A brief history of Parallelisation in classical computing

Cyril BAUDRY

Scientific Information System Architect, EDF R&D

HPC Senior expert

AQADOC

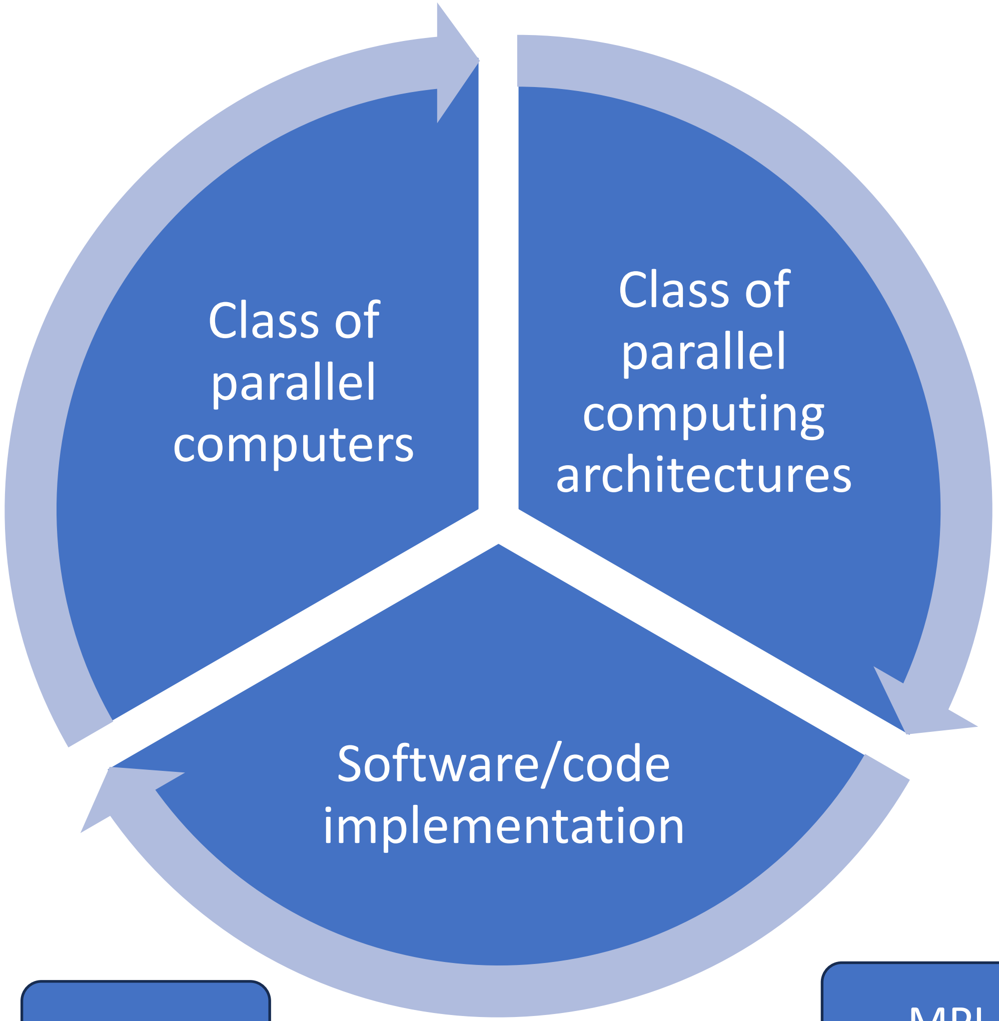
2024/02/10

What is parallel computing ?

- Divide a problem into smaller tasks and run them concurrently
- Goal : faster than a sequential computation
- To solve large, complex problems in a much shorter time.

The link between hardware and software

- ✓ Multi-Core Computing
- ✓ Symmetric Multiprocessing
- ✓ Distributed Computing
- ✓ Cluster Computing
- ✓ Massively Parallel Computing
- ✓ Grid Computing


Class of parallel computers
Class of parallel computing architectures
Software/code implementation

- ✓ Shared Memory Systems
- ✓ Distributed Memory Systems
- ✓ Hybrid Systems

kokkos

openMP

CUDA

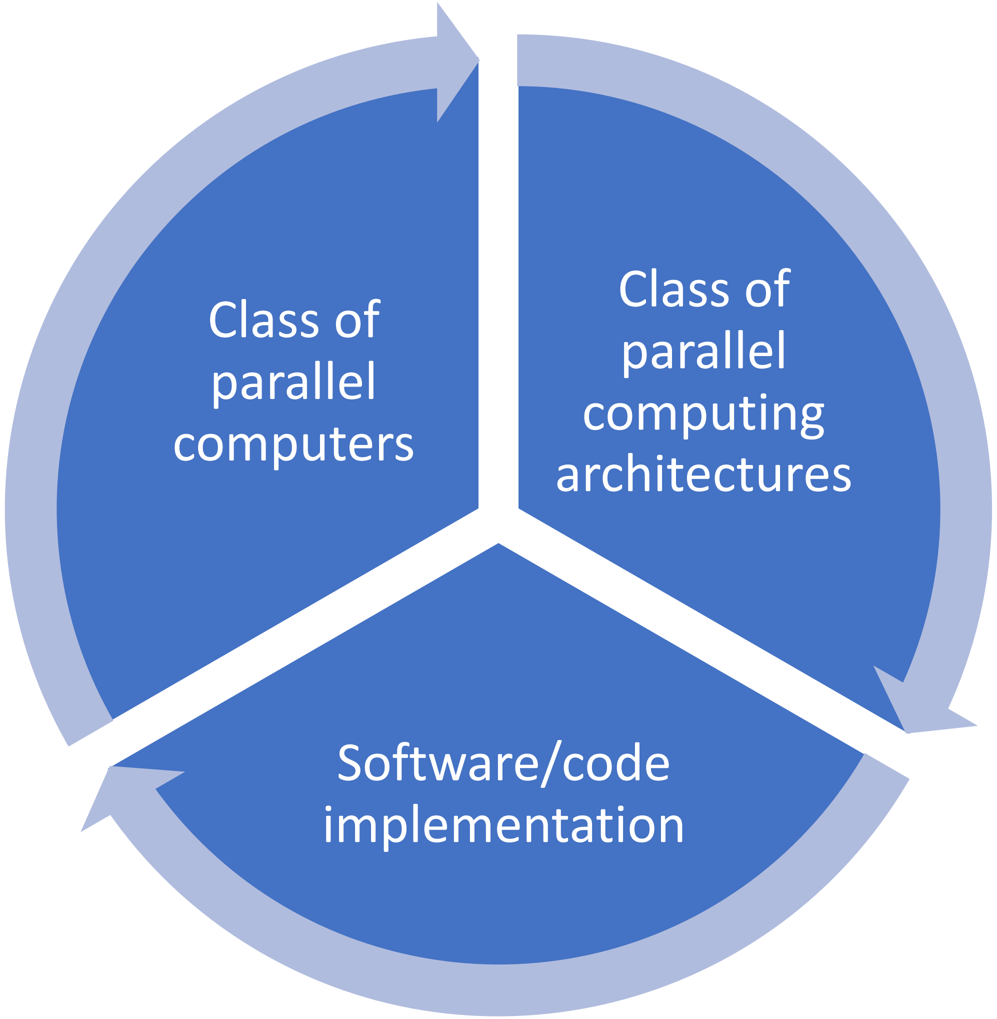
openACC

MPI

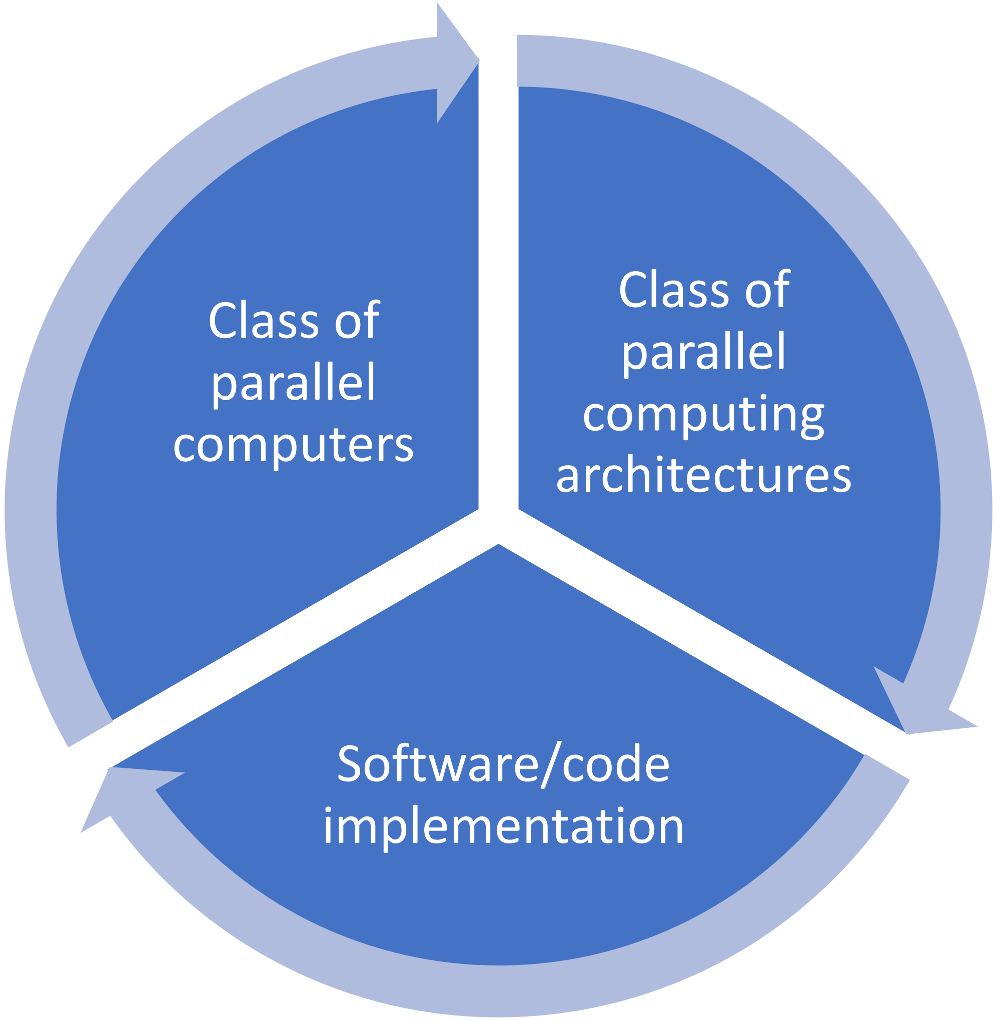
TBB

The link between hardware and software

- ✓ Multi-Core Computing: a single computing component with two or more independent processing units (cores), multitasking environments, several programs run concurrently
- ✓ Symmetric Multiprocessing (SMP): two or more identical processors are connected to a single shared main memory, efficient for multiple tasks with frequent inter-processor communication
- ✓ Distributed Computing: divided a single task into many smaller subtasks, distributed across multiple computers
- ✓ Cluster Computing: group of computers (nodes) are linked together to form a single, unified computing resource
- ✓ Massively Parallel Computing: hundreds or thousands of processors are used to perform a set of coordinated computations simultaneously
- ✓ Grid Computing: virtual supercomputer composed of networked, loosely coupled computers, used to perform large tasks


Class of parallel computers
Class of parallel computing architectures
Software/code implementation

The link between hardware and software


Class of parallel computers
Class of parallel computing architectures
Software/code implementation

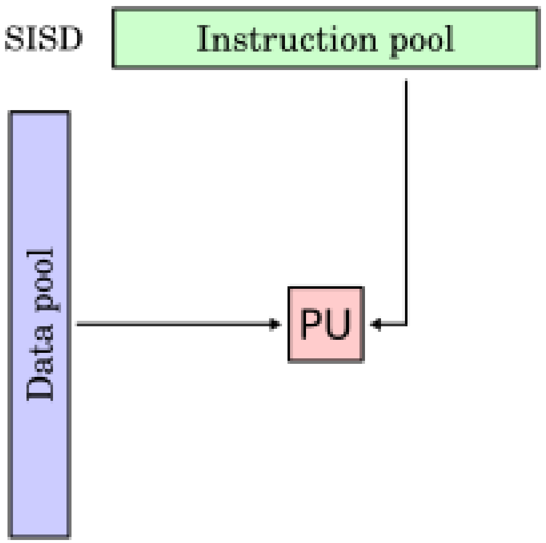
- ✓ Shared Memory:
  - ✓ multiple processors access the same physical memory
  - ✓ efficient communication between processors but scalability and memory contention: memory access can lead to bottlenecks

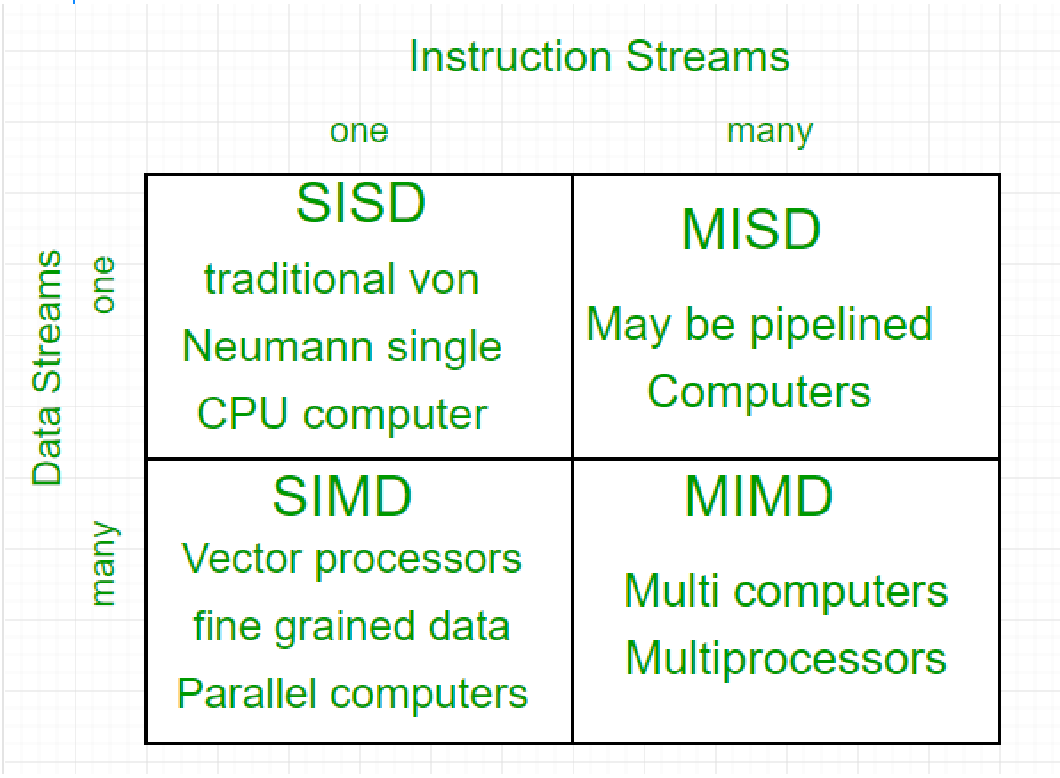
- ✓ Distributed Memory:
  - ✓ multiple processors, each with its own private memory
  - ✓ processors communicate by passing messages over a network
  - ✓ complexity of communication and synchronization: explicitly manage data distribution and message passing (MPI)

- ✓ Hybrid Memory:
  - ✓ combine elements of shared and distributed memory architectures
  - ✓ nodes use shared memory, interconnected by a distributed memory network

Flynn's taxonomy: a classification of computer architectures

- Single-instruction, single-data (SISD) systems:
  - ✓ uniprocessor machine, executing a single instruction, operating on a single data stream
  - ✓ Sequential execution
  - ✓ Ex.: workstations



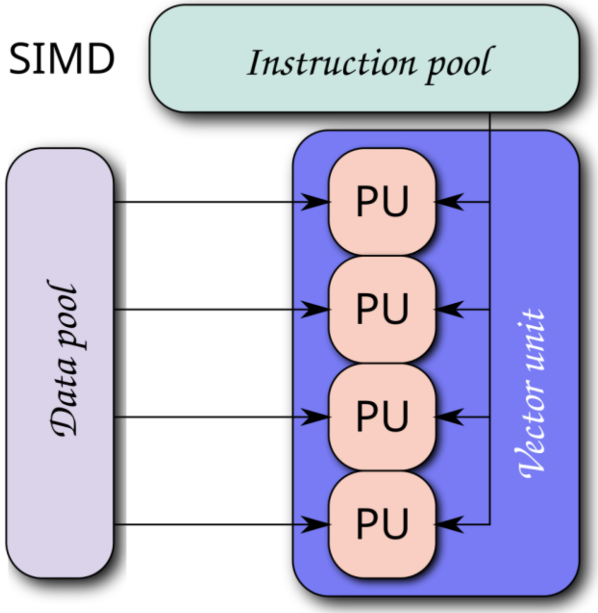


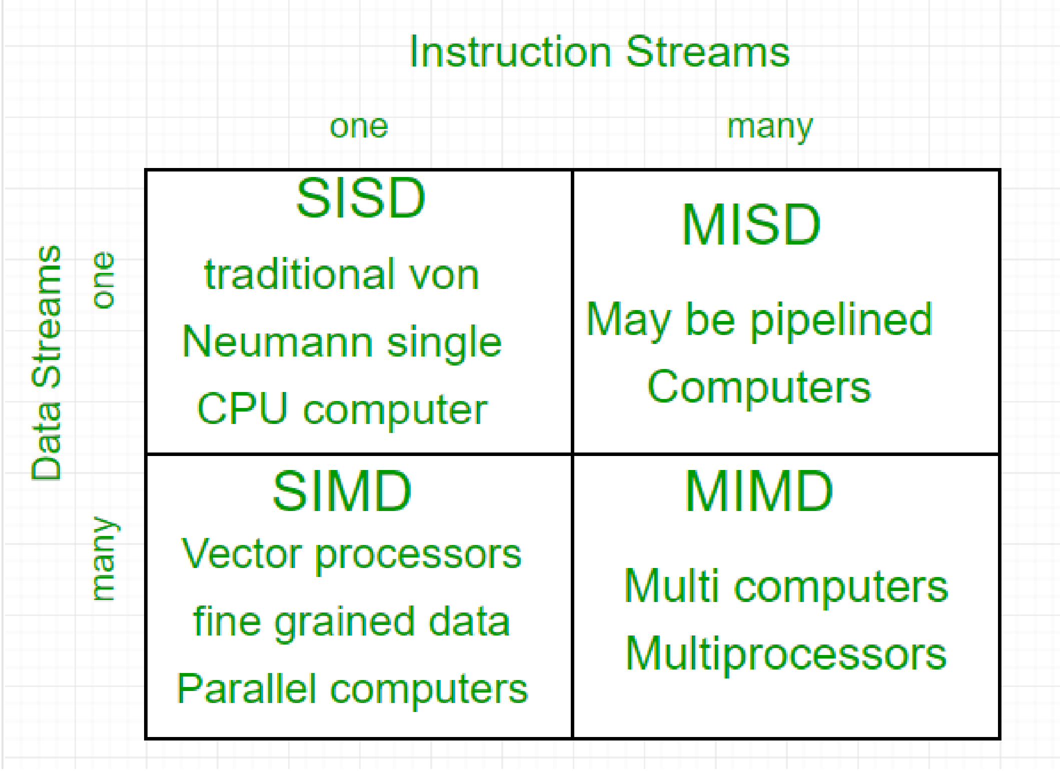
		Instruction Streams	
		one	many
Data Streams	one	SISD traditional von Neumann single CPU computer	MISD May be pipelined Computers
	many	SIMD Vector processors fine grained data Parallel computers	MIMD Multi computers Multiprocessors

source https://www.geeksforgeeks.org/computer-architecture-flynns-taxonomy/

Flynn's taxonomy: a classification of computer architectures

- Single-instruction, single-data (SISD) systems
- Single-instruction, multiple-data (SIMD) systems:
  - multiprocessor machine capable of executing the same instruction on all the CPUs but operating on different data streams
  - ex.: Cray's vector processing machine



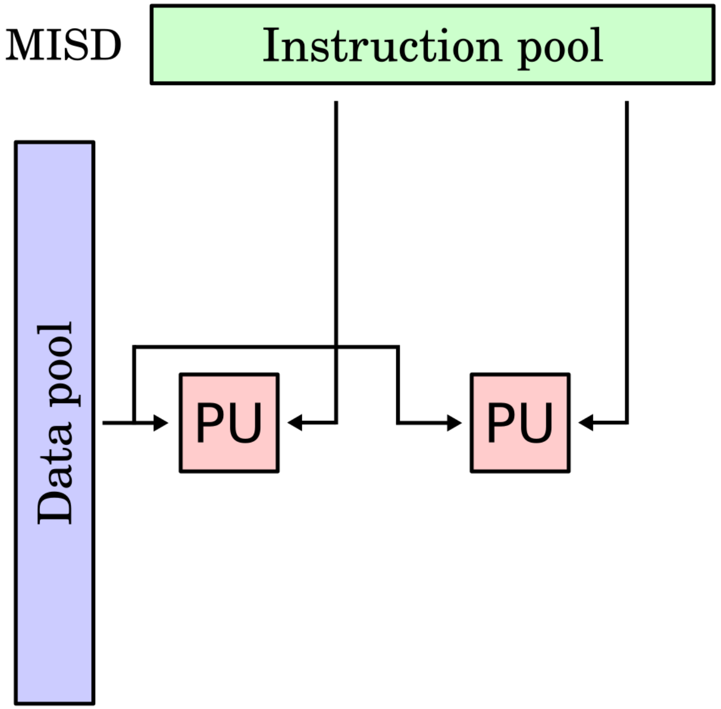


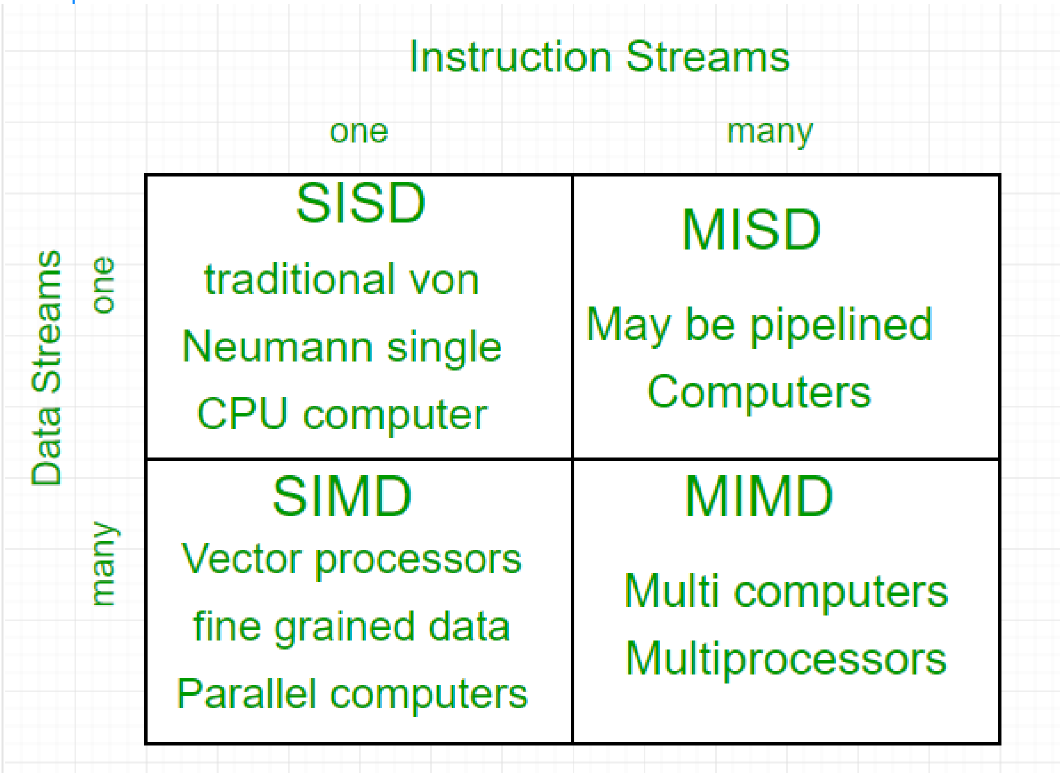
		Instruction Streams	
		one	many
Data Streams	one	SISD traditional von Neumann single CPU computer	MISD May be pipelined Computers
	many	SIMD Vector processors fine grained data Parallel computers	MIMD Multi computers Multiprocessors

source https://www.geeksforgeeks.org/computer-architecture-flynns-taxonomy/

Flynn's taxonomy: a classification of computer architectures

- Single-instruction, single-data (SISD) systems
- Single-instruction, multiple-data (SIMD) systems
- Multiple-instruction, single-data (MISD) systems:
  - multiprocessor machine capable of executing different instructions on different PUs but all of them operating on the same dataset
  - Machines built using the MISD model are not useful in most of the application, a few machines are built, but none of them are available commercially



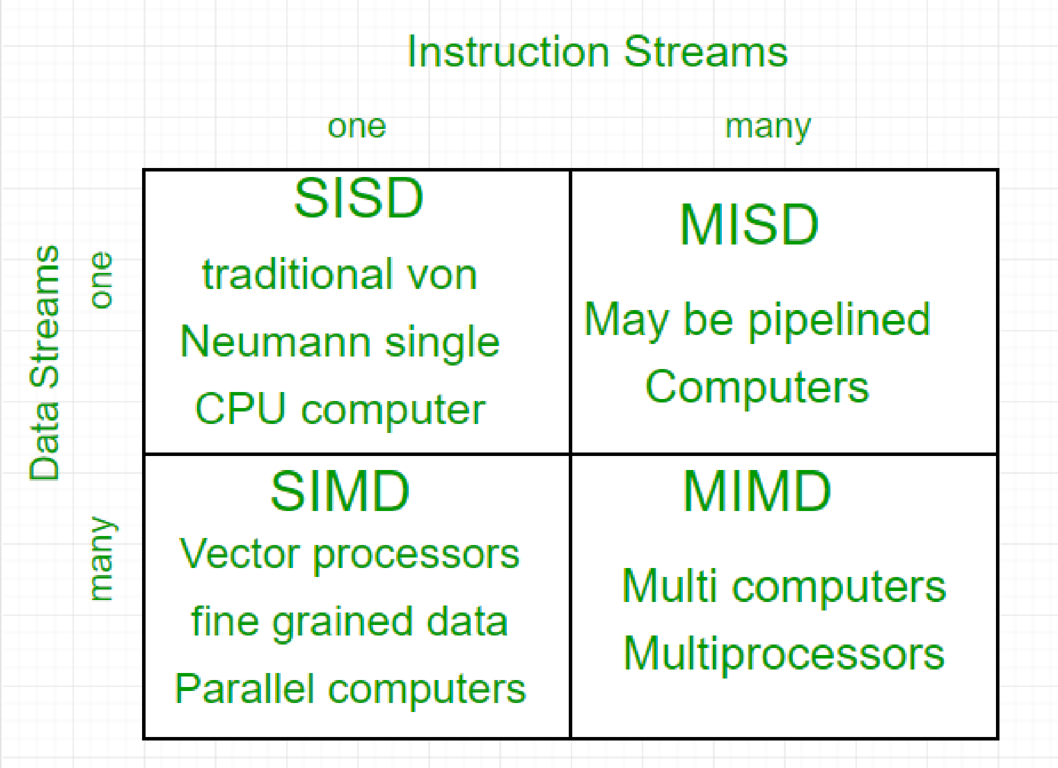


		Instruction Streams	
		one	many
Data Streams	one	SISD traditional von Neumann single CPU computer	MISD May be pipelined Computers
	many	SIMD Vector processors fine grained data Parallel computers	MIMD Multi computers Multiprocessors

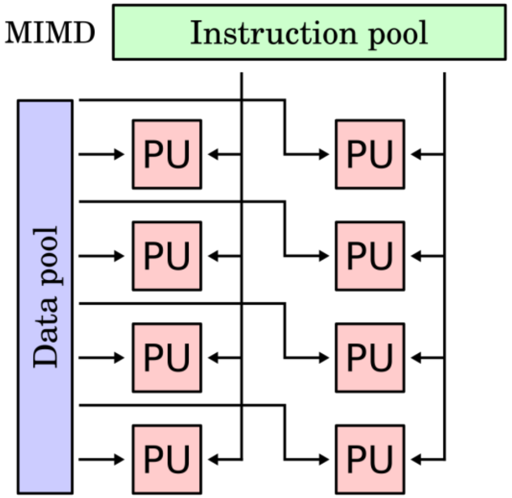
source https://www.geeksforgeeks.org/computer-architecture-flynns-taxonomy/

Flynn's taxonomy: a classification of computer architectures

- Single-instruction, single-data (SISD) systems
- Single-instruction, multiple-data (SIMD) systems
- Multiple-instruction, single-data (MISD) systems
- Multiple-instruction, multiple-data (MIMD) systems:
  - multiprocessor machine which is capable of executing multiple instructions on multiple data sets
  - Shared memory MIMD model (tightly coupled multiprocessor systems):
    - single global memory, modification of the data stored in the global memory by one PU is visible to all other PUs.
    - Ex. Silicon Graphics machines and Sun/IBM's SMP (Symmetric Multi-Processing).
  - Distributed memory MIMD machines (loosely coupled multiprocessor systems):
    - all PUs have a local memory.

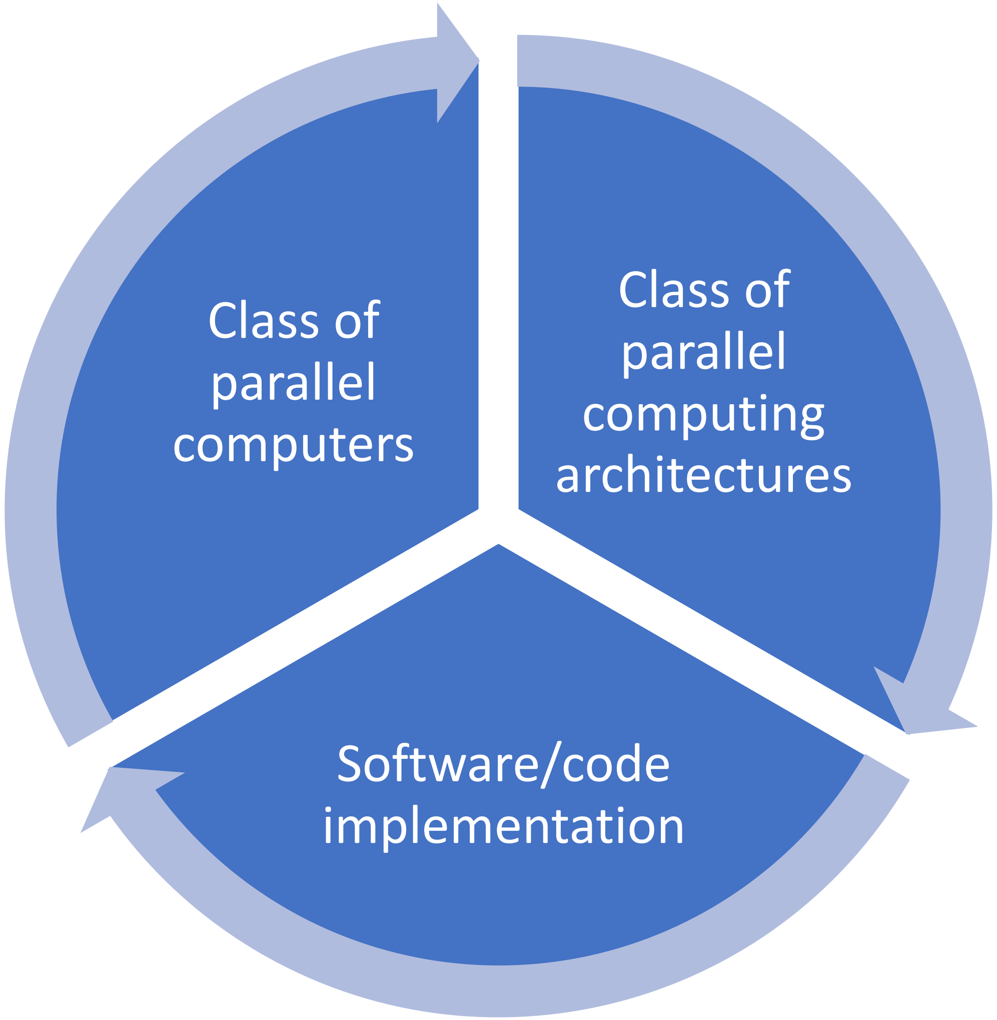


		Instruction Streams	
		one	many
Data Streams	one	SISD traditional von Neumann single CPU computer	MISD May be pipelined Computers
	many	SIMD Vector processors fine grained data Parallel computers	MIMD Multi computers Multiprocessors



source https://www.geeksforgeeks.org/computer-architecture-flynns-taxonomy/

The link between hardware and software: Parallel Computing Techniques

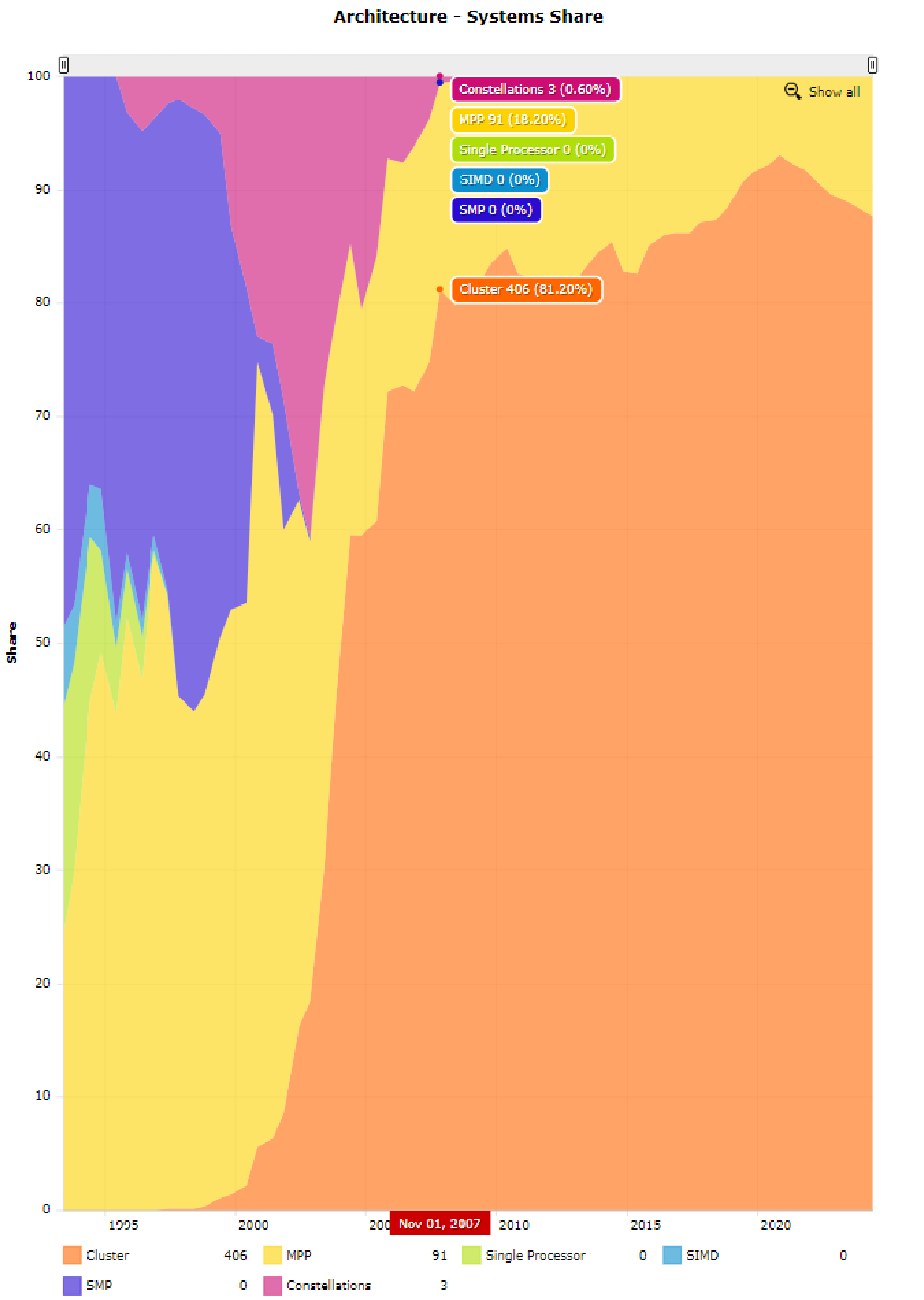

Class of parallel computers
Class of parallel computing architectures
Software/code implementation

- ✓ Bit-Level Parallelism:
  - ✓ Increase the number of bits processed in a single instruction
  - ✓ Using larger word sizes could significantly speed up computation
  - ✓ By increasing the register size, more bits can be handled simultaneously, thus increasing computational speed
  - ✓ Transparent to the programmer
  - ✓ Ex.: shift from 32-bit to 64-bit
- ✓ Instruction-Level Parallelism:
  - ✓ Execute the next instruction even before the first one has completed (pipelining)
  - ✓ Allows for the simultaneous execution of instructions
  - ✓ Limited by dependency between instructions
  - ✓ Transparent to the programmer
- ✓ Superword Level Parallelism:
  - ✓ Vectorizing operations on data stored in short vector registers
  - ✓ SIMD operations, one instruction is applied to multiple pieces of data simultaneously
- ✓ Task Parallelism:
  - ✓ Distributing tasks across different processors

History: build more power-efficient processors

- ➢ In the 1950s, 1960s and 1970s: shared memory space, run parallel operations on datasets
- ➢ In the 1990s, the ASCI Red supercomputer, using massively parallel processors (MPPs), achieved an unprecedented trillion operations per second, ushering in an era of MPP dominance in computing power

vector parallel processor : code developer use #PRAGMA directive for parallelization

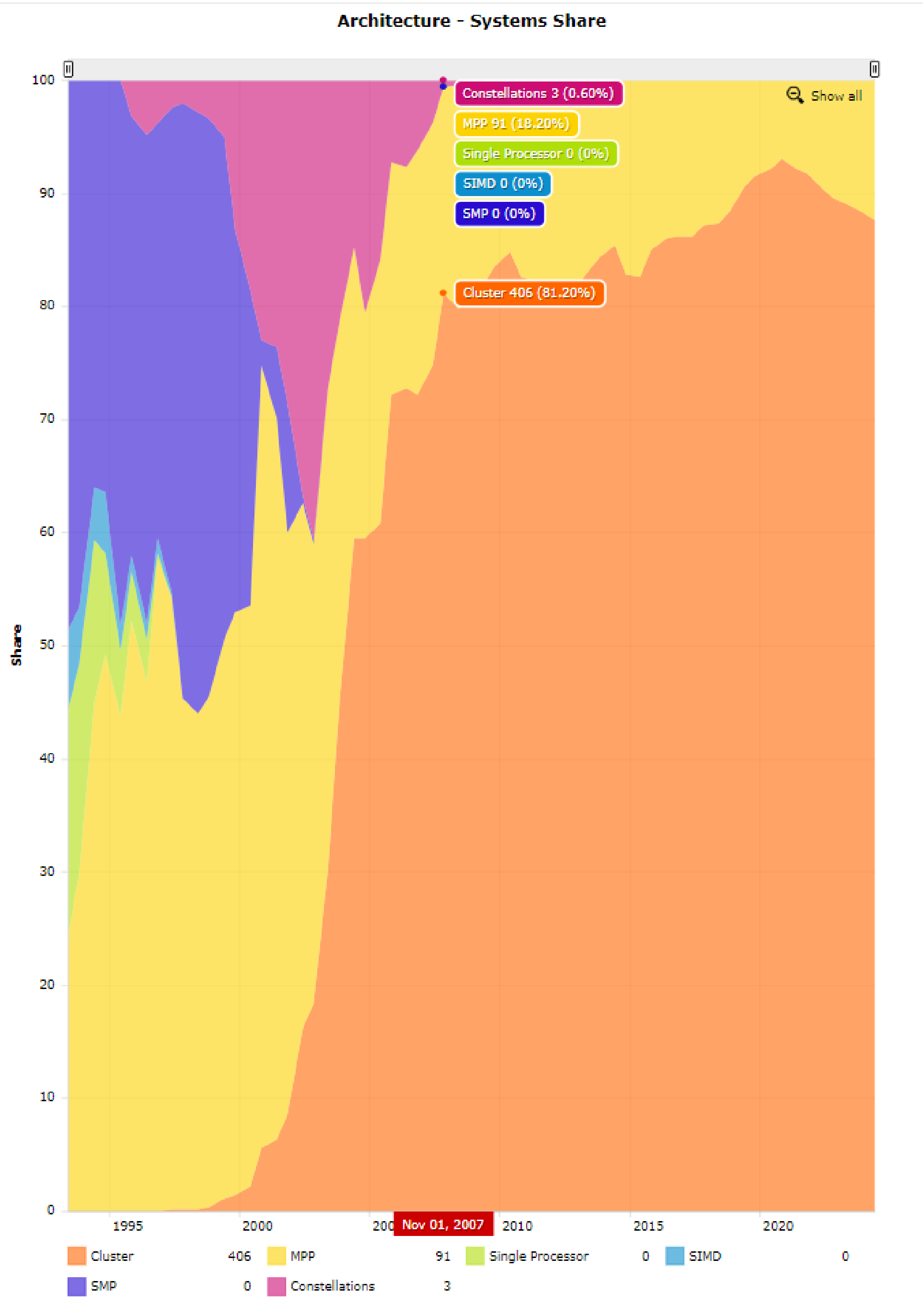


History: build more power-efficient processors

- In the 1950s, 1960s and 1970s: shared memory space, run parallel operations on datasets
- In the 1990s, the ASCI Red supercomputer, using massively parallel processors (MPPs), achieved an unprecedented trillion operations per second, ushering in an era of MPP dominance in computing power
- In the 2000s, clusters were introduced to the market

Message passing: send, receive, wait, …

- Parallel Virtual Machine (PVM) (1989-2009): software system, enables a collection of heterogeneous computers to be used as a coherent and flexible concurrent computational resource
- CORBA (1991-): enables communication between software written in different languages and running on different computers
- Message Passing Interface (openMPI, intelMPI, Cray MPICH, …)
- …



History: build more power-efficient processors

- In the 1950s, 1960s and 1970s: shared memory space, run parallel operations on datasets
- In the 1990s, the ASCI Red supercomputer, using massively parallel processors (MPPs), achieved an unprecedented trillion operations per second, ushering in an era of MPP dominance in computing power
- In the 2000s, clusters were introduced to the market
- Since 2010: mainly cluster with multi-core processors
- Recently: GPUs are often deployed concurrently with CPUs: expand data throughput and run more calculations at once

The return of #PRAGMA ?

kokkos

CUDA

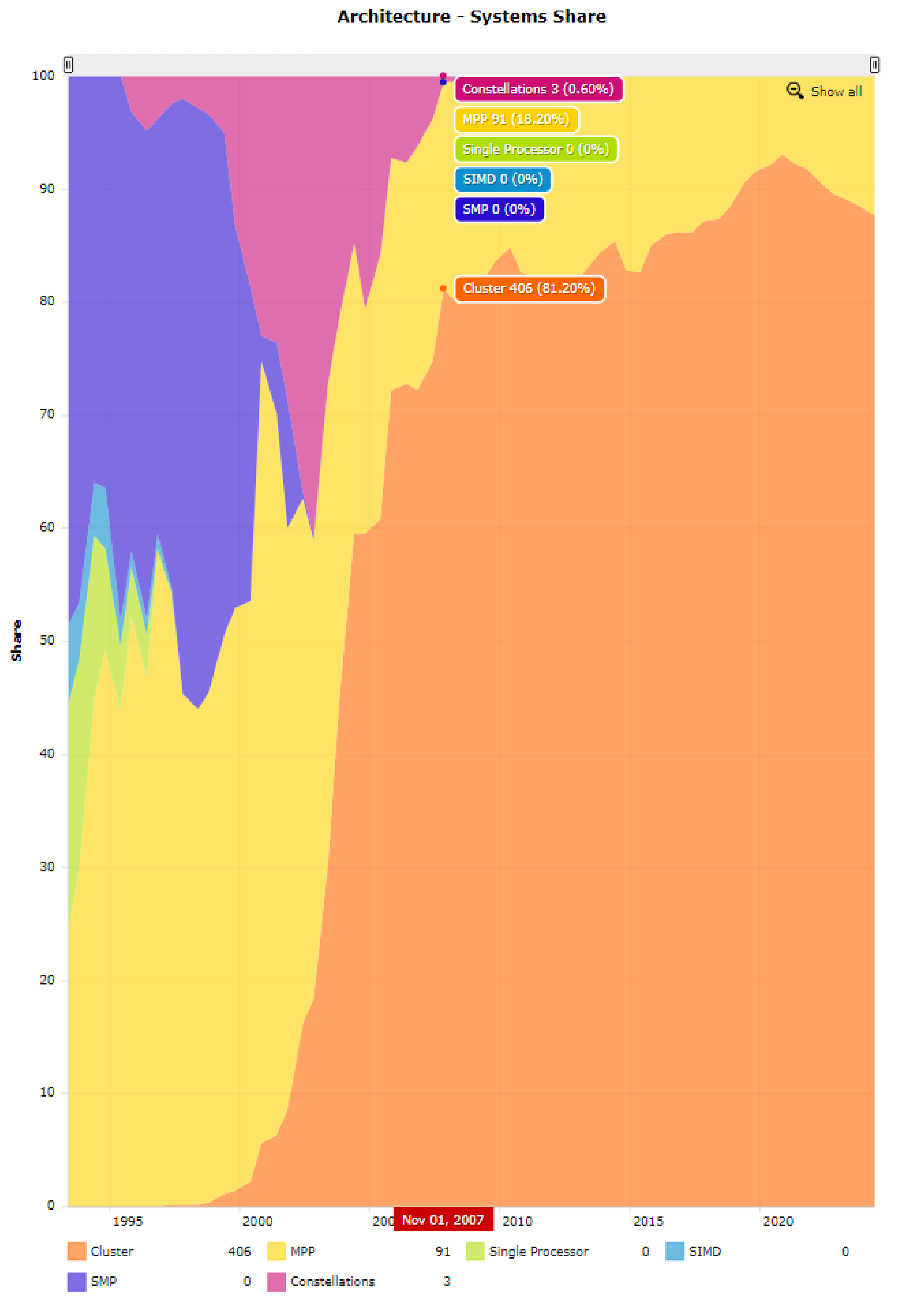
openMP

TBB

openMP +MPI

openACC

MPI

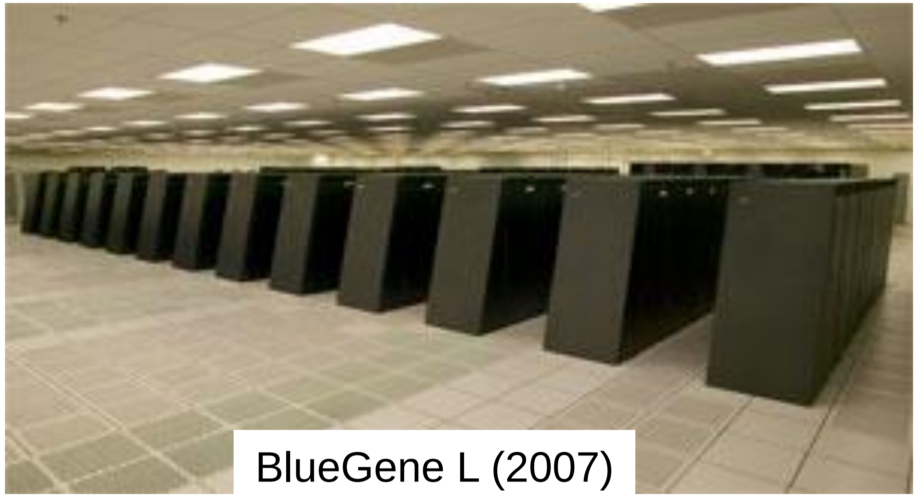


Famous HPC systems

CRAY 90s

Earth Simulator (2003)

ES : 40 TF (40.10^{12} op/s), 3,2 MW
Pentium 4 (2004, 3 GHz) : 3 GF (3.10^{9} op/s)



BlueGene L (2007)

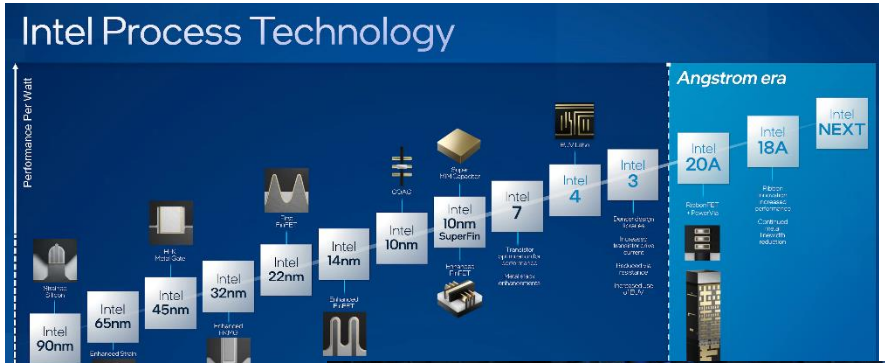
213 000 cores
478,2 TF ($478,2.10^{12}$ op/s), 2,3 MW
USA

Tianhe-2 (2013)

« rivière céleste »
3 120 000 cores
33,86 PF ($33,86.10^{15}$ op/s), 17,8 MW,Chine

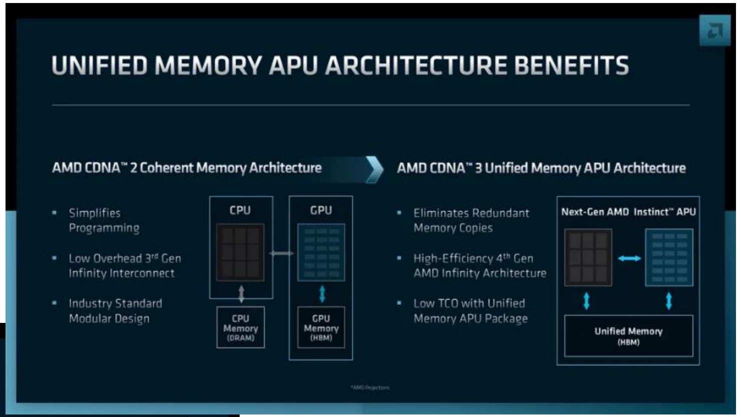
HPC today is ….. a supercomputer

➢Thousands of processors and GPGPU - to process data in parallel



(Image credit: Intel®)

intel
DATA CENTER
GPU
MAX SERIES
Intel® Data Center GPU Max Series
Up to 128 Xe Cores
52TF Peak FP64 Throughput
16 Xe Links for GPU-to-GPU communication



UNIFIED MEMORY APU ARCHITECTURE BENEFITS

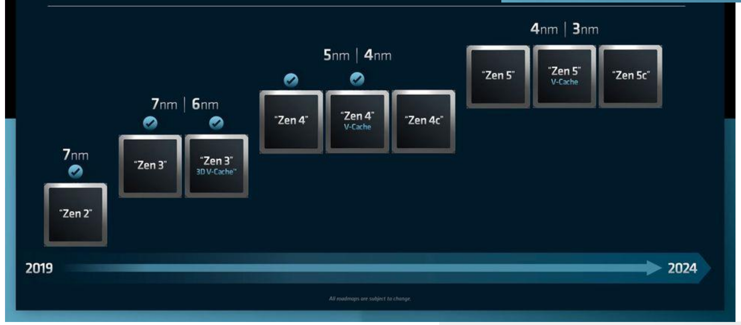
AMD CDNA™ 2 Coherent Memory Architecture

- Simplifies Programming
- Low Overhead 3rd Gen Infinity Interconnect
- Industry Standard Modular Design

AMD CDNA™ 3 Unified Memory APU Architecture

- Eliminates Redundant Memory Copies
- High-Efficiency 4th Gen AMD Infinity Architecture
- Low TCO with Unified Memory APU Package

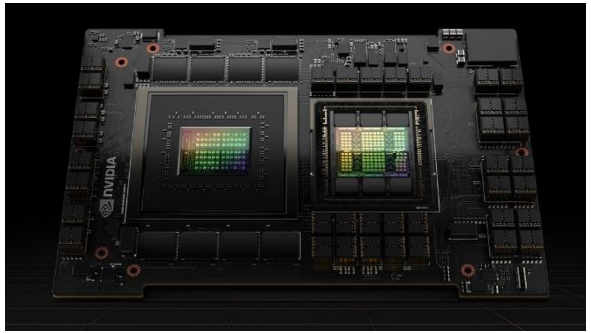
*AMD Projections



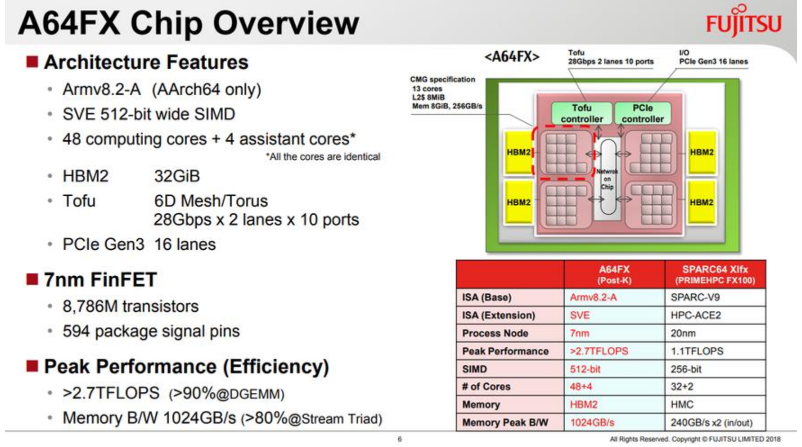
HIGH PERFORMANCE
CPU CORE ROADMAP

All roadmaps are subject to change

(Image credit: AMD®)



(Image credit: NVIDIA®)



A64FX Chip Overview

FUJITSU

■ Architecture Features

- Armv8.2-A (AArch64 only)
- SVE 512-bit wide SIMD
- 48 computing cores + 4 assistant cores*

*All the cores are identical

HBM2	32GiB
Tofu	6D Mesh/Torus 28Gbps x 2 lanes x 10 ports
PCIe Gen3	16 lanes

■ 7nm FinFET

- 8,786M transistors
- 594 package signal pins

■ Peak Performance (Efficiency)

- >2.7TFLOPS (>90%@DGEMM)
- Memory B/W 1024GB/s (>80%@Stream Triad)

	A64FX (Post-K)	SPARC64 XIfx (PRIMEHPC FX100)
ISA (Base)	Armv8.2-A	SPARC-V9
ISA (Extension)	SVE	HPC-ACE2
Process Node	7nm	20nm
Peak Performance	>2.7TFLOPS	1.1TFLOPS
SIMD	512-bit	256-bit
# of Cores	48+4	32+2
Memory	HBM2	HMC
Memory Peak B/W	1024GB/s	240GB/s x2 (in/out)

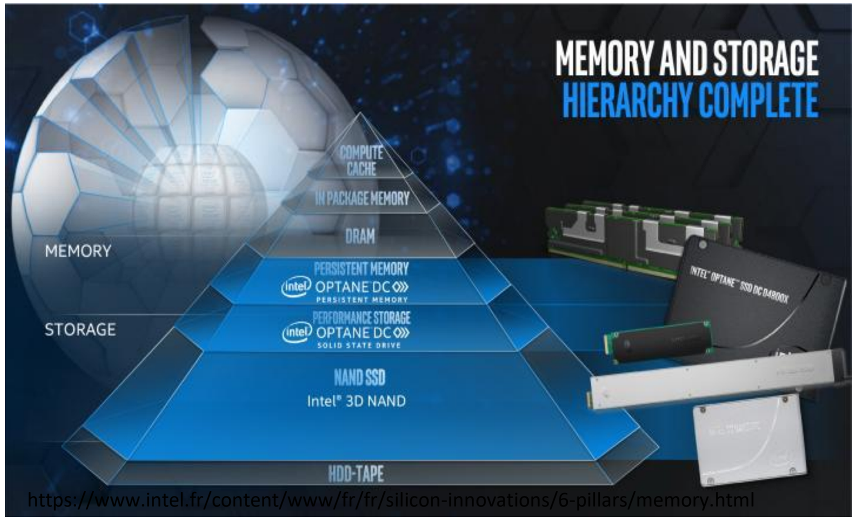
6

All Rights Reserved. Copyright © FUJITSU LIMITED 2018

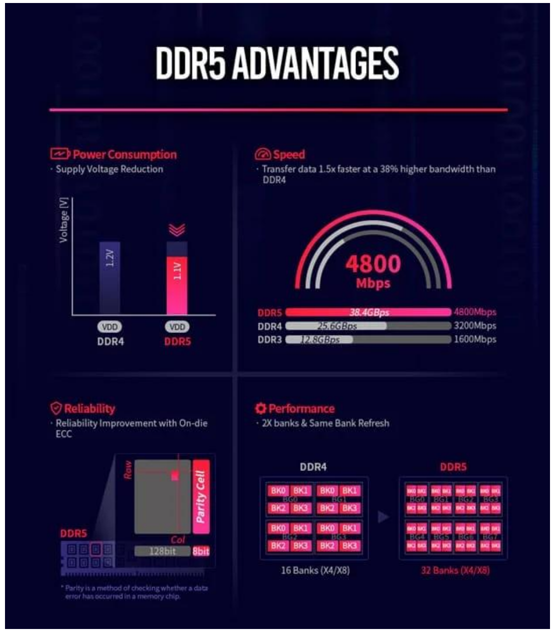
(Image credit: FUJITSU®)

HPC today is … a supercomputer

- Memory (a lot !)



https://www.intel.fr/content/www/fr/fr/silicon-innovations/6-pillars/memory.html



DDR5 ADVANTAGES

Power Consumption
· Supply Voltage Reduction

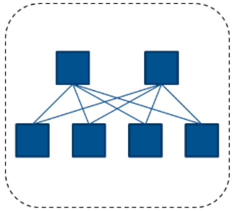
Speed
· Transfer data 1.5x faster at a 38% higher bandwidth than DDR4

Reliability
· Reliability Improvement with On-die ECC

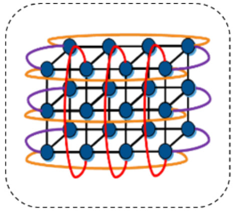
* Parity is a method of checking whether a data error has occurred in a memory chip.

Performance
· 2X banks & Same Bank Refresh

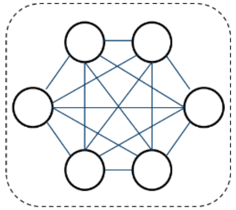
- Interconnection



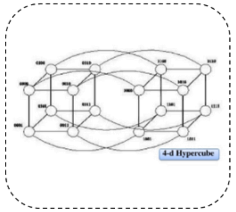
Fat Tree



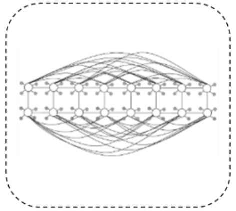
Torus



Dragonfly



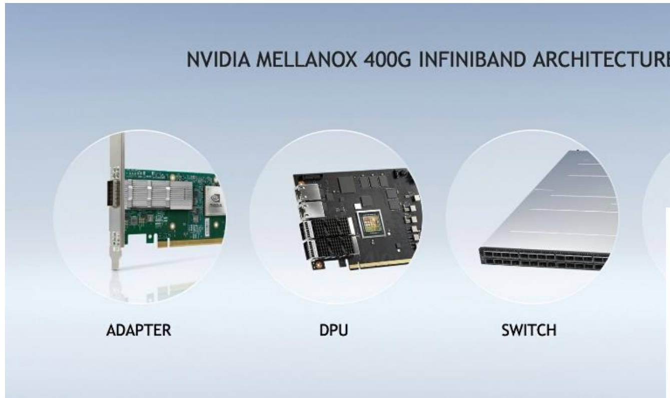
Hypercube

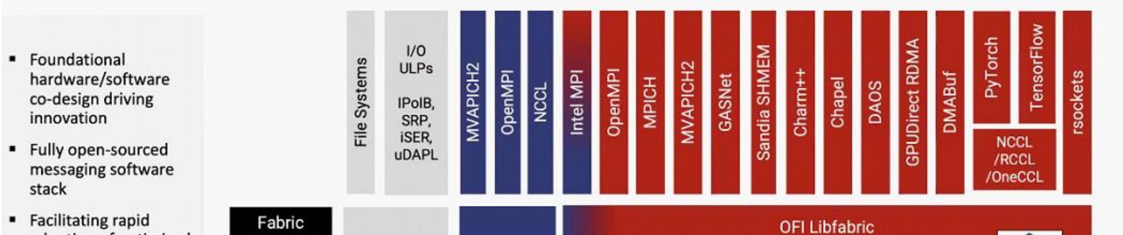


HyperX

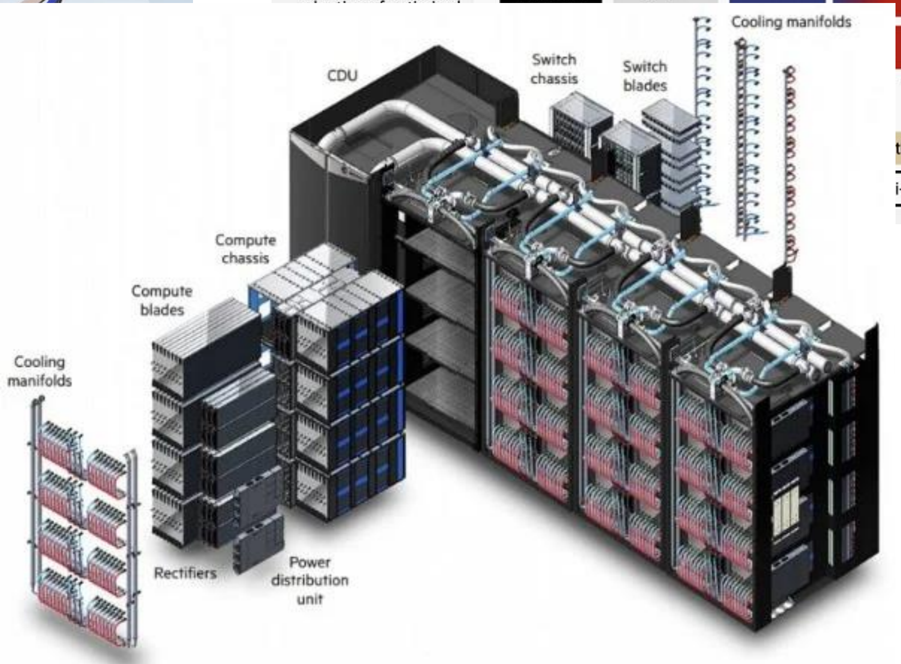
HPC today is ….. a supercomputer

- ➢Thousands of processors and GPGPU - to process data in parallel
- ➢Memory
- ➢A fast, **low-latency interconnection** network to exchange data between processors

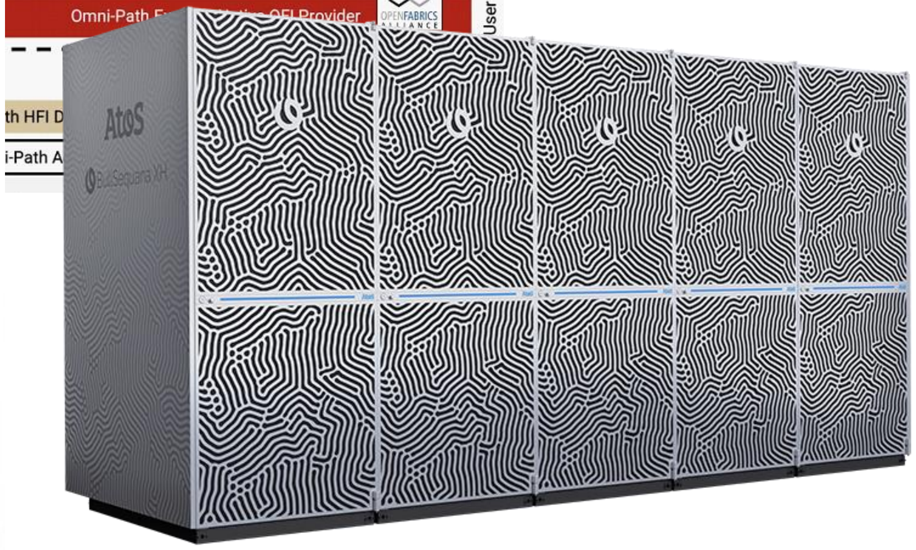




- Foundational hardware/software co-design driving innovation
- Fully open-sourced messaging software stack
- Facilitating rapid



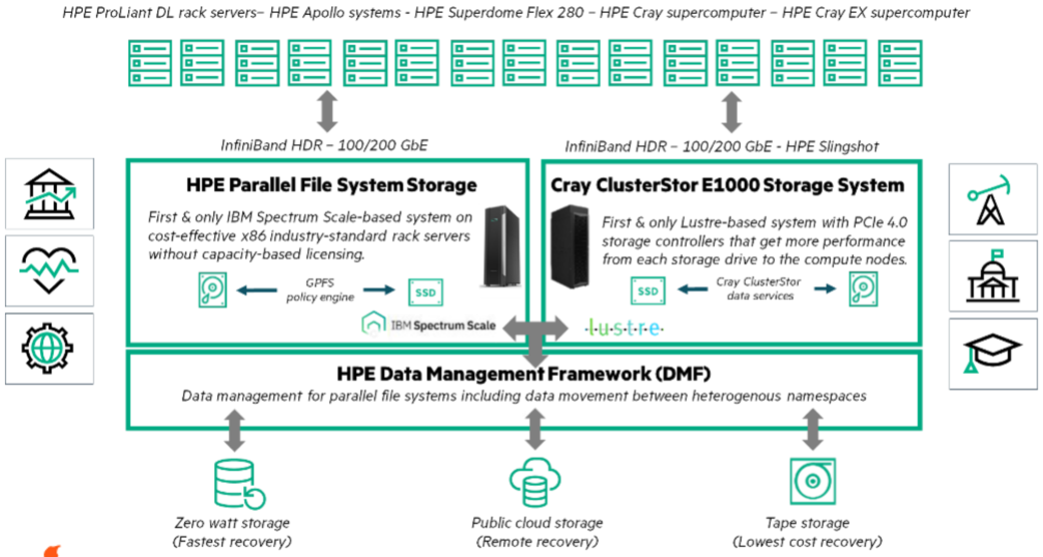
(Image credit: HPE®)



who make racks

HPC today is ….. a supercomputer

- Thousands of processors and GPGPU - to process data in parallel
- Memory
- A fast, **low-latency interconnection** network to exchange data between processors
- Parallel storage



HPE ProLiant DL rack servers– HPE Apollo systems - HPE Superdome Flex 280 – HPE Cray supercomputer – HPE Cray EX supercomputer

InfiniBand HDR – 100/200 GbE

InfiniBand HDR – 100/200 GbE - HPE Slingshot

HPE Parallel File System Storage

First & only IBM Spectrum Scale-based system on cost-effective x86 industry-standard rack servers without capacity-based licensing.

GPFS policy engine

SSD

IBM Spectrum Scale

Cray ClusterStor E1000 Storage System

First & only Lustre-based system with PCIe 4.0 storage controllers that get more performance from each storage drive to the compute nodes.

SSD

Cray ClusterStor data services

lustre

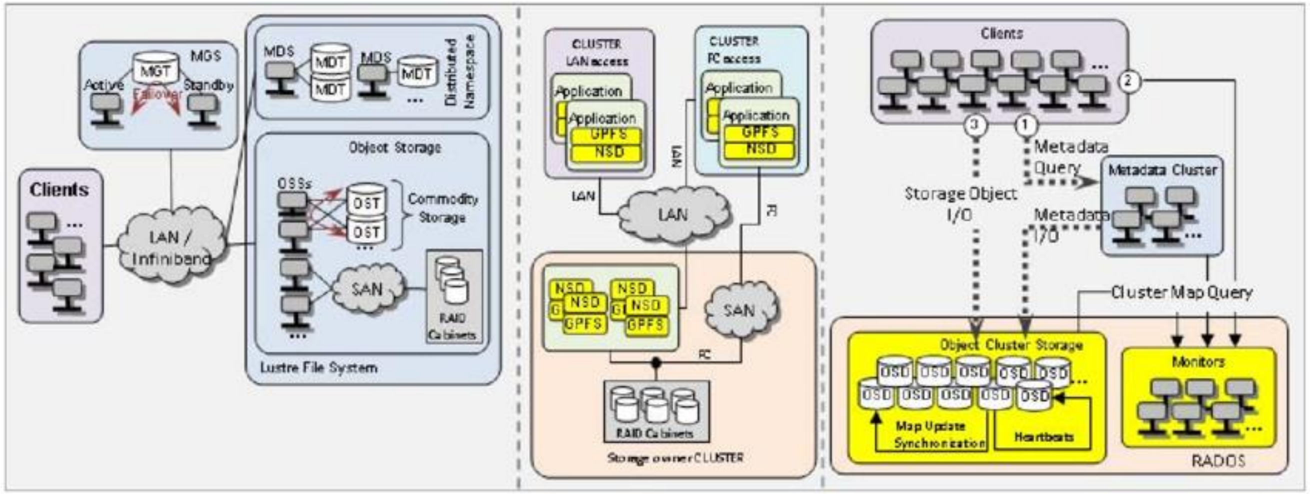
HPE Data Management Framework (DMF)

Data management for parallel file systems including data movement between heterogenous namespaces

Zero watt storage (Fastest recovery)

Public cloud storage (Remote recovery)

Tape storage (Lowest cost recovery)



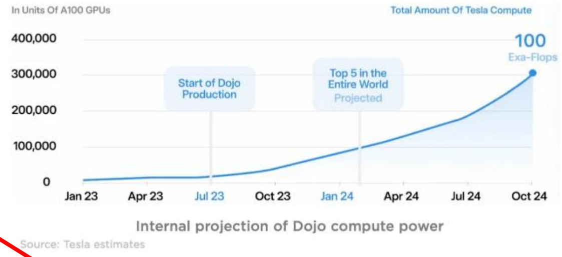
Lustre Architecture **GPFS Architecture** **Ceph Architecture**

And now ?

- 3 kinds of codes : memory bandwidth, CPU bandwidth or network bandwidth … but
  - Number of cores per CPU increase and frequency decrease (power consumption)
  - Number of memory channel per CPU increase
  - DDR5
  - New network and topology
- System are bigger and bigger and heterogeneous:
  - mix threads and MPI
  - memory synchronization between CPU and GPU
  - …
- Concurrency introduces several new classes of potential software bugs
- CPU/GPU system are very hard to debug
- Long term support and reproducibility ?

Communication and synchronization: greatest obstacles to getting optimal parallel program performance

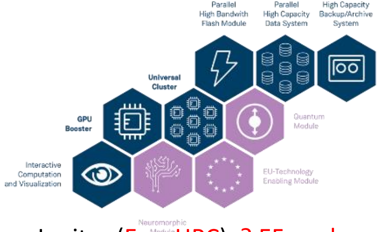
# 2, Aurora, ~~>2 EF peak~~0,6 EF, 2023, 25MW, Argonne National Laboratory, HPE Cray, ~65K, Intel GPU Max, 21K Intel Max CPU



Internal projection of Dojo compute power

Source: Tesla estimates

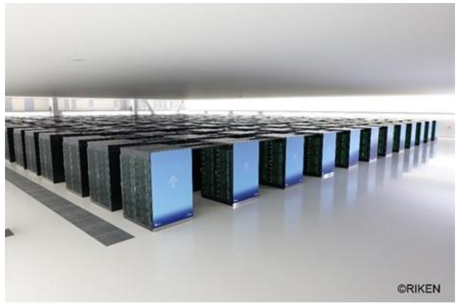
Jules Verne project (EuroHPC), Q4/2025, TGCC CEA, France


Parallel High Bandwith Flash Module
Parallel High Capacity Data System
High Capacity Backup/Archive System
Universal Cluster
GPU Booster
Quantum Module
Interactive Computation and Visualization
EU-Technology Enabling Module
Neuromorphic Module

Jupiter (EuroHPC), ? EF peak, Q4/2024, 20MW, Hybrid, Jülich Supercomputing Centre (JSC), Germany

# 5, LUMI-G (EuroHPC), ~380 PF HPL, 2023, 7 MW, CSC, Finland , HPE Cray, ~10K AMD MI250X

El Capitan, >2 EF peak, Q1/2024, ~30 MW, Lawrence Livermore National Lab HPE Cray, AMD MI300A APU



#4, Fugaku, ~440 PF HPL, 2020, 30 MW, RIKEN Center for Computational Science, Japan. Fujitsu A64FX 7,630,848 cores

# 1, Frontier, 1.2 EF HPL, 2022, 22MW, Oak Ridge National Laboratory (ORNL) HPE Cray, ~38K AMD MI250X

10 system in 2025 ?

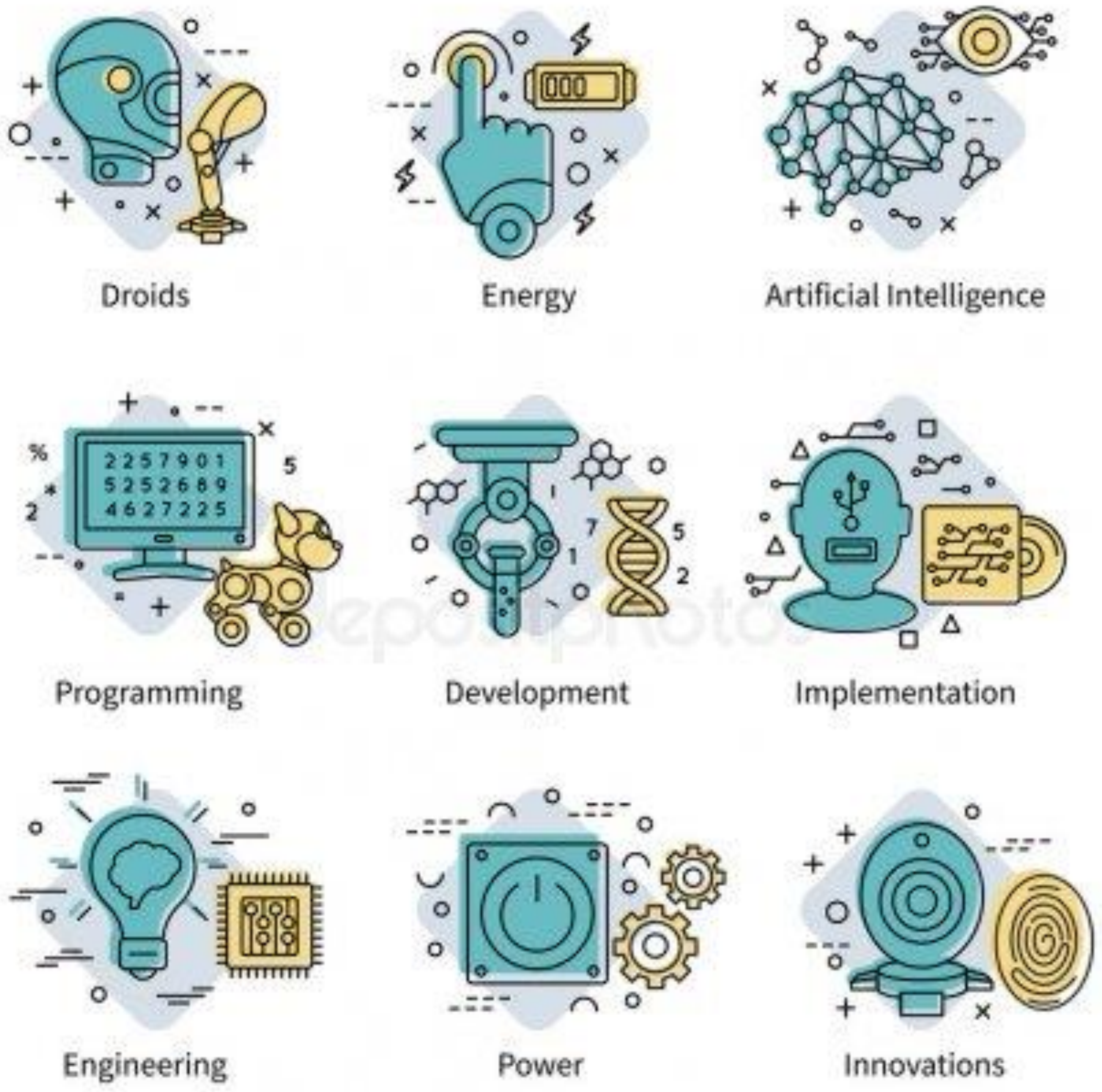
W,

# 3, Eagle, 560 PF HPL, 2023, Microsoft Azure, Microsoft, 14,4K Nvidia H100

#8, MareNostrum 5 (EuroHPC), ~140 PF HPL, 2023, 2,5 MW, BSC Spain, Atos, Nvidia H100 GPUs

#6, Leonardo, ~240 PF HPL, 2023, 7,5 MW, (EuroHPC), CINECA, Italy, Atos XH2000, ~14K Nvidia A100

Sunway-OceanLight, ~1 EF HPL, 2021, ~35 MW, ~37M cores SW26010-Pro



Thanks you for your attention