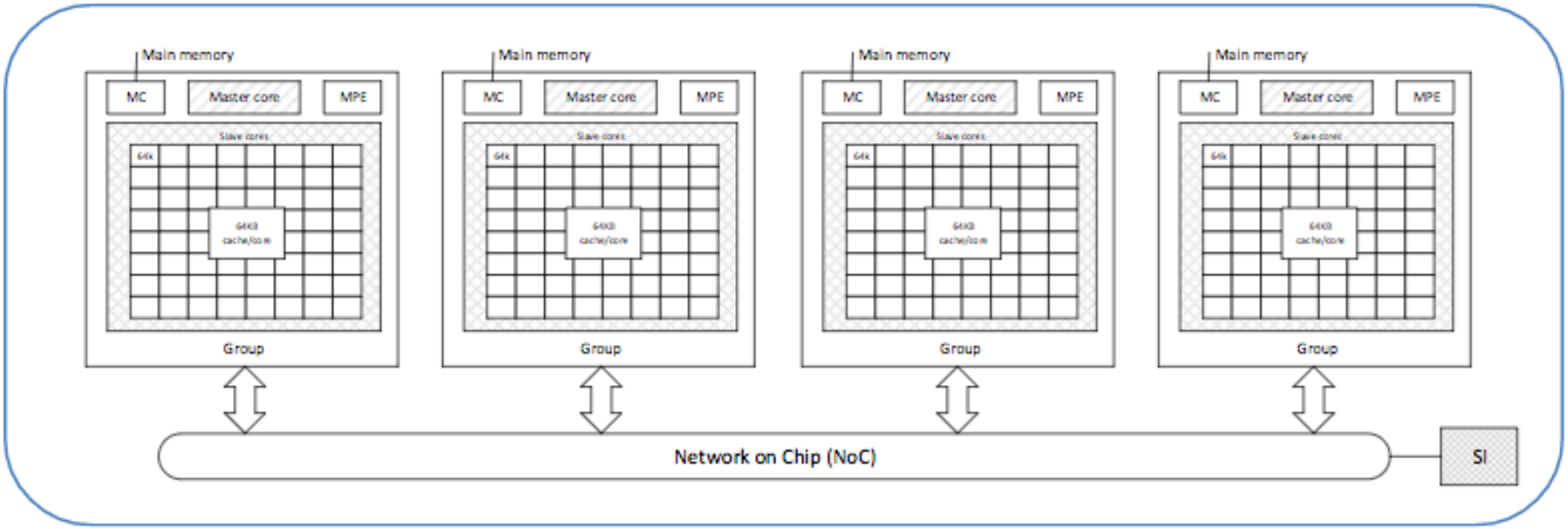
The Sunway TaihuLight

#1 Jun'16-Nov'18 TOP500

One card with two nodes
(two SW26010 chips)



SW26010: the 4x64-core 64-bit RISC processor (w/ **256-bit vector** instructions & only cache L1)



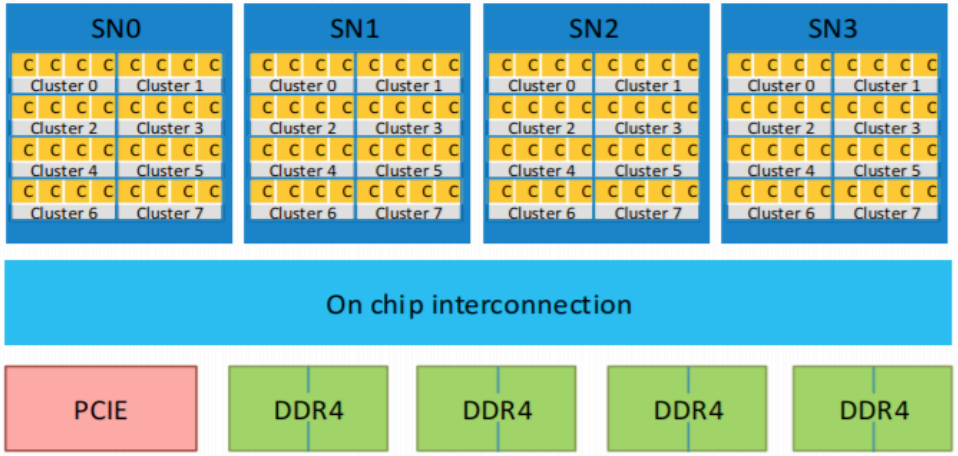
http://www.netlib.org/utk/people/JackDongarra/PAPERS/sunway-report-2016.pdf

Replacing the KNC in Tianhe-2A: the Matrix-2000 accelerator

Matrix-2000 accelerator

- **Chip specification**
  - 128cores
    - 4 super-nodes (SN)
    - 8 clusters per SN
    - 4 cores per cluster
    - Core
      - Self-defined 256-bit vector ISA
      - 16 DP flops/cycle per core
  - Peak performance: 2.4576Tflops@1.2GHz

    4 SNs x 8 clusters x 4cores x 16 flops x 1.2 GHz = 2.4576 Tflops
  - Peak power dissipation: ~240w
  - Interface
    - 8 DDR4-2400 channels
    - X16 PCIE 3.0 EP Port

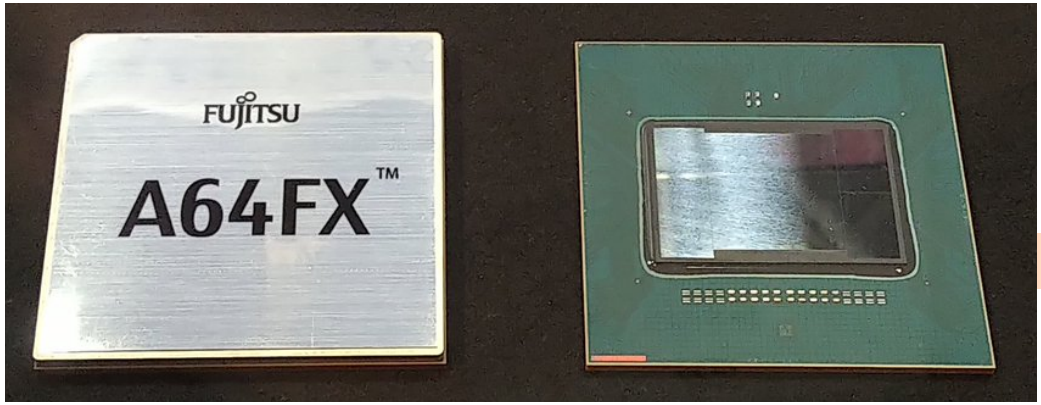


https://www.nextplatform.com/2018/02/14/look-whats-store-chinas-tianhe-2a-supercomputer/

Fujitsu's A64FX Arm & PEZY-SC2 in Green500

Green500 List for November 2019

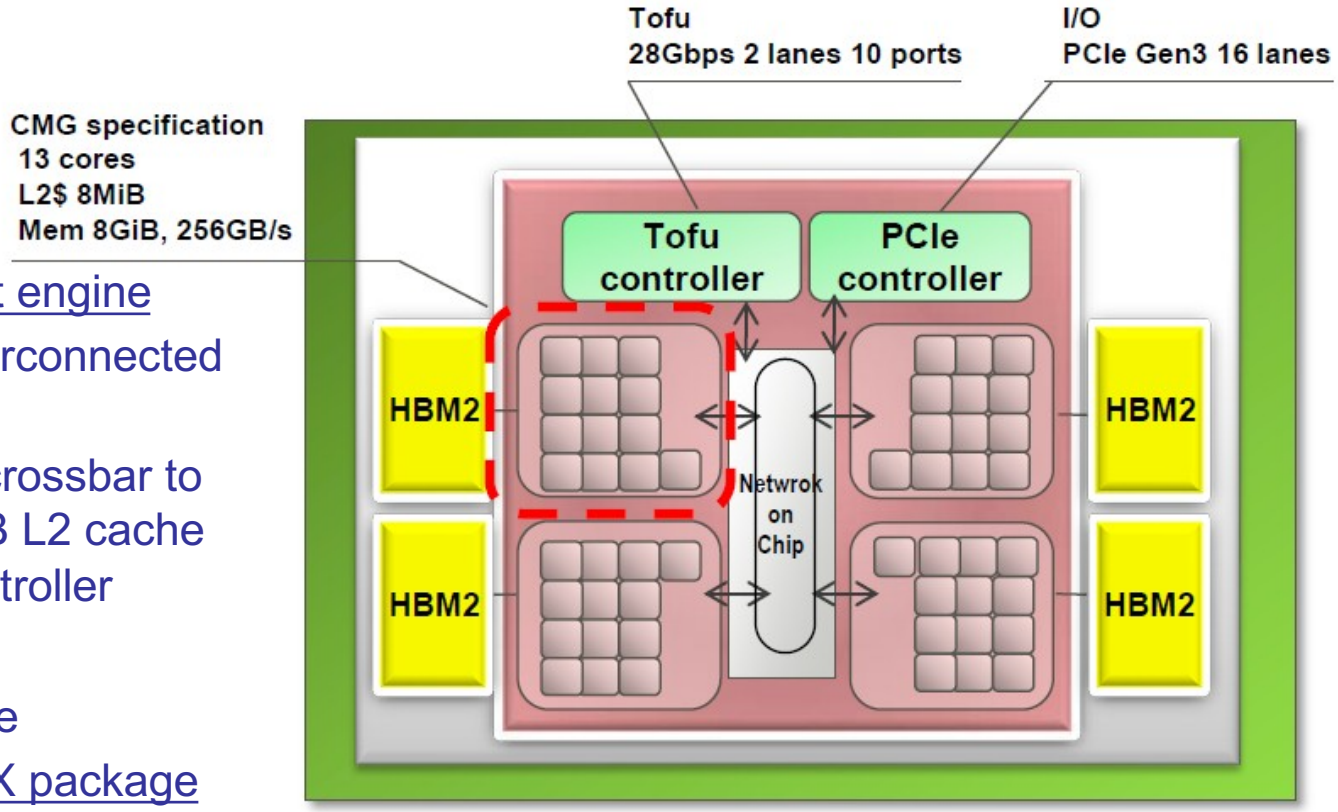
Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	159	**A64FX prototype** - Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, Tofu interconnect D , Fujitsu Fujitsu Numazu Plant Japan	36,864	1,999.5	118	16.876
2	420	**NA-1** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	1,271,040	1,303.2	80	16.256
3	24	**AiMOS** - IBM Power System AC922, IBM POWER9 20C 3.45GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100 , IBM Rensselaer Polytechnic Institute Center for Computational Innovations (CCI) United States	130,000	8,045.0	510	15.771
4	373	**Satori** - IBM Power System AC922, IBM POWER9 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100 SXM2 , IBM MIT/MGHPCC Holyoke, MA United States	23,040	1,464.0	94	15.574
5	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	10,096	14.719



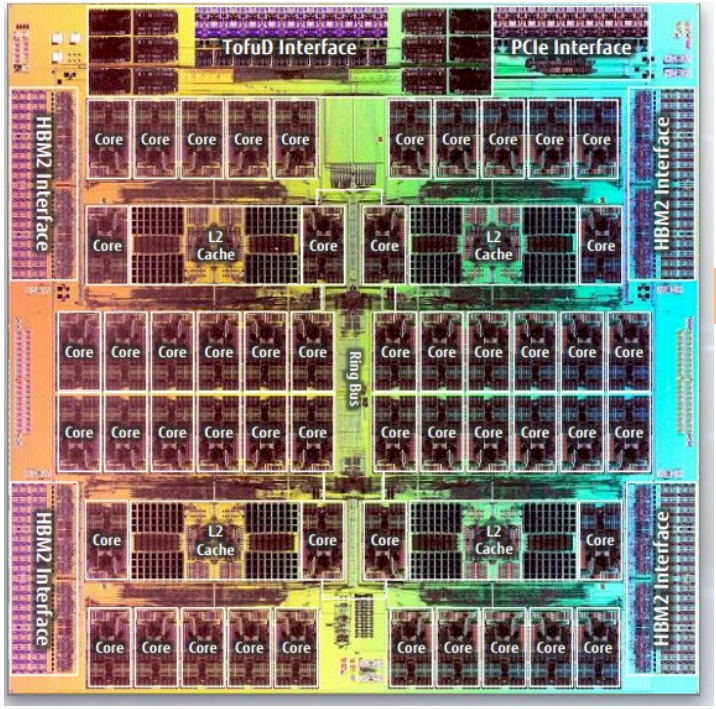
Fujitsu's A64FX Arm Chip: 48+4 cores

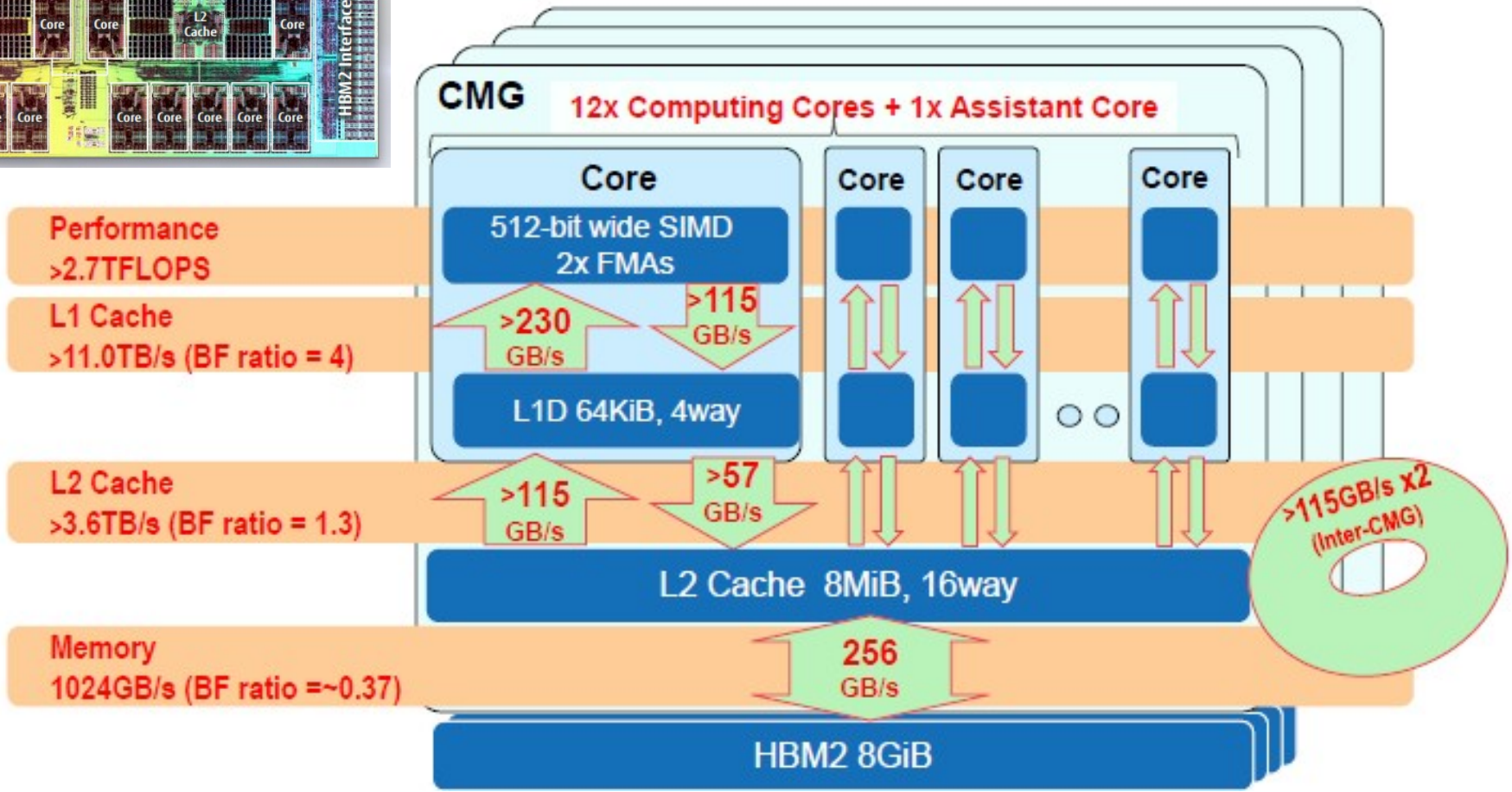
A64FX Arm:

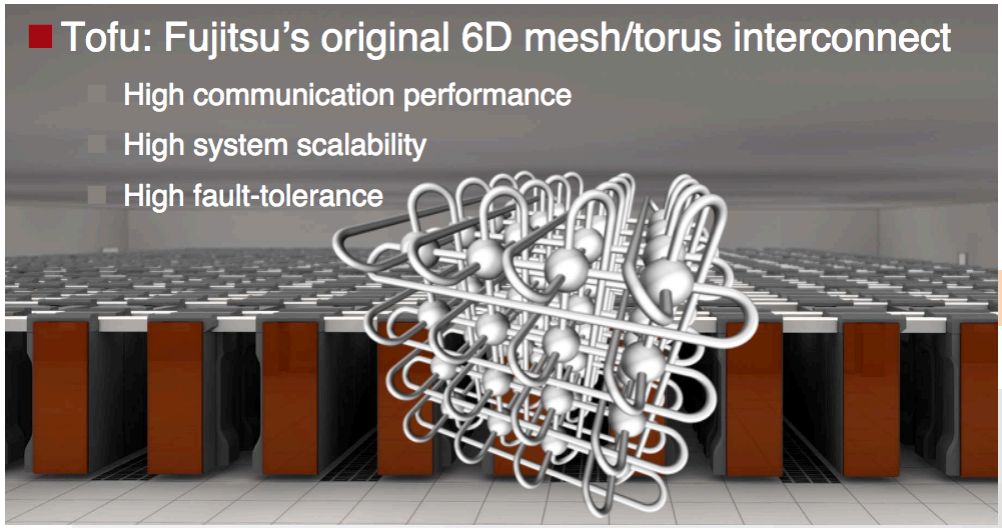
- Armv8.2-A spec with 512-bit SVE extensions
- HP math and a dot-product engine
- 4 core memory groups interconnected with a double ring bus
- cores in CMG linked by a crossbar to a 16-way associative 8 MiB L2 cache and to the HBM2 mem controller
- No L3 cache
- a Tofu3 controller on the die
- Cray CS500 will use A64FX package



Block diagram of the A64FX chip



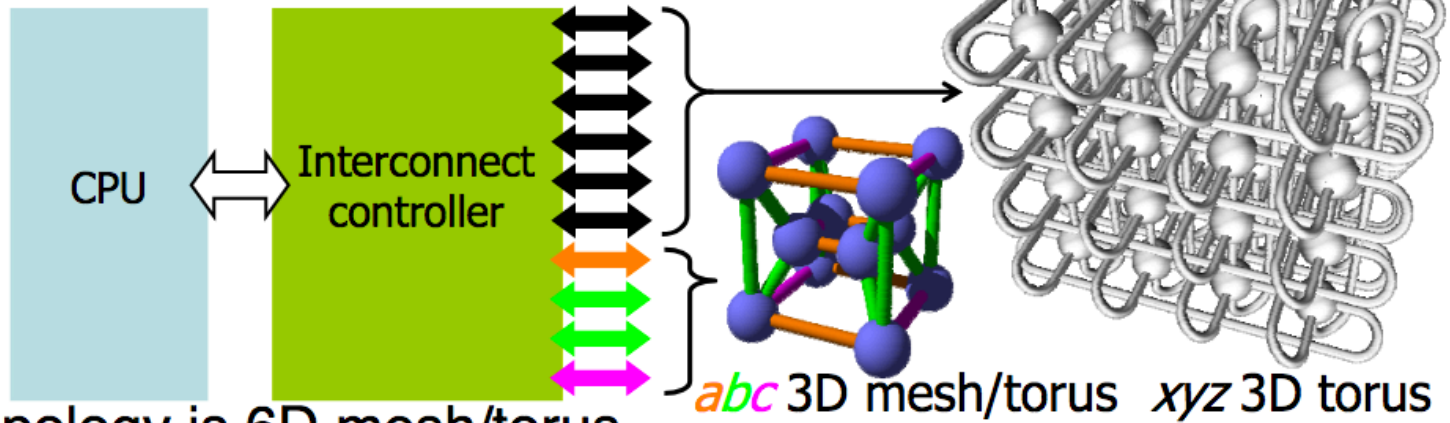




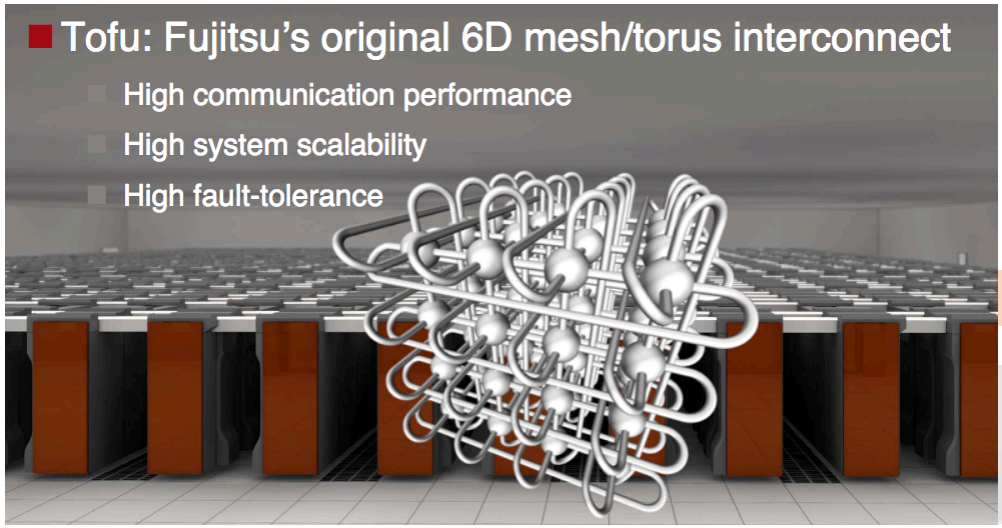
Tofu 3: 6D mesh/torus interconnect

FUJITSU

- 6 links ⇨ Scalable *xyz* 3D torus
- 4 links ⇨ Fixed size *abc* 3D mesh/torus
  - |*a*|=2, |*b*|=3, |*c*|=2 ⇨ 12 nodes



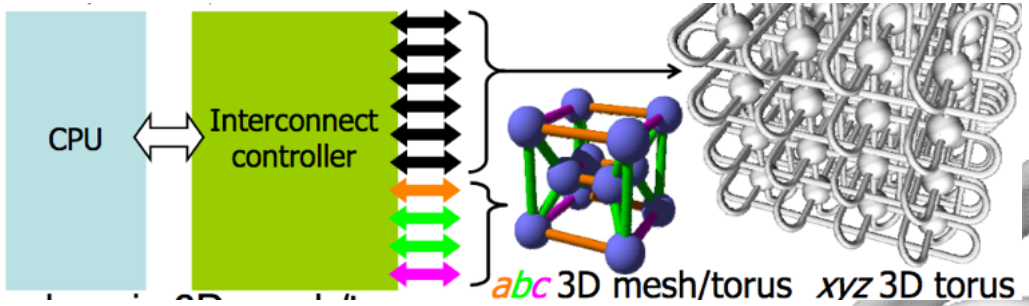
- Total topology is 6D mesh/torus
  - Cartesian product of *xyz* and *abc* mesh/torus

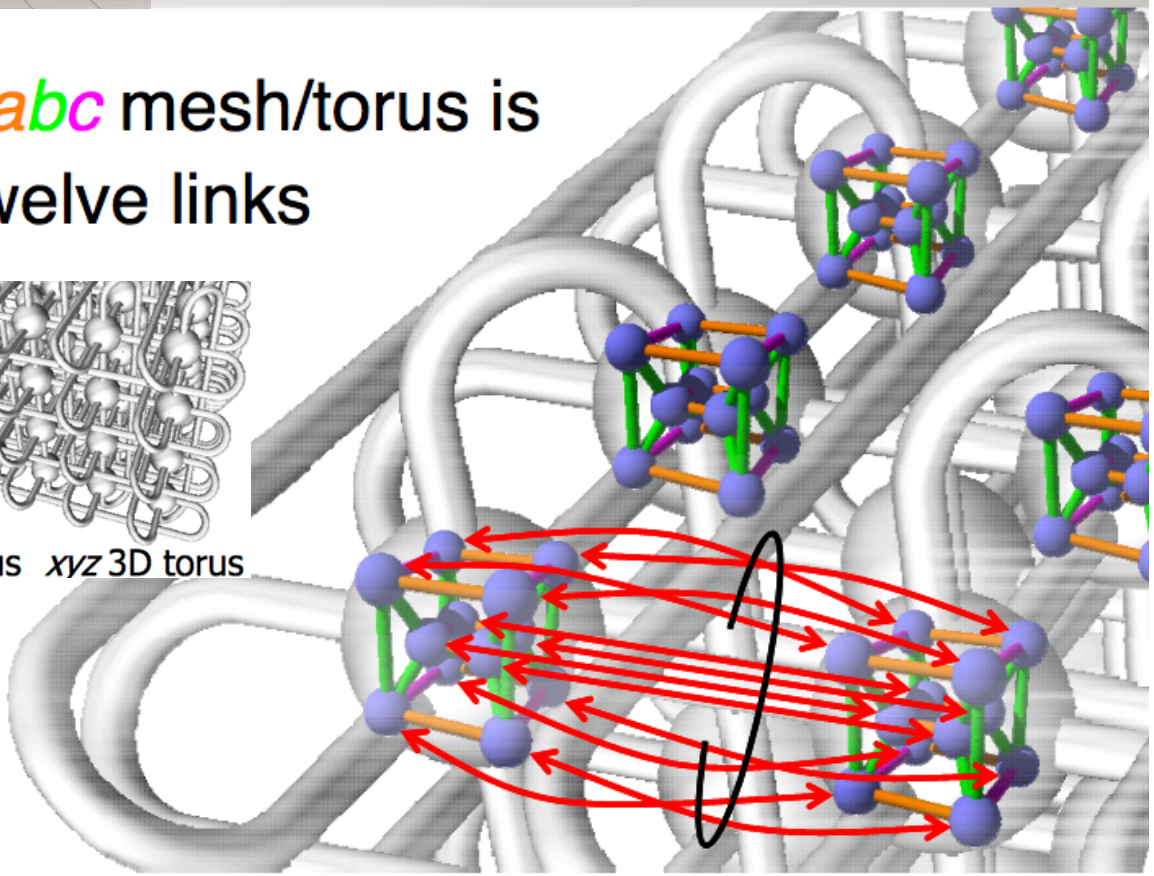


Tofu 3: 6D mesh/torus interconnect

FUJITSU

- Each pair of adjacent *abc* mesh/torus is interconnected with twelve links



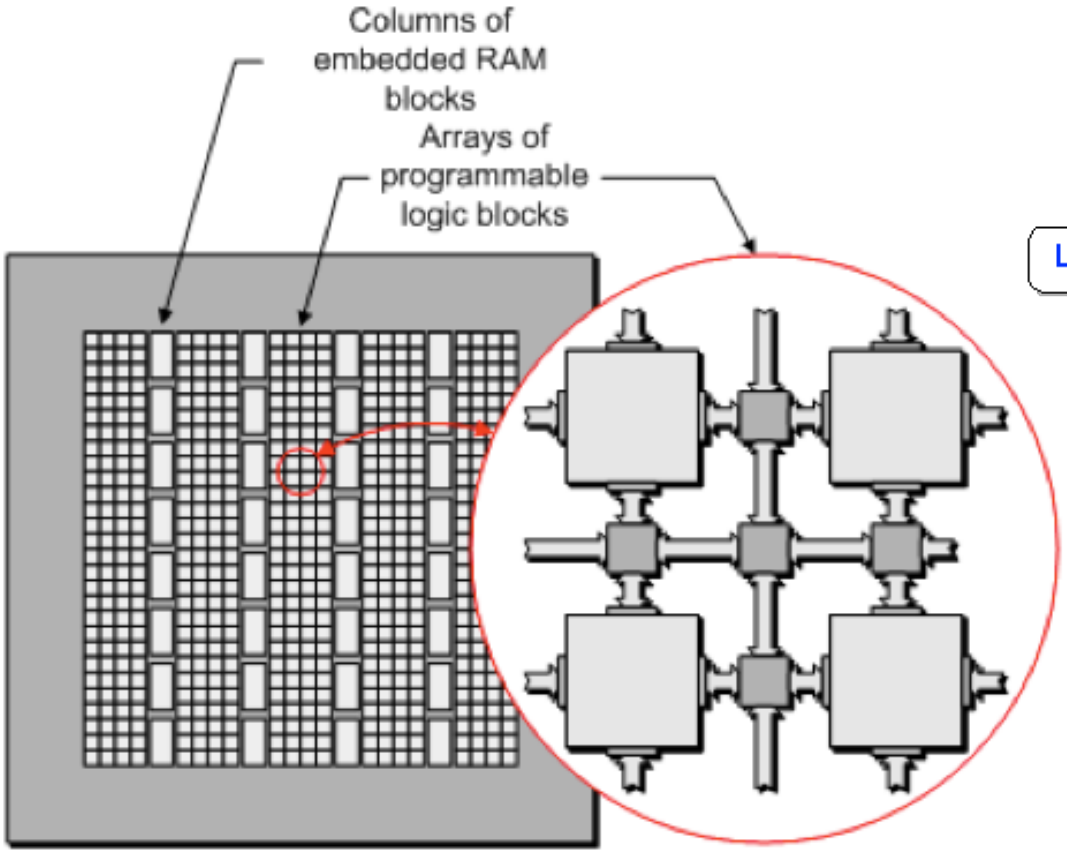


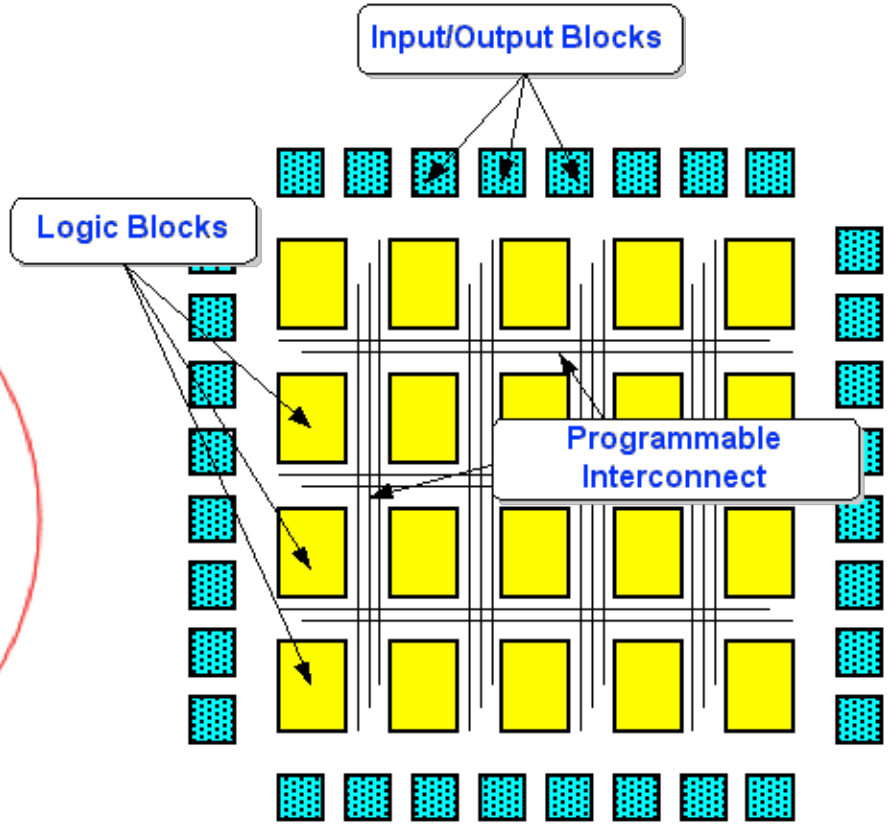
Beyond Vector/SIMD architectures

- Vector/SIMD-extended architectures are hybrid approaches
  - mix (super)**scalar + vector** op capabilities on a single device
  - **highly pipelined** approach to reduce memory access penalty
  - **tightly-closed access to shared memory**: lower latency
- Evolution of Vector/SIMD-extended architectures
  - **PU (Processing Unit) cores with wider vector units**
    - x86 many-core: **Intel** MIC / **Xeon KNL**
    - others: **IBM** BlueGene/Q Compute, **ShenWay** 260, **Matrix-2000, A64FX Arm**
  - **coprocessors (require a host scalar processor): accelerator devices**
    - on disjoint physical memories (e.g., **Xeon KNC** with PCI-Expr, **PEZY-SC**)
    - ISA-free architectures, code compiled to silica: **FPGA**
    - ...
  - **heterogeneous PUs in a SoC: multicore PUs with GPU-cores**
    - ...

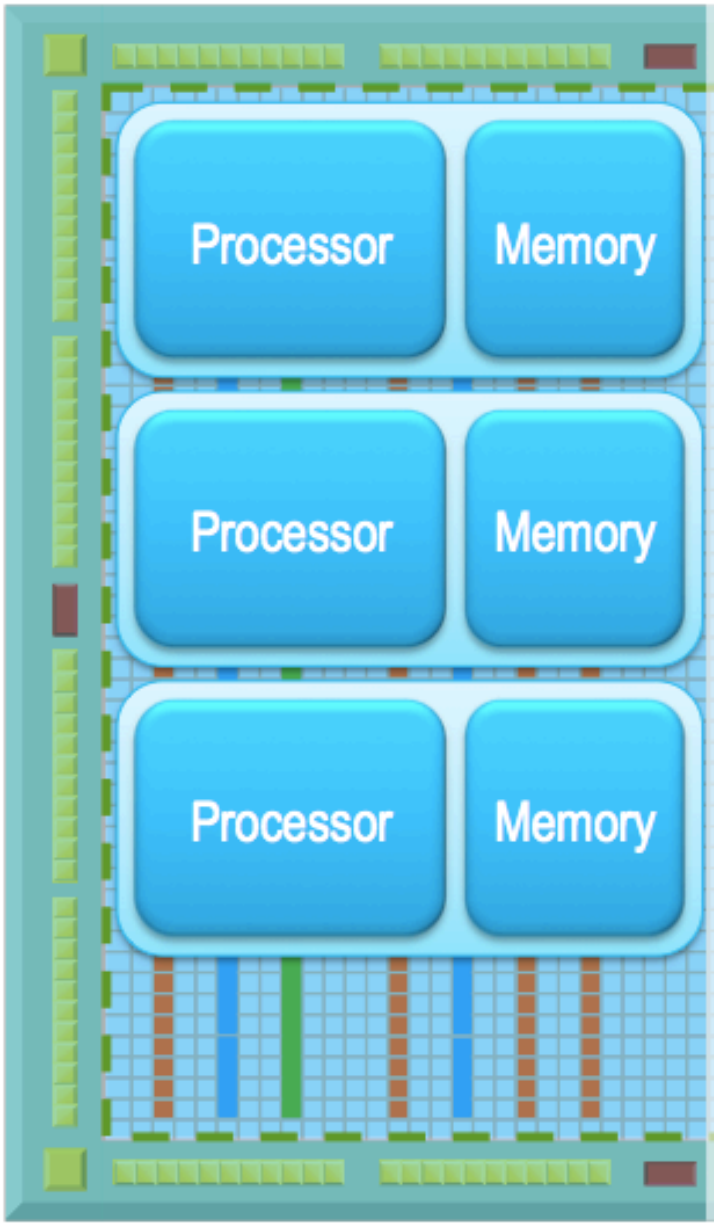
Field-Programmable Gate Arrays (FPGA)

A fabric with 1000s of simple configurable logic cells with LUTs, on-chip SRAM, configurable routing and I/O cells



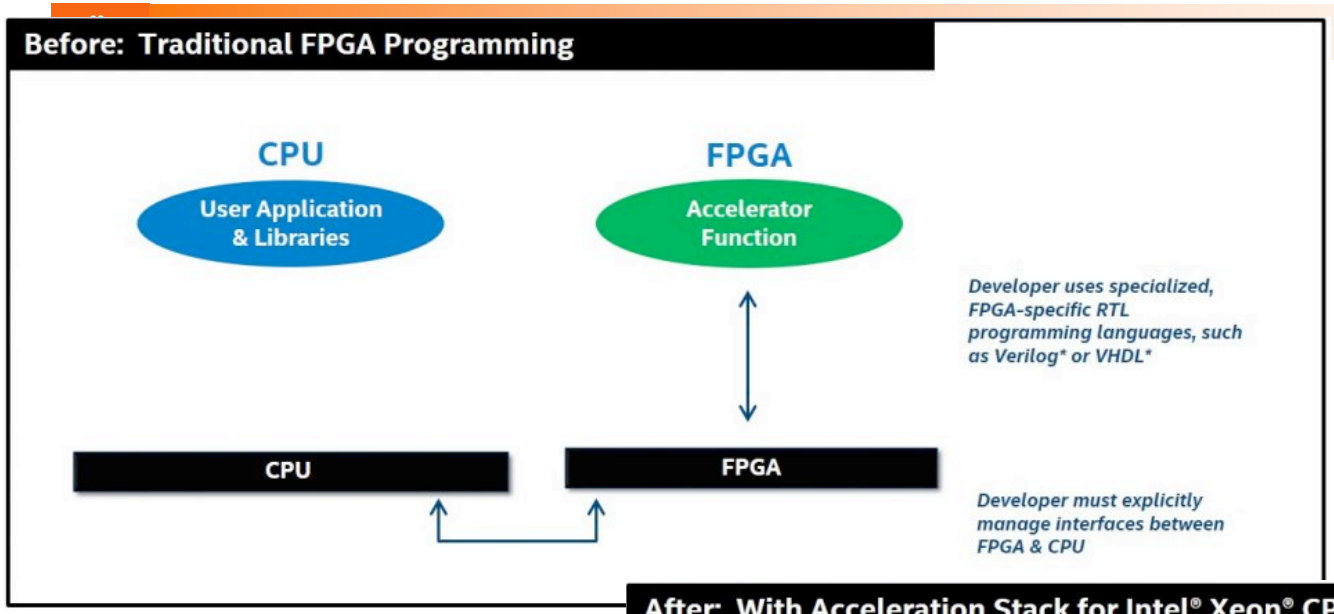


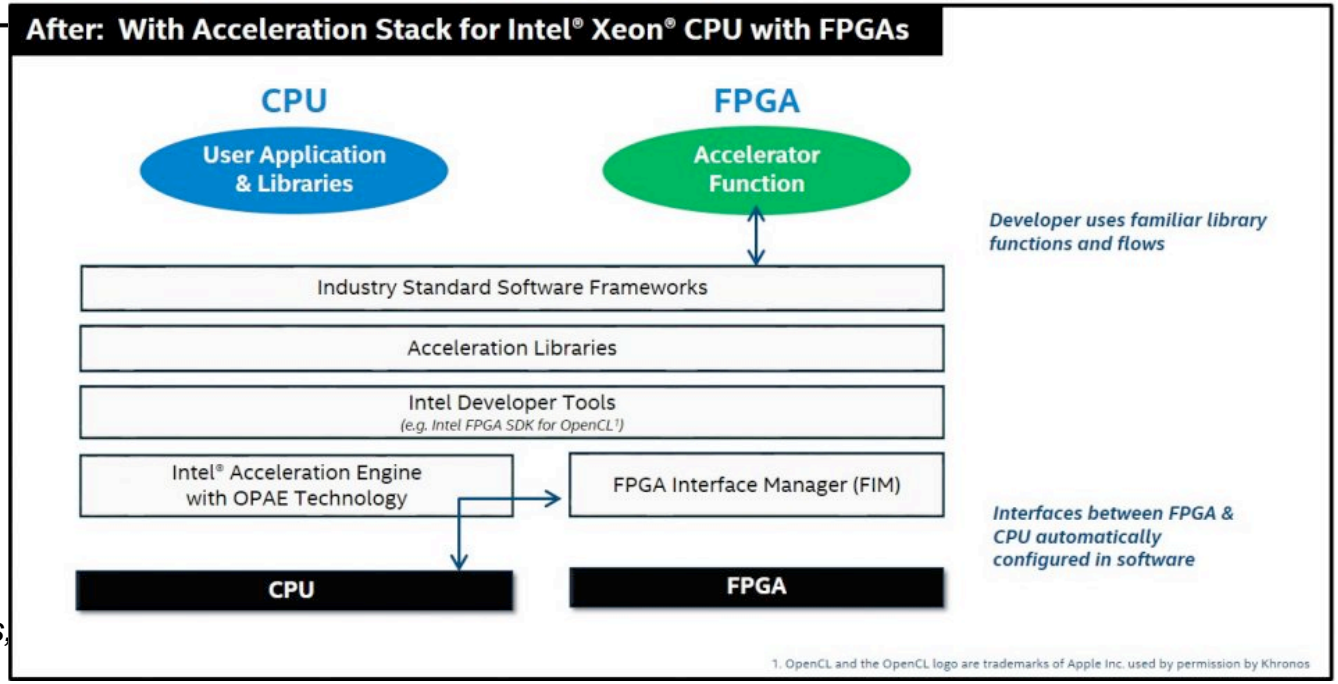
FPGA as a multiple configurable ISA



- Many coarse-grained processors
  - Different Implementation Options
    - Small soft scalar processor
    - or Larger vector processor
    - or Customized hardware pipeline
  - Each with local memory
- Each processor can exploit the fine grained parallelism of the FPGA to more efficiently implement it's "program"
- Possibly heterogeneous
  - Optimized for different tasks
- Customizable to suit the needs of a particular application

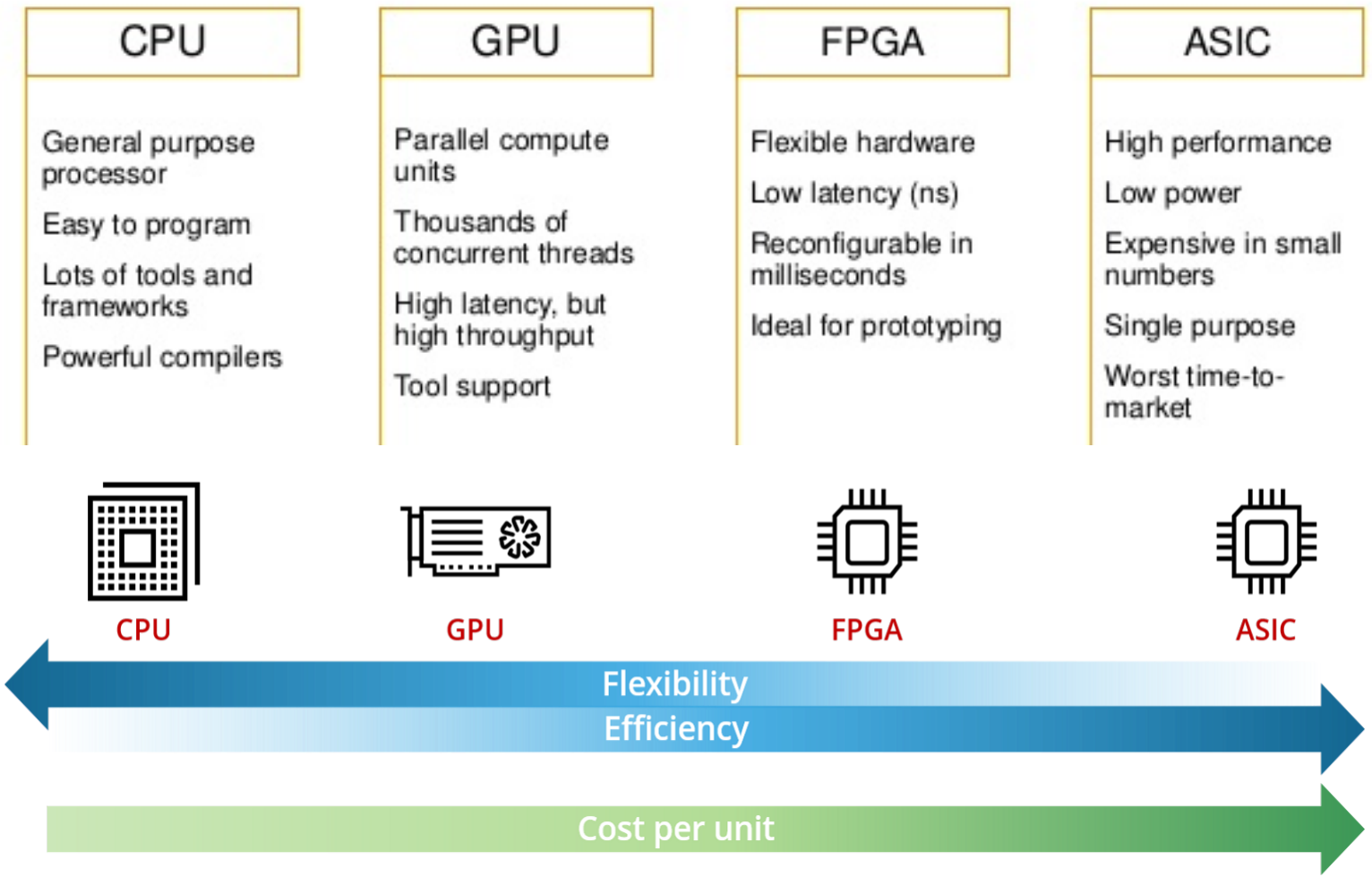
FPGA as a computing accelerator





https://builders.intel.com/blog/fpga-in-the-data-center-programming-for-all/

CPU vs. GPU vs. FPGA vs. ASIC

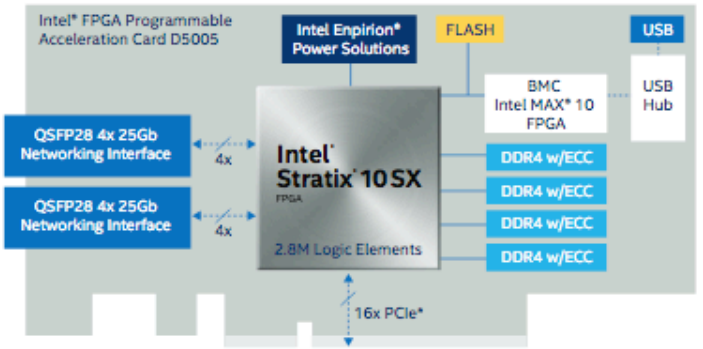


Intel FPGAs

Intel® FPGA Programmable Acceleration Card (PAC) D5005

Introduction

This high-performance FPGA acceleration card for data centers offers both inline and lookaside acceleration. Expanding upon the Intel® FPGA Programmable Acceleration Card (PAC) portfolio, it offers inline high-speed interfaces up to 100 Gbps for video transcode and streaming analytics applications. It provides the performance and versatility of FPGA acceleration and is one of several platforms supported by the Acceleration Stack for Intel Xeon® CPUs with FPGAs. This acceleration stack provides a common developer interface for both application and accelerator function developers, and includes drivers, application programming interfaces (APIs), and an FPGA interface manager. Together with acceleration libraries and development tools, the acceleration stack saves developer's time and enables code re-use across multiple Intel FPGA platforms.

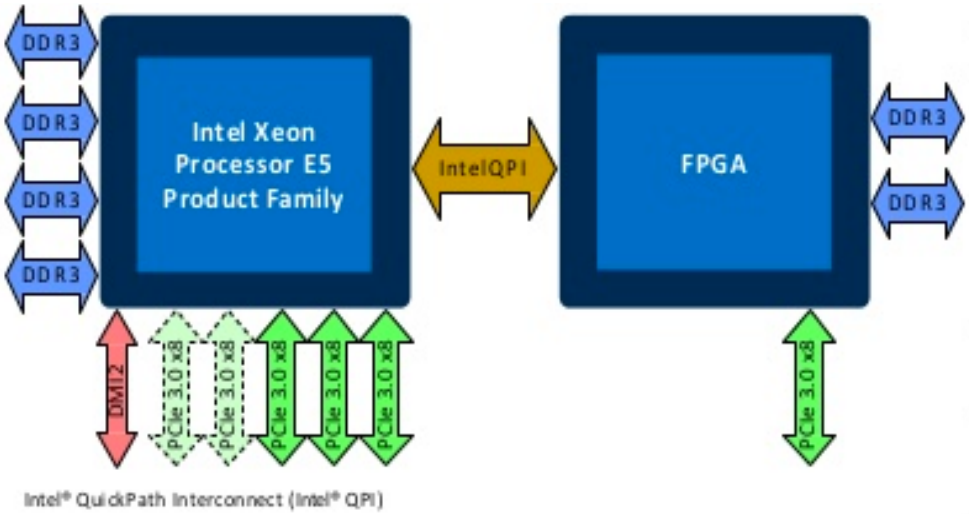


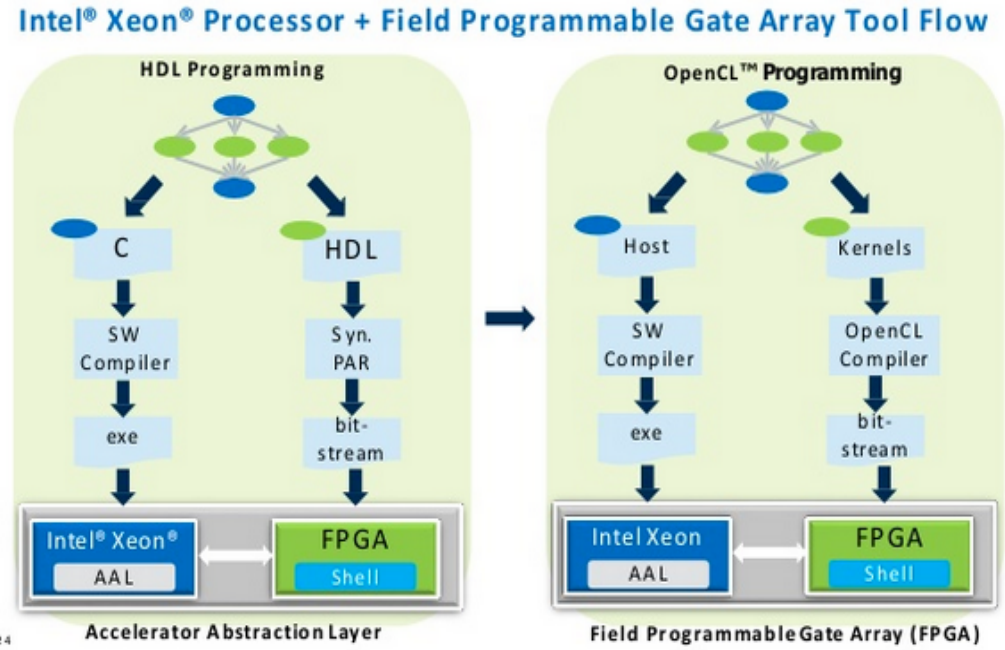
Targeted Workloads

Power and Thermals

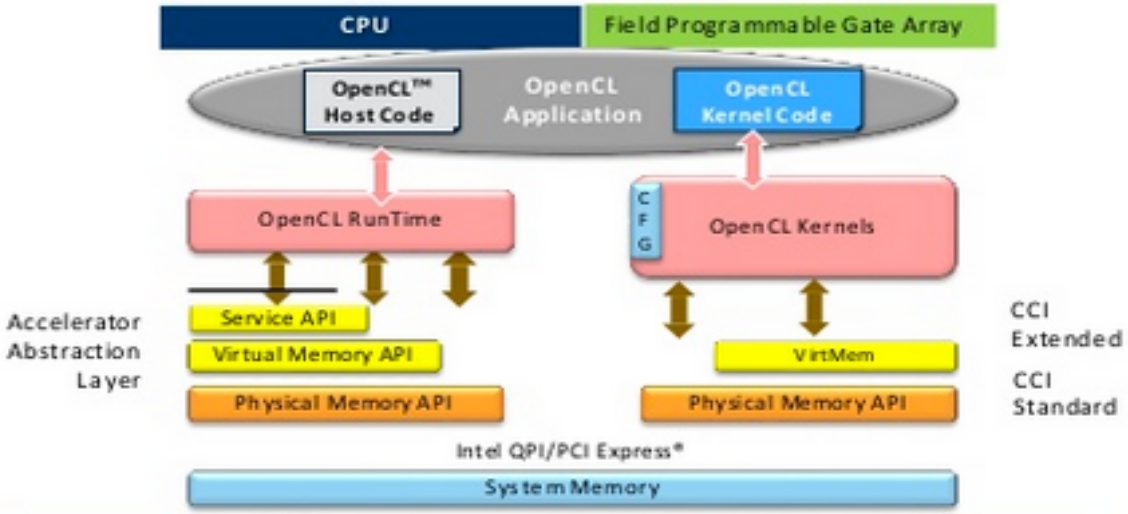
AJProença, A

Integrating programmable acceleration cards at Intel





Programming Interfaces: OpenCL



Unified application code abstracted from the hardware environment
Portable across generations and families of CPUs and FPGAs

https://www.slideshare.net/insideHPC/using-xeon-fpga-for-accelerating-hpc-workloads