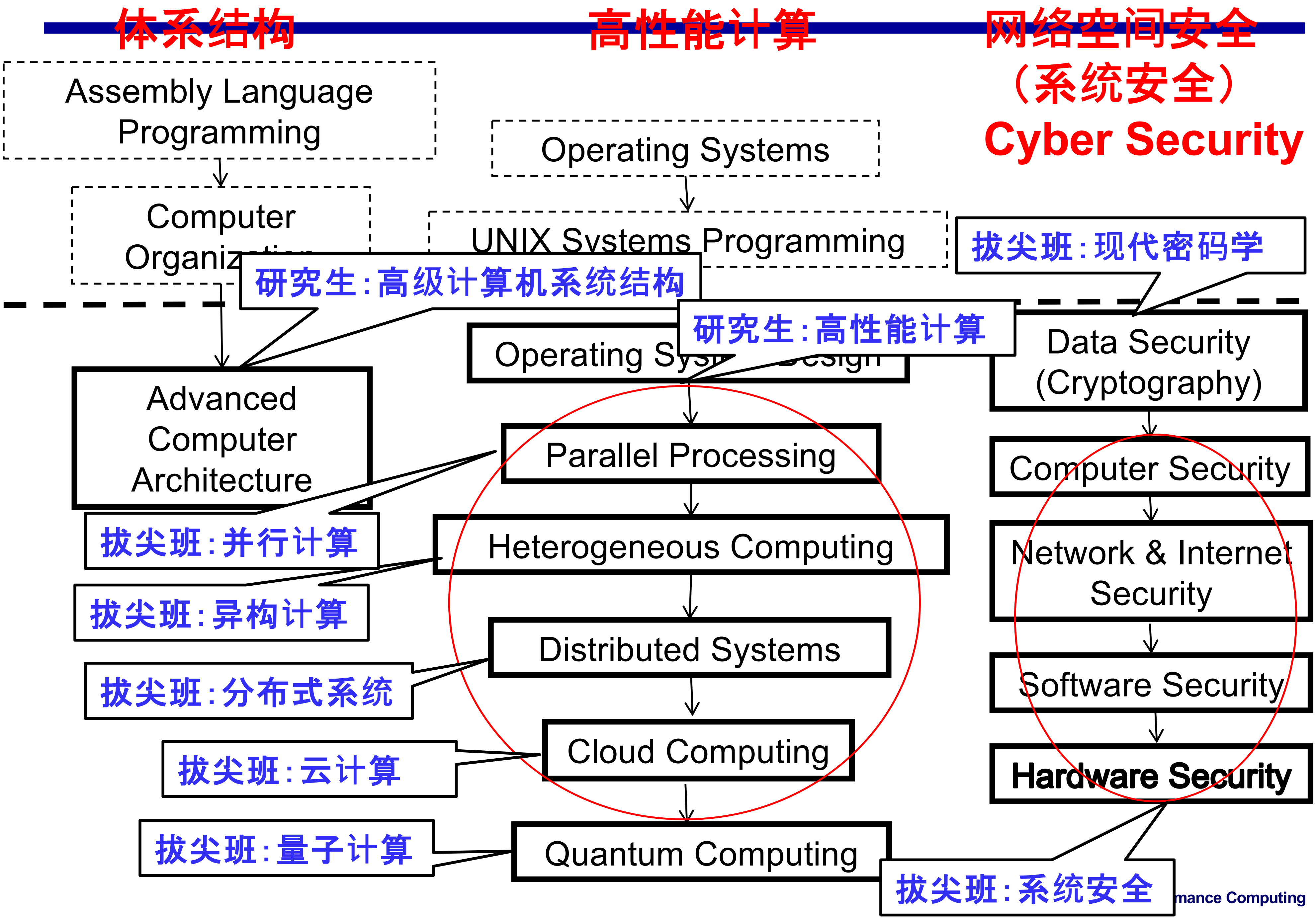
High Performance Computing
高性能计算

Hai Jiang（姜海）, Ph. D.

北京邮电大学 计算机学院 教授

Email: hai.jiang@bupt.edu.cn
办公室：科研楼807

国外教学与科研

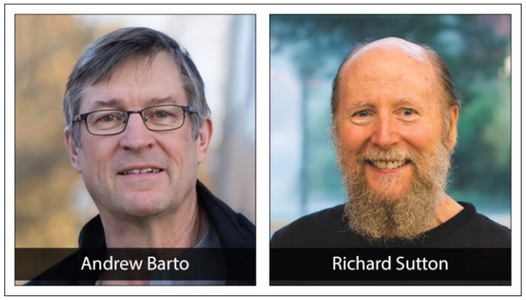


体系结构
高性能计算
网络空间安全
（系统安全）
Cyber Security
Assembly Language Programming
Computer Organization
Advanced Computer Architecture
Operating Systems
UNIX Systems Programming
Operating System Design
Parallel Processing
Heterogeneous Computing
Distributed Systems
Cloud Computing
Quantum Computing
Data Security (Cryptography)
Computer Security
Network & Internet Security
Software Security
Hardware Security
研究生:高级计算机系统结构
研究生:高性能计算
拔尖班:现代密码学
拔尖班:并行计算
拔尖班:异构计算
拔尖班:分布式系统
拔尖班:云计算
拔尖班:量子计算
拔尖班:系统安全

March 5th, 2025, Turing Award

ACM Bulletin

Andrew Barto and Richard Sutton Receive the 2024 ACM A.M. Turing Award



ACM has named Andrew G. Barto and Richard S. Sutton as the recipients of the 2024 ACM A.M. Turing Award for developing the conceptual and algorithmic foundations of reinforcement learning. In a series of papers beginning in the 1980s, Barto and Sutton introduced the main ideas, constructed the mathematical foundations, and developed important algorithms for reinforcement learning—one of the most important approaches for creating intelligent systems.

Barto is Professor Emeritus of Information and Computer Sciences at the University of Massachusetts, Amherst. Sutton is a Professor of Computer Science at the University of Alberta, a Research Scientist at Keen Technologies, and a Fellow at Amii (Alberta Machine Intelligence Institute).

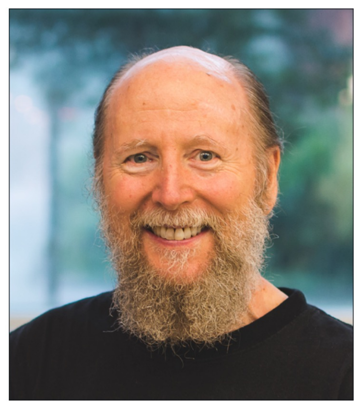
The ACM A.M. Turing Award, often referred to as the “Nobel Prize in Computing,” carries a $1 million prize with financial support provided by Google, Inc. The award is named for Alan M. Turing, the British mathematician who articulated the mathematical foundations of computing.

“Barto and Sutton’s work demonstrates the immense potential of applying a multidisciplinary approach to longstanding challenges in our field,” explains ACM President Yannis Ioannidis. “Research areas ranging from cognitive science and psychology to neuroscience inspired the development of reinforcement learning, which has laid the foundations for some of the most important advances in AI and has given us greater insight into how the brain works. Barto and Sutton’s work is not a stepping stone that we have now moved on from. Reinforcement learning continues to grow and offers great potential for further advances in computing and many other disciplines. It is fitting that we are honoring them with the most prestigious award in our field.”

2024 ACM A.M. Turing Award Laureates

Andrew Barto is Professor Emeritus, Department of Information and Computer Sciences, University of Massachusetts, Amherst. He began his career at UMass Amherst as a postdoctoral Research Associate in 1977, and has subsequently held various positions including Associate Professor, Professor, and Department Chair. Barto received a BS degree in Mathematics (with distinction) from the University of Michigan, where he also earned his MS and PhD degrees in Computer and Communication Sciences.

Barto's honors include the UMass Neurosciences Lifetime Achievement Award, the IJCAI Award for Research Excellence, and the IEEE Neural Network Society Pioneer Award. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), and a Fellow of the American Association for the Advancement of Science (AAAS).



Richard Sutton is a Professor in Computing Science at the University of Alberta, a Research Scientist at Keen Technologies (an artificial general intelligence company based in Dallas, Texas) and Chief Scientific Advisor of the Alberta Machine Intelligence Institute (Amii). Sutton was a Distinguished Research Scientist at Deep Mind from 2017 to 2023. Prior to joining the University of Alberta, he served as a Principal Technical Staff Member in the Artificial Intelligence Department at the AT&T Shannon Laboratory in Florham Park, New Jersey, from 1998 to 2002. Sutton's collaborations with Andrew Barto began in 1978 at the University of Massachusetts at Amherst, where Barto was Sutton's PhD and postdoctoral advisor. Sutton received his BA in Psychology from Stanford University and earned his MS and PhD degrees in Computer and Information Science from the University of Massachusetts at Amherst.

Sutton's honors include receiving the IJCAI Research Excellence Award, a Lifetime Achievement Award from the Canadian Artificial Intelligence Association, and an Outstanding Achievement in Research Award from the University of Massachusetts at Amherst. Sutton is a Fellow of the Royal Society of London, a Fellow of the Association for the Advancement of Artificial Intelligence, and a Fellow of the Royal Society of Canada.

计算机的诺贝尔奖:图灵奖 (since 2000)

Frequency	Topics
6	Parallel & Distributed Systems (HPC)
5	Cryptography/Security
4	Programming languages + OOP + Compiler
3	Artificial intelligence
	Internet + Networking
2	Architecture
1	Verification
	Complexity Theory
	Databases
	Computer Graphics

习近平同志《论教育》

《加强基础研究，实现高水平科技自立自强》是2023年2月21日习近平同志主持中共二十届中央政治局第三次集体学习时的讲话。指出，加强基础研究，是实现高水平科技自立自强的迫切要求，是建设世界科技强国的必由之路。应对国际科技竞争、实现高水平科技自立自强，推动构建新发展格局、实现高质量发展，迫切需要我们加强基础研究，从源头和底层解决关键技术问题。要强化基础研究前瞻性、战略性、系统性布局，坚持"四个面向"，深化基础研究体制机制改革，建设基础研究高水平支撑平台，加强基础研究人才队伍建设，广泛开展基础研究国际合作，塑造有利于基础研究的创新生态。各级党委和政府要把加强基础研究纳入科技工作重要日程，加强统筹协调，加大政策支持力度，推动基础研究实现高质量发展。

Why HPC?

美国务院资助澳“反华智库”报告称：44项最关键新兴技术，中国37项全球第一

观察者网
2023-03-03 08:57 | 上海观察者信息技术有限公司官方帐号

关注

【文/观察者网 熊超然】当地时间3月2日，知名“反华智库”——澳大利亚战略政策研究所（ASPI）发布一份报告称，中国在44项最关键和最新兴技术中，有多达37项“惊人领先于全球”，包括国防、太空、机器人、能源、环境、生物技术和人工智能等领域。

路透社报道称，作为ASPI长期的海外经费最大来源，美国国务院同样资助了这一研究。结果发现，尽管美国在高性能计算、量子计算、小型卫星和疫苗等领域引领全球发展，但通常情况下只能位居次席。这一研究还发现，中国在关键技术方面要比以前认为的更先进，在某些领域，世界排名前十的研究机构都在中国。

借此报告，ASPI渲染“西方民主国家”在全球研究竞争中失利，敦促各国政府加大研究投资。而澳媒《澳大利亚人报》也据此报告声称，在全球科技领域，中国正朝着超越西方的方向“急速发展”，将引发其很快会钳制某些最关键技术供应的担忧。

Comparison

科技	领先国家	垄断风险
先进材料与制造		
纳米材料与制造	中国	高
涂料	中国	高
智能材料	中国	中
先进复合材料	中国	中
新型超材料	中国	中
高规格加工工艺	中国	中
先进炸药和高能材料	中国	中
关键矿物提取和加工	中国	低
先进的磁体和超导体	中国	低
高级保护设备	中国	低
连续流动化学合成	中国	低
增材制造（包括3D打印）	中国	低

Comparison

人工智能、计算和通信		
先进的射频通信（包括 5G 和 6G）	中国	高
先进的光通信	中国	中
人工智能 算法和硬件加速器	中国	中
分布式账本	中国	中
高级数据分析	中国	中
机器学习（包括神经网络和深度学习）	中国	低
护性网络安全技术	中国	低
高性能计算	美国	低
先进的集成电路设计与制造	美国	低
自然语言处理（包括语音和文本识别和分析）	美国	低
能源与环境		
氢和氨的供能	中国	高
超级电容器	中国	高
电池	中国	高
光伏	中国	中
核废料管理和回收	中国	中
定向能源技术	中国	中
生物燃料	中国	低
核能	中国	低

Comparison

量子信息技术		
量子计算	美国	中
后量子密码学	中国	低
量子通信（包括量子密钥分发）	中国	低
量子传感器	中国	低
生物技术、基因技术与疫苗		
合成生物学	中国	高
生物制造	中国	中
疫苗和医疗对策	美国	中
传感、定时和导航		
光子传感器	中国	高
国防、太空、机器人与运输		
先进飞机发动机（包括高超音速）	中国	中
无人机、集群和协作机器人	中国	中
小型卫星	美国	低
自主系统运行技术	中国	低
先进机器人	中国	低
航天发射系统	美国	低

未尽研究

Course Description

- **This course is about hardware and software aspects of High Performance Computing (HPC)**
  - Homogeneous Computing
    - CPU architecture
    - Parallel computing over classical CPUs
  - Cluster Computing
    - Distributed systems across multiple workstations
  - Heterogeneous Computing
    - Nvidia GPU architecture and generations
    - Other accelerators (ASIC, FPGA) and co-processors
    - CUDA and OpenCL programming
    - Quantum computing
  - Computing Networks 算力网络/算力网
- **Designed for graduate students**

Course Contents

Chapter 1 Introduction ***1 hrs***

1.1 Computing Trends in Supercomputers

1.2 Computing Trends in Clouds

1.3 Computing Platforms for AI

Chapter 2 Multi-Core Architecture ***1 hrs***

2.1 CPU Computing

2.2 CPU Architecture

2.3 CPU Multicores

Chapter 3 Parallel Computing ***3 hrs***

3.1 Dichotomy of Parallel Computing Platforms

3.2 Parallel Algorithms

3.3 Characteristics of Tasks and interactions

3.4 Mapping and Scheduling

3.5 Scalability, Efficiency and Speedup

Course Contents

Chapter 4 Parallel Programming **4 hrs**

4.1 Threads vs. Processes

4.2 Pthread

4.3 Mutex locks

4.4 Condition Variables

4.5 MPI Programming

4.6 OpenMP Programming

Chapter 5 Distributed Systems **3 hrs**

5.1 Complexity

5.2 Clock Synchronization

5.3 Cloud Computing

5.4 Big Data Processing

Course Contents

Chapter 6   GPU Ecosystem **1 hrs**

6.1 Languages and APIs

6.2 Development Tools

6.3 Application Design Patterns

6.4 Libraries

6.5 Cluster & Grid Management

6.6 Developer Resources

6.7 Connections in Computers

6.8 CPU/GPU Arrangement

Course Contents

Chapter 7 Nvidia GPU Architecture ***3 hrs***

7.1 Nvidia GPU History

7.2 G80 Architecture

7.3 Memory Hardware in G80

7.4 GT 200 Architecture

7.5 Fermi Architecture

7.6 Kepler Architecture

7.7 Maxwell Architecture

7.8. Pascal Architecture

7.9 Volta Architecture

7.10 Ampere Architecture

7.11 Hopper Architecture

7.12 Blackwell Architecture

7.13 Nvidia GPU Architectural Comparison

7.14 Application Tuning

Course Contents

Chapter 8   CUDA Programming **4 hrs**

Course Contents

Chapter 9 OpenCL Programming **3 hrs**

9.1 OpenCL Introduction

9.2 OpenCL APIs

9.3 The OpenCL C Language

9.4 Programming Advice

9.5 OpenCL Examples

9.6 CUDA Driver API vs. OpenCL

Chapter 10 ASIC-FPGA **3 hrs**

10.1 Application Specific Integrated Circuits (ASIC)

10.2 Design of ASIC

10.3 Programmable Logic Device (PLD)

10.4 Field-Programmable Gate Arrays (FPGA)

10.5 Design and Configuration of FPGA

10.6 languages and Programmability on Hardware

Course Contents

Chapter 11   Heterogeneous System Architecture (HSA)   2 hrs

11.1 HSA Hardware Components

11.2 HSA Scheduling

11.3 HSA Software Components

11.4 Accelerated Processing Unit (APU)

11.5 Tensor Processing Unit (TPU)

11.6 Intel Co-Processor Xeon Phi

Chapter 12   Computing Networks   2 hrs

12.1 Definition and Concepts

12.2 National Computing Resources

12.3 Infrastructure

12.3 Related Technologies

Course Materials

- **Reference book**
  - None
- **PPT Slides**
  - Main focus

Course Grading

- **Attendance** 10%
- **Term paper** 40%
  - Write a survey paper and make ppt slides
  - Select a High Performance Computing related topic
    - Possible language topics（maybe two): OneAPI, HPCS (High Productivity Computing Systems) Languages (such as X10, Chapel and Fotress, PGAS), Khronos SYCL, Khronos SPIR, OpenACC, PyCUDA, Jcuda, Verilog, VHDL …
    - Possible architecture topics: Nvidia Ampere, GPU Virtualization, FPGA scheduling, Quantum Computers …
    - Contact us for other possible topics (MUST be in HPC scope)
  - Survey length: 10 pages (follow the template), ppt: 20 slides
  - Grading is based on difficulty levels (research oriented)
- **Final exam** 50%

Top Ways to Survive This Course

- **Read the slides really carefully**
  - Abstraction for memorization
  - Search Internet for any unknown word or concept
- **Lecture attendance:**
  - Do not hesitate to ask questions, point out weaknesses, make observations...
- **Discussion:**
  - Skills and involvement
- **Very important: Select a *good* survey paper topic**
  - Browse Internet for topics（mentioned in our class)
  - Challenging topics win

Useful Links for Paper Search

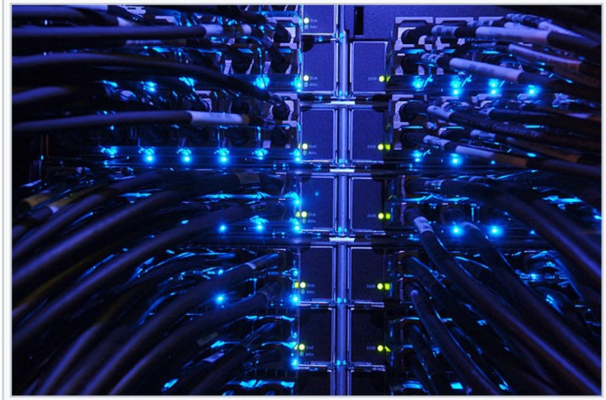
- Google (http://www.google.com/)
- Google Scholar (http://scholar.google.com/)
- DBLP (http://dblp.uni-trier.de/)
- CiteSeer (http://citeseer.ist.psu.edu/cs)
- Microsoft Academic Search (http://academic.research.microsoft.com/?SearchDomain=2)
- ACM Digital Library (http://dl.acm.org/)
- Wikipedia (http://en.wikipedia.org/wiki/Main_Page)
- IEEE Digital Library

Introduction to HPC

Wikipedia

Not to be confused with High-throughput computing or Many-task computing.

High-performance computing (**HPC**) is the use of supercomputers and computer clusters to solve advanced computation problems.



The Center for Nanoscale Materials at the Advanced Photon Source

Overview [edit]

HPC integrates systems administration (including network and security knowledge) and parallel programming into a multidisciplinary field that combines digital electronics, computer architecture, system software, programming languages, algorithms and computational techniques.[1] HPC technologies are the tools and systems used to implement and create high performance computing systems.[2] Recently[when?], HPC systems have shifted from supercomputing to computing clusters and grids.[1] Because of the need of networking in clusters and grids, High Performance Computing Technologies are being promoted[by whom?] by the use of a collapsed network backbone, because the collapsed backbone architecture is simple to troubleshoot and upgrades can be applied to a single router as opposed to multiple ones. HPC integrates with data analytics in AI engineering workflows to generate new data streams that increase simulation ability to answer the "what if" questions.[3]

The term is most commonly associated with computing used for scientific research or computational science. A related term, high-performance technical computing (HPTC), generally refers to the engineering applications of cluster-based computing (such as computational fluid dynamics and the building and testing of virtual prototypes). HPC has also been applied to business uses such as data warehouses, line of business (LOB) applications, and transaction processing.

High-performance computing (HPC) as a term arose after the term "supercomputing".[4] HPC is sometimes used as a synonym for supercomputing; but, in other contexts, "supercomputer" is used to refer to a more powerful subset of "high-performance computers", and the term "supercomputing" becomes a subset of "high-performance computing". The potential for confusion over the use of these terms is apparent.

CS Rankings - HPC

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go to a faculty member's home page. **Click on a chart icon** (the 📊 after a name or institution) to see the distribution of their publication areas as a bar chart. **Click on a Google Scholar icon** to see publications, and **click on the DBLP logo** to go to a DBLP entry. *Applying to grad school? Read this first.* For info on grad stipends, check out CSStipendRankings.org. **Do you find CSrankings useful? Sponsor CSrankings on GitHub.**

Rank institutions in USA by publications from 2014 to 2024

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence ☐
- ▶ Computer vision ☐
- ▶ Machine learning ☐
- ▶ Natural language processing ☐
- ▶ The Web & information retrieval ☐

Systems [off | on]

- ▶ Computer architecture ☐
- ▶ Computer networks ☐
- ▶ Computer security ☐
- ▶ Databases ☐
- ▶ Design automation ☐
- ▶ Embedded & real-time systems ☐
- ▶ High-performance computing ☑
- ▶ Mobile computing ☐
- ▶ Measurement & perf. analysis ☐
- ▶ Operating systems ☐
- ▶ Programming languages ☐
- ▶ Software engineering ☐

Theory [off | on]

- ▶ Algorithms & complexity ☐
- ▶ Cryptography ☐
- ▶ Logic & verification ☐

Interdisciplinary Areas [off | on]

- ▶ Comp. bio & bioinformatics ☐
- ▶ Computer graphics ☐
- ▶ Computer science education ☐
- ▶ Economics & computation ☐
- ▶ Human-computer interaction ☐
- ▶ Robotics ☐
- ▶ Visualization ☐

#	Institution	Count	Faculty
1	▶ North Carolina State University	14.5	10
2	▶ Univ. of Illinois at Urbana-Champaign	13.6	18
3	▶ University of Utah	12.5	11
4	▶ Georgia Institute of Technology	11.9	14
5	▶ Ohio State University	10.1	12
6	▶ Univ. of California - Riverside	9.0	9
7	▶ Univ. of California - Merced	8.7	4
8	▶ Northeastern University	8.2	5
9	▶ University of Chicago	7.9	7
10	▶ Indiana University	7.4	7
11	▶ Rice University	7.2	2
12	▶ University of Texas at Arlington	6.5	7
13	▶ Purdue University	6.4	6
13	▶ University of Minnesota	6.4	8
15	▶ Virginia Tech	5.9	6
16	▶ University of Delaware	5.6	6
17	▶ University of Massachusetts Amherst	5.5	5
18	▶ Univ. of California - Berkeley	5.1	5
18	▶ University of Georgia	5.1	4
18	▶ University of Houston	5.1	3
21	▶ University of Maryland - College Park	4.7	2
22	▶ University of Iowa	4.6	2
23	▶ Washington State University	4.5	5

CS Rankings - HPC

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go to a faculty member's home page. **Click on a chart icon** (the after a name or institution) to see the distribution of their publication areas as a bar chart. **Click on a Google Scholar icon** () to see publications, and **click on the DBLP logo** () to go to a DBLP entry. *Applying to grad school? Read this first.* For info on grad stipends, check out CSStipendRankings.org. **Do you find CSrankings useful? Sponsor CSrankings on GitHub.**

Rank institutions in USA by publications from 2014 to 2024

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence
- ▶ Computer vision
- ▶ Machine learning
- ▶ Natural language processing
- ▶ The Web & information retrieval

Systems [off | on]

- ▶ Computer architecture
- ▶ Computer networks
- ▶ Computer security
- ▶ Databases
- ▶ Design automation
- ▶ Embedded & real-time systems
- ▶ High-performance computing ☑
- ▶ Mobile computing
- ▶ Measurement & perf. analysis
- ▶ Operating systems
- ▶ Programming languages
- ▶ Software engineering

Theory [off | on]

- ▶ Algorithms & complexity
- ▶ Cryptography
- ▶ Logic & verification

Interdisciplinary Areas [off | on]

- ▶ Comp. bio & bioinformatics
- ▶ Computer graphics
- ▶ Computer science education
- ▶ Economics & computation
- ▶ Human-computer interaction
- ▶ Robotics
- ▶ Visualization

#	Institution	Count	Faculty
1	▼ North Carolina State University	14.5	10

Faculty	*# Pubs*	*Adj. #*
Xu Liu 0001 HPC	13	3.2
Frank Mueller 0001 HPC	10	2.7
Xipeng Shen ARCH	9	2.6
Jiajia Li 0001 HPC	7	1.5
Huiyang Zhou HPC,ARCH	6	1.5
Amro Awad ARCH	3	0.9
Greg Byrd	1	0.3
Zhishan Guo EMBEDDED	1	0.3
Sharma V. Thankachan COMP. BIO,THEORY	1	0.3
James Tuck 0001 ARCH	1	0.3

#	Institution	Count	Faculty
2	▶ Univ. of Illinois at Urbana-Champaign	13.6	18
3	▶ University of Utah	12.5	11
4	▶ Georgia Institute of Technology	11.9	14
5	▶ Ohio State University	10.1	12
6	▶ Univ. of California - Riverside	9.0	9
7	▶ Univ. of California - Merced	8.7	4
8	▶ Northeastern University	8.2	5
9	▶ University of Chicago	7.9	7
10	▶ Indiana University	7.4	7
11	▶ Rice University	7.2	2
12	▶ University of Texas at Arlington	6.5	7
13	▶ Purdue University	6.4	6

HPC vs. HTC

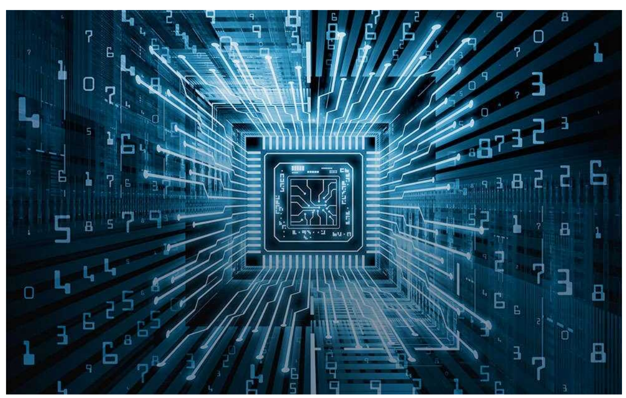
- **Software Execution Speedup**
  - High Performance Computing (HPC)
  - How to run programs faster
  - Concerns of Programmers/Users
  - Sequential vs. Parallel vs. Distributed Computing

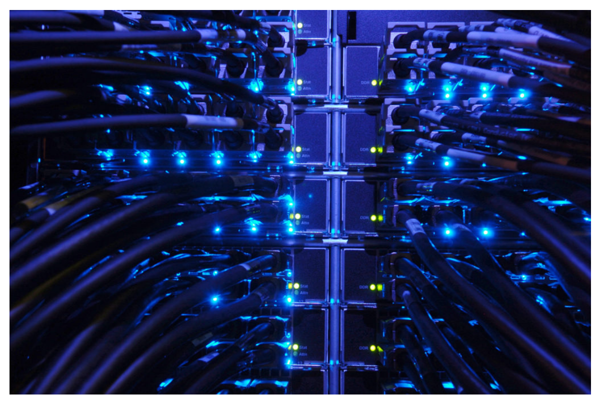
- **Resource Utilization Efficiency**
  - High Throughput Computing (HTC)
  - How to keep systems busy all the time
  - Concerns of System Administrators
  - Centralized vs. Distributed Control/Management
  - Time-Sharing vs. Space-Sharing

High Performance Computing (HPC)

- **Provide compute power**
  - Parallel processing is the key (beyond frequency limits)
  - Uses supercomputers and computer clusters to solve advanced computation problems
- **Full-Domain HPC = Full-Stack HPC + Full-Network HPC**
  - Full-Stack HPC: within compute nodes
  - Full-Network HPC: computations and communications across compute nodes/data centers
- **Two key weapons: Partitioning & Duplication**

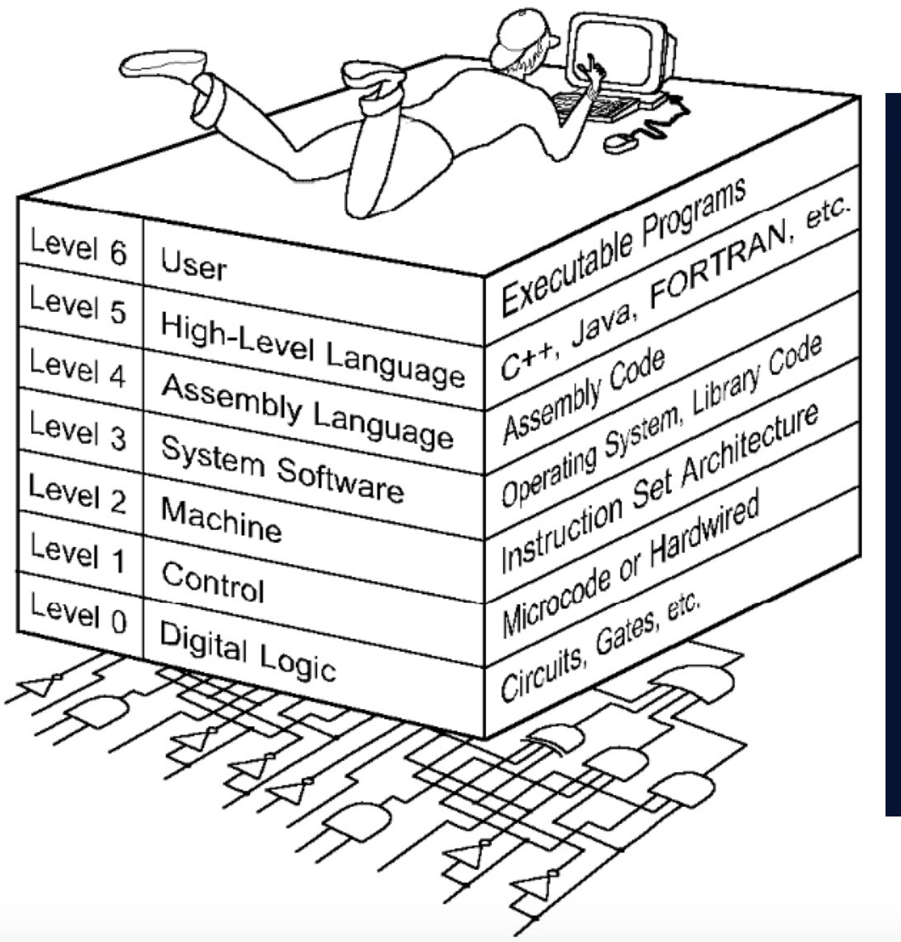
==> Increase granularity (inside & outside)





Full-Stack HPC on Compute Nodes

- **Exploring parallelism across all hardware/software layers**
  - Digital electronics, computer architecture, system software, programming languages, algorithms, and computational techniques

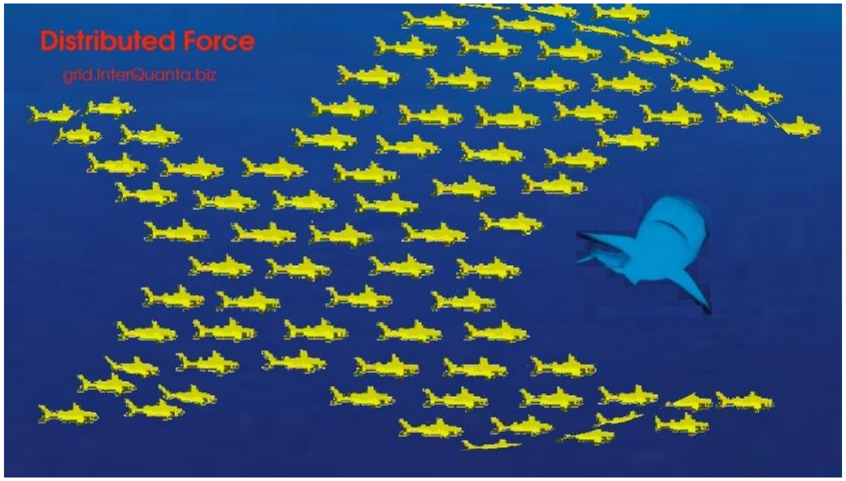


Distributed Heterogeneous Systems

- **Chat GPT 4 used at least 285,000 CPUs and 10,000 GPUs**
  - Decentralized (GPU, TPU or quantum clusters)

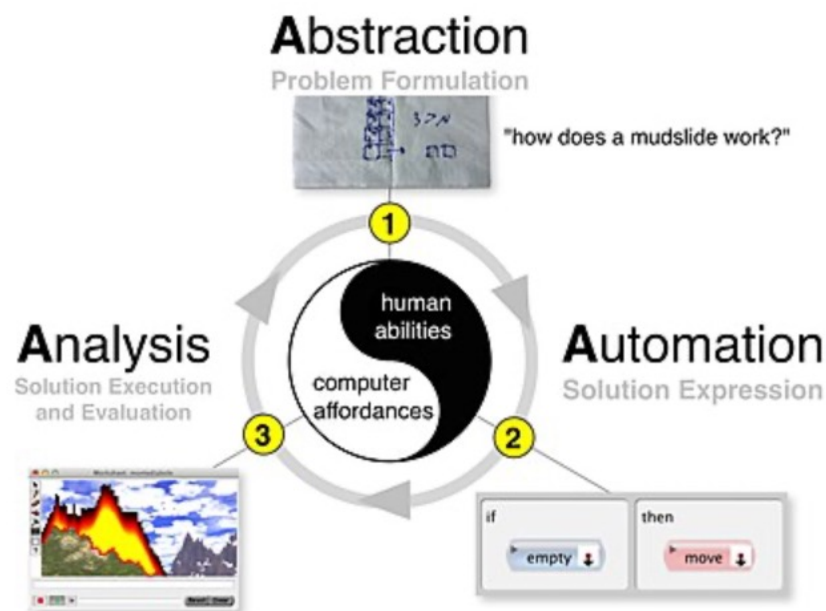
- **Full-Network HPC**
  - Distributed operating systems (control)
  - High performance communications
    - Partitioning communication for pipelining
    - Duplicating channels for parallelization
- **Goal: Full-Domain HPC = Full-Stack HPC + Full-Network HPC**

(within nodes) **(among nodes)**

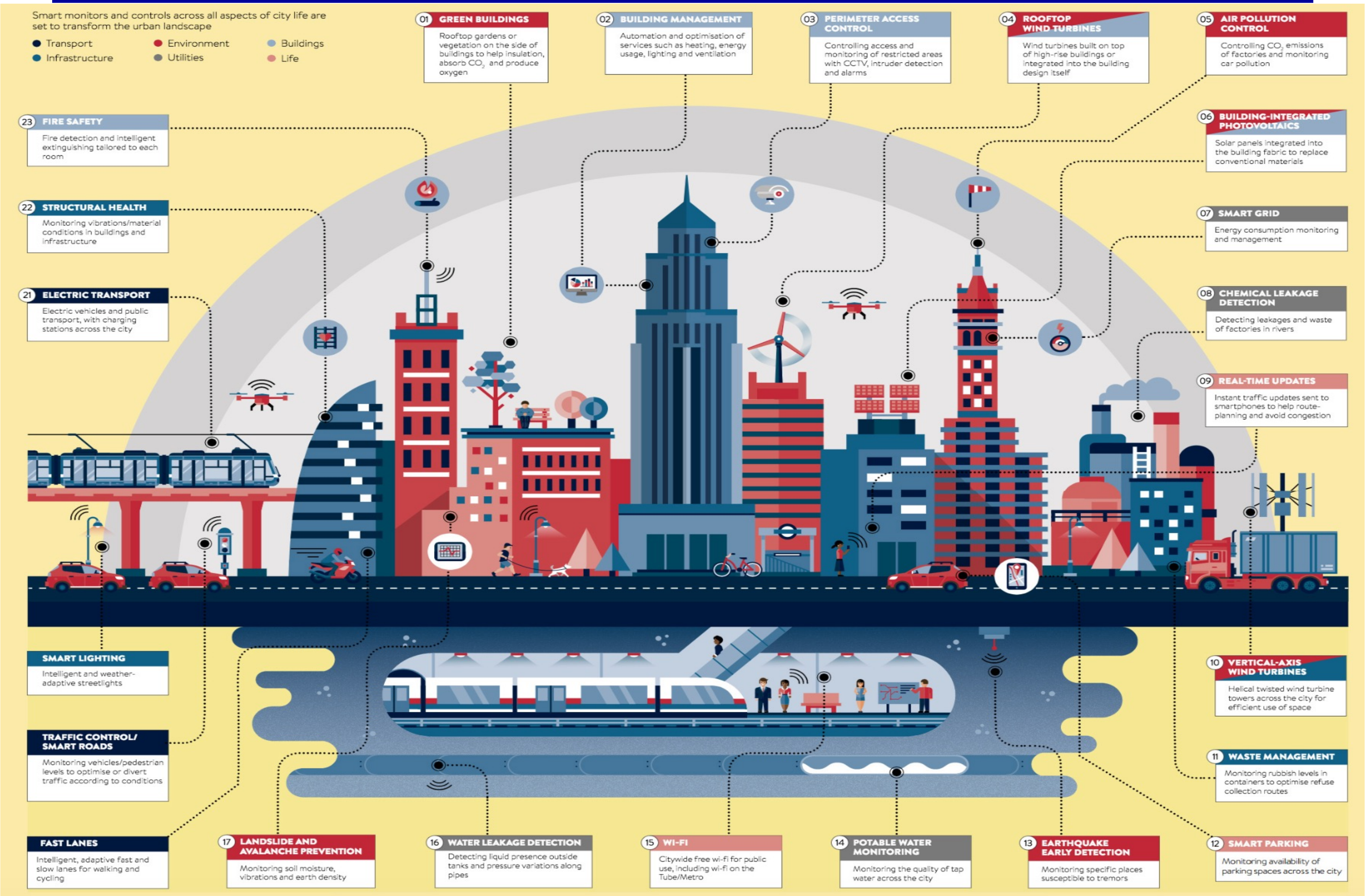


Computational Thinking

- **The mental skill to apply concepts, methods, problem solving techniques, and logic reasoning, derived from computing and computer science, to solve problems in all areas, including our daily lives**
  - Computational thinking, since it is present essentially everywhere, may be similar to data science: everyone does it, and it is hard to define without excluding any of its use cases
- **Characteristics:**
  - Decomposition, pattern recognition/data representation, generalization/abstraction and algorithms



Smart Cities

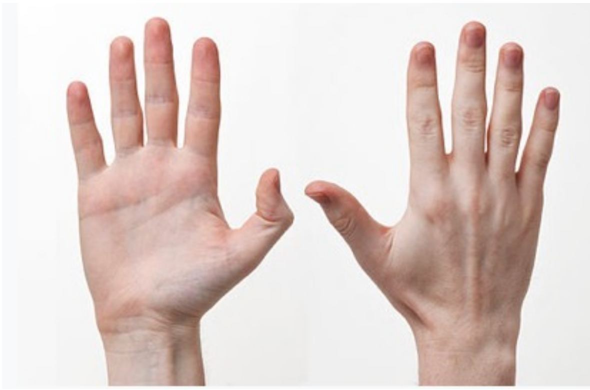


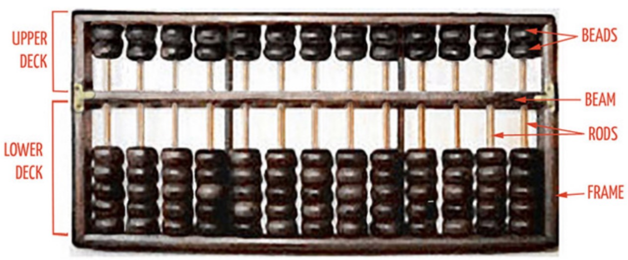
HPC Applications

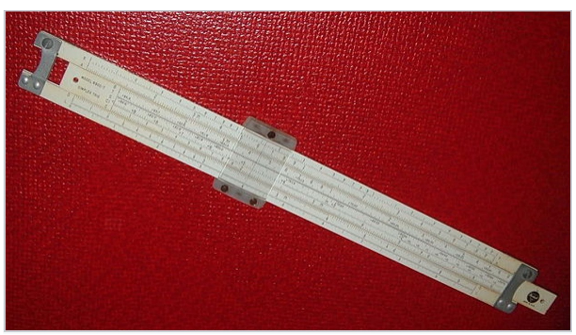
1. **Applications in Engineering and Design**
   - Design of airfoils
   - Design of internal combustion engines
   - Design of high-speed circuits
   - Design of microelectromechanical and nanoelectromechanical systems CAD/CAM
2. **Scientific Applications**
   - Bioinformatics (sequence matching, protein folding…)
   - Computational physics
   - Computational chemistry
3. **Commercial Applications**
   - Web servers
   - Database servers
   - Servers for data mining and decision making
4. **Application in Computer Systems**
   - Intrusion detection
   - Cryptography: factoring extremely large integers
   - Embedded systems (cars, tanks…)

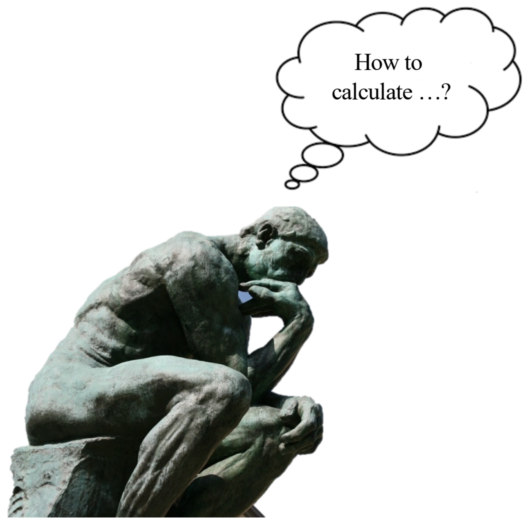
History

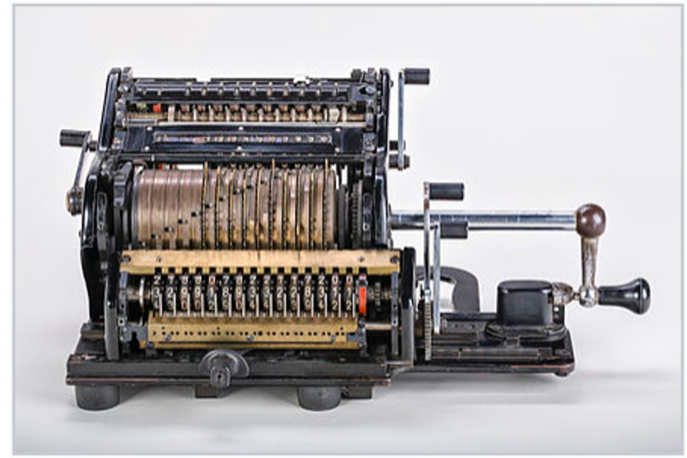
Before Modern Computers …

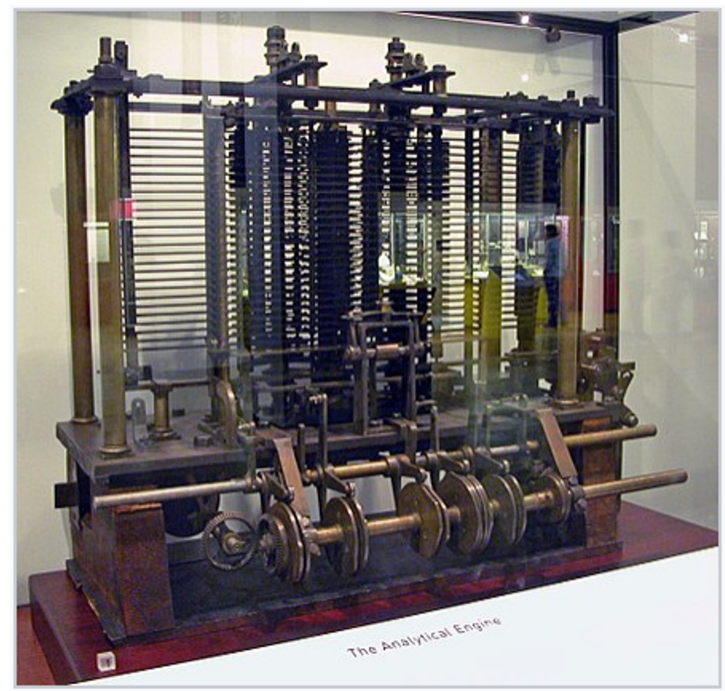








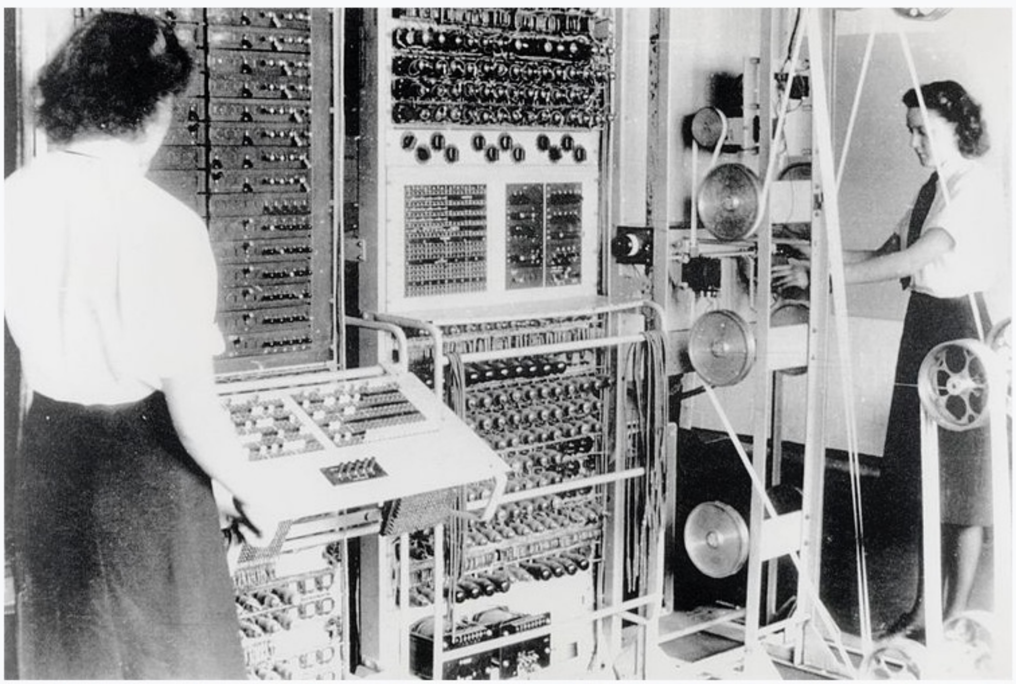




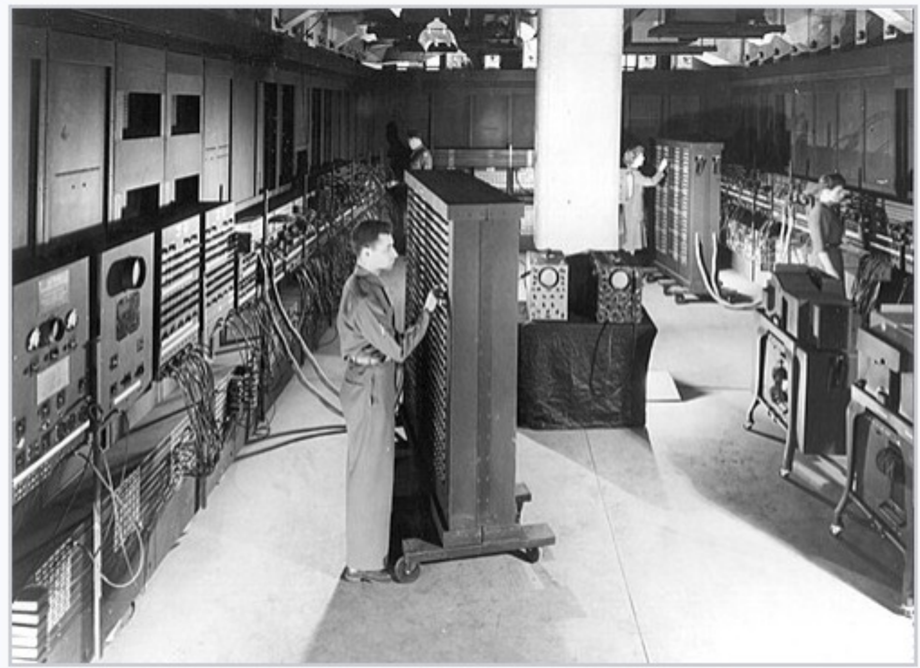
1940's: Electromechanical and Electronic Computers

- **The first functioning programmable computers**
  - Colossus in England used for breaking German codes
  - ENIAC in United States for calculating ranges of artillery shells (and later in development of the hydrogen bomb)

Colossus computer



A Colossus Mark 2 computer being operated by Wrens Dorothy Du Boisson (left) and Elsie Booker. The slanted control panel on the left was used to set the "pin" (or "cam") patterns of the Lorenz. The "bedstead" paper tape transport is on the right.



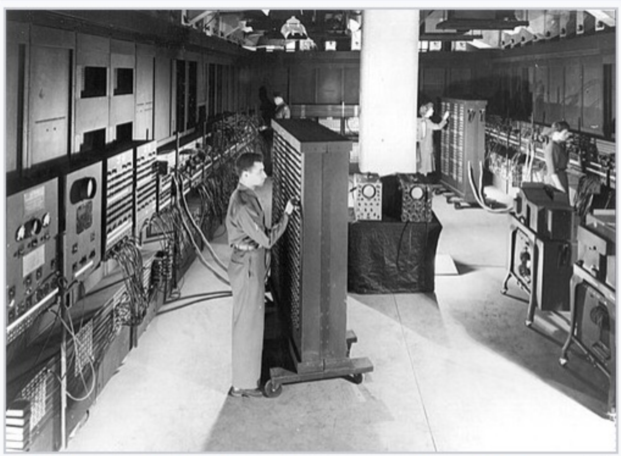
Cpl. Irwin Goldstein (foreground) sets the switches on one of ENIAC's function tables at the Moore School of Electrical Engineering.

1945: John von Neumann write First Draft for EDVAC

- In a widely circulated paper, mathematician John von Neumann outlines the architecture of a stored-program computer, including electronic storage of programming information and data -- which eliminates the need for more clumsy methods of programming such as plugboards, punched cards and paper
- Hungarian-born von Neumann demonstrated prodigious expertise in hydrodynamics, ballistics, meteorology, game theory, statistics, and the use of mechanical devices for computation
- After the war, he concentrated on the development of Princeton´s Institute for Advanced Studies computer

1946: Public Unveiling of ENIAC

- **Started in 1943, the ENIAC computing system was built by John Mauchly and J. Presper Eckert at the Moore School of Electrical Engineering of the University of Pennsylvania**
  - Because of its electronic, as opposed to electromechanical, technology, it is over 1,000 times faster than any previous computer
  - ENIAC used panel-to-panel wiring and switches for programming, occupied more than 1,000 square feet, used about 18,000 vacuum tubes and weighed 30 tons
  - It was believed that ENIAC had done more calculation over the ten years it was in operation than all of humanity had until that time



Cpl. Irwin Goldstein (foreground) sets the switches on one of ENIAC's function tables at the Moore School of Electrical Engineering.

1964: IBM Announces System/360

- **System/360 is a major event in the history of computing**
- **On April 7, IBM announced five models of System/360, spanning a 50-to-1 performance range**
  - At the same press conference, IBM also announced 40 completely new peripherals for the new family
  - System/360 was aimed at both business and scientific customers and all models could run the same software, largely without modification
- **At the time IBM released the System/360, the company had just made the transition from discrete transistors to integrated circuits, and its major source of revenue began to move from punched card equipment to electronic computer systems**

1976: Cray-1 Supercomputer Introduced

- The fastest machine of its day, The Cray-1's speed comes partly from its shape, a "C," which reduces the length of wires and thus the time signals need to travel across them
- High packaging density of integrated circuits and a novel Freon cooling system also contributed to its speed
- Each Cray-1 took a full year to assemble and test and cost about $10 million
- Typical applications included US national defense work, including the design and simulation of nuclear weapons, and weather forecasting

1978: The DEC VAX Introduced

- **Beginning with the VAX-11/780, the Digital Equipment Corporation (DEC) VAX family of computers rivals much more expensive mainframe computers in performance and features the ability to address over 4 GB of virtual memory, hundreds of times the capacity of most minicomputers**
  – Called a “complex instruction set computer,” VAX systems were backward compatible and so preserved the investment owners of previous DEC computers had in software
  – The success of the VAX family of computers transformed DEC into the second-largest computer company in the world, as VAX systems became the de facto standard computing system for industry, the sciences, engineering, and research

2010: China’s Tianhe Supercomputers Operate

- **With a peak speed of over a petaflop (one thousand trillion calculations per second), the Tianhe 1 (translation: Milky Way 1) is developed by the Chinese National University of Defense Technology using Intel Xeon processors combined with AMD graphic processing units (GPUs)**
  - The upgraded and faster Tianhe-1A used Intel Xeon CPUs as well, but switched to nVidia's Tesla GPUs and added more than 2,000 Fei-Tang (SPARC-based) processors
  - The machines were used by the Chinese Academy of Sciences to run massive solar energy simulations, as well as some of the most complex molecular studies ever undertaken

Mainframe Computers

- **Originally referred to the large cabinets that housed units.**
- **Later, the term was used to distinguish high-end machines from less powerful units**
  - IBM has traditionally dominated this portion of the market
  - Their dominance grew out of their 700/7000 series and, later, the development of the 360 series mainframes

An IBM 704 computer at NACA in 1957

Minicomputers

- **Class of computers that developed in the mid-1960s that were smaller and cheaper than mainframe and mid-size computers from IBM**
- **Later "minicomputer" came to mean a machine that is smaller than the mainframe computers but larger than the microcomputers**
  - Digital Equipment Corporation's (DEC) PDP-8
  - The term *minicomputer* is no longer widely used
    - The term *midrange computer* is now preferred

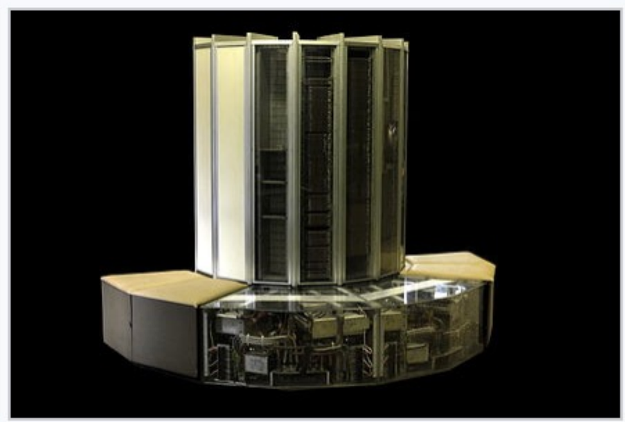
A PDP-8 on display at the Smithsonian's National Museum of American History in Washington, D.C.. This example is from the first generation of PDP-8s, built with discrete transistors and later known as the *Straight 8*.

Supercomputers

- **The first computer to be commonly referred to as a *supercomputer* was the CDC6600, released in 1964**
  - It was designed by Seymour Cray
  - In 1976 the Cray 1 became one of the most successful supercomputers in history
- **The latest supercomputer list can be found at:**

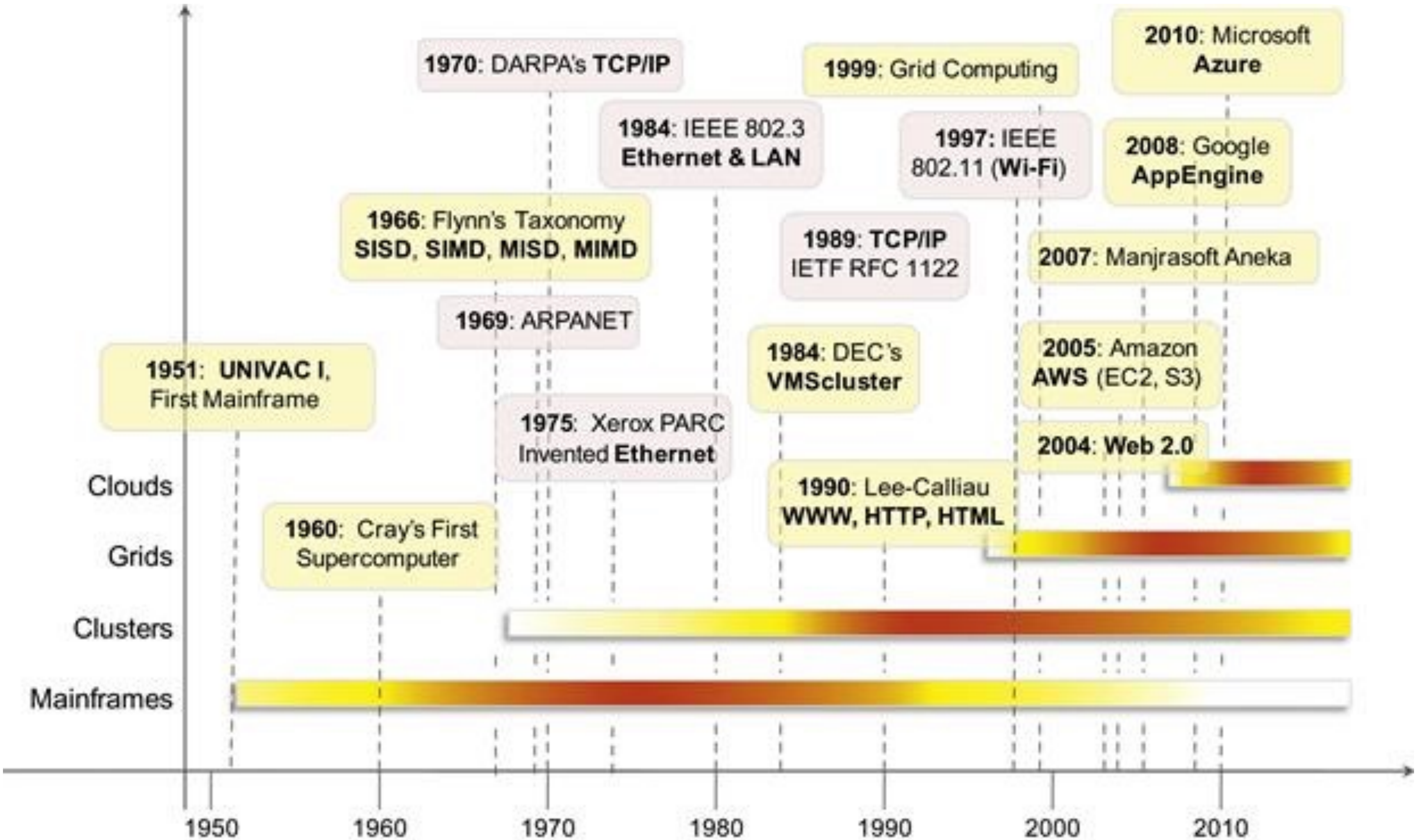
  https://top500.org

The CDC 6600. Behind the system console are two of the "arms" of the plus-sign shaped cabinet with the covers opened. Individual modules can be seen inside. The racks holding the modules are hinged to give access to the racks behind them. Each arm of the machine had up to four such racks. On the right is the cooling system.



Cray-1 with internals exposed at EPFL

The Evolution of HPC



Top500

Turing Award 2021

Jack Dongarra

ACM TURING AWARD HONORS JACK J. DONGARRA FOR PIONEERING CONCEPTS AND METHODS WHICH HAVE RESULTED IN WORLD-CHANGING COMPUTATIONS

Dongarra's Algorithms and Software Fueled the Growth of High-Performance Computing and Had Significant Impacts in Many Areas of Computational Science from AI to Computer Graphics

ACM, the Association for Computing Machinery, today named Jack J. Dongarra recipient of the 2021 ACM A.M. Turing Award for pioneering contributions to numerical algorithms and libraries that enabled high performance computational software to keep pace with exponential hardware improvements for over four decades. Dongarra is a University Distinguished Professor of Computer Science in the Electrical Engineering and Computer Science Department at the University of Tennessee. He also holds appointments with Oak Ridge National Laboratory and the University of Manchester.

The ACM A.M. Turing Award, often referred to as the "Nobel Prize of Computing," carries a $1 million prize, with financial support provided by Google, Inc. It is named for Alan M. Turing, the British mathematician who articulated the mathematical foundation and limits of computing.

Dongarra has led the world of high-performance computing through his contributions to efficient numerical algorithms for linear algebra operations, parallel computing programming mechanisms, and performance evaluation tools. For nearly forty years, Moore's Law produced exponential growth in hardware performance. During that same time, while most software failed to keep pace with these hardware advances, high performance numerical software did—in large part due to Dongarra's algorithms, optimization techniques, and production-quality software implementations.

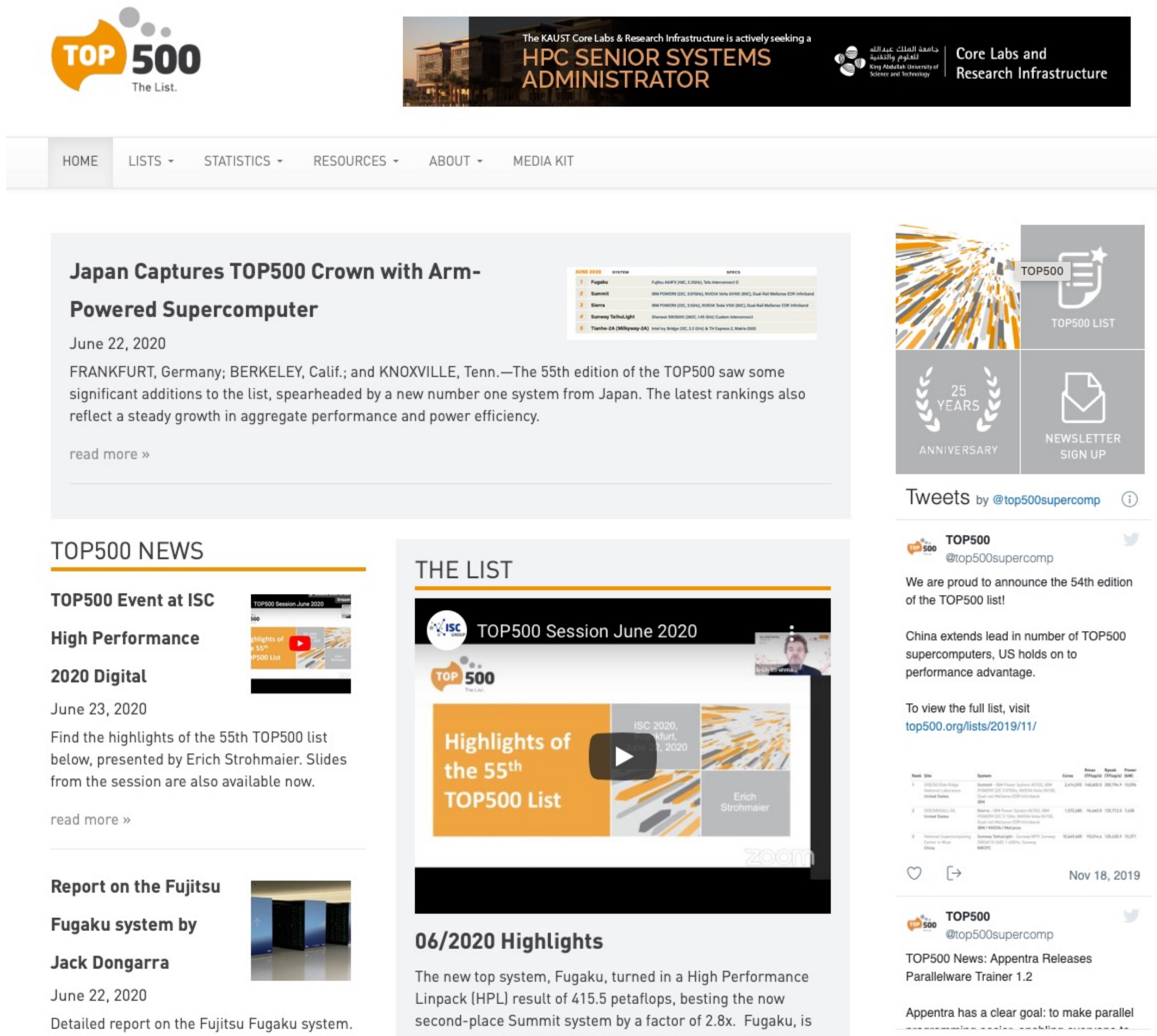
These contributions laid a framework from which scientists and engineers made important discoveries and game-changing innovations in areas including big data analytics, healthcare, renewable energy, weather prediction, genomics, and economics, to name a few. Dongarra's work also helped facilitate leapfrog advances in computer architecture and supported revolutions in computer graphics and deep learning.

Biographical Background

Jack J. Dongarra has been a University Distinguished Professor at the University of Tennessee and a Distinguished Research Staff Member at the Oak Ridge National Laboratory since 1989. He has also served as a Turing Fellow at the University of Manchester (UK) since 2007. Dongarra earned a B.S. in Mathematics from Chicago State University, an M.S. in Computer Science from the Illinois Institute of Technology, and a Ph.D. in Applied Mathematics from the University of New Mexico.

Top500

- https://www.top500.org



June 2004

	ite Country/Year	Computer / Processors Manufacturer	Computer Family Model	Inst. type Installation Area	R_{max} R_{peak}	N_{ma} n_{half}
1	Earth Simulator Center Japan/2002	**Earth-Simulator** / 5120 NEC	**NEC Vector** SX6	**Research**	35860 40960	1.0752e 266240
2	Lawrence Livermore National Laboratory United States/2004	*Thunder* **Intel Itanium2 Tiger4 1.4GHz - Quadrics** / 4096 California Digital Corporation	**NOW - Intel Itanium** Itanium2 Tiger4 Cluster - Quadrics	**Research**	19940 22938	975000 110000
3	Los Alamos National Laboratory United States/2002	**ASCI Q - AlphaServer SC45, 1.25 GHz** / 8192 HP	**HP AlphaServer** Alpha-Server-Cluster	**Research**	13880 20480	633000 225000
4	IBM - Rochester United States/2004	**BlueGene/L DD1 Prototype (0.5GHz PowerPC 440 w/Custom)** / 8192 IBM/ LLNL	**IBM BlueGene/L** BlueGene/L	**Vendor**	11680 16384	331775
5	NCSA United States/2003	*Tungsten* **PowerEdge 1750, P4 Xeon 3.06 GHz, Myrinet** / 2500 Dell	**Dell Cluster** PowerEdge 1750, Myrinet	**Academic**	9819 15300	630000
6	ECMWF United Kingdom/2004	**eServer pSeries 690 (1.9 GHz Power4+)** / 2112 IBM	**IBM SP** SP Power4+, Federation	**Research** Weather and Climate Research	8955 16051	350000
7	Institute of Physical and Chemical Res. (RIKEN) Japan/2004	**RIKEN Super Combined Cluster** / 2048 Fujitsu	**Fujitsu Cluster** Fujitsu Cluster	**Research**	8728 12534	474200 120000
8	IBM - Thomas Watson Research Center United States/2004	**BlueGene/L DD2 Prototype (0.7 GHz PowerPC 440)** / 4096 IBM/ LLNL	**IBM BlueGene/L** BlueGene/L	**Research**	8655 11469	294911
9	Pacific Northwest National Laboratory United States/2003	*Mpp2* **Integrity rx2600 Itanium2 1.5 GHz, Quadrics** / 1936 HP	**HP Cluster** Integrity rx2600 Itanium2 Cluster	**Research**	8633 11616	835000 140000
	Shanghai Supercomputer ...nter	**Dawning 4000A, Opteron 2.2 GHz, Myrinet** / 2560	**NOW - AMD** NOW Cluster - AMD -	**Research**	8061 11264	7284... 18...

June 2006

Rank	Site	Computer	Processors	Year	R_{max}	R_{peak}
1	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution IBM	131072	2005	280600	367000
2	IBM Thomas J. Watson Research Center United States	BGW - eServer Blue Gene Solution IBM	40960	2005	91290	114688
3	DOE/NNSA/LLNL United States	ASC Purple - eServer pSeries p5 575 1.9 GHz IBM	12208	2006	75760	92781
4	NASA/Ames Research Center/NAS United States	Columbia - SGI Altix 1.5 GHz, Voltaire Infiniband SGI	10160	2004	51870	60960
5	Commissariat a l'Energie Atomique (CEA) France	Tera-10 - NovaScale 5160, Itanium2 1.6 GHz, Quadrics Bull SA	8704	2006	42900	55705.6
6	Sandia National Laboratories United States	Thunderbird - PowerEdge 1850, 3.6 GHz, Infiniband Dell	9024	2006	38270	64972.8
7	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME Grid Cluster - Sun Fire X64 Cluster, Opteron 2.4/2.6 GHz, Infiniband NEC/Sun	10368	2006	38180	49868.8
8	Forschungszentrum Juelich (FZJ) Germany	JUBL - eServer Blue Gene Solution IBM	16384	2006	37330	45875
9	Sandia National Laboratories United States	Red Storm Cray XT3, 2.0 GHz Cray Inc.	10880	2005	36190	43520
10	The Earth Simulator Center Japan	Earth-Simulator NEC	5120	2002	35860	40960

June 2008

Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz , Voltaire Infiniband / 2008 IBM	122400	1026.00	1375.78	2345.50
2	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution / 2007 IBM	212992	478.20	596.38	2329.60
3	Argonne National Laboratory United States	Blue Gene/P Solution / 2007 IBM	163840	450.30	557.06	1260.00
4	Texas Advanced Computing Center/Univ. of Texas United States	Ranger - SunBlade x6420, Opteron Quad 2Ghz, Infiniband / 2008 Sun Microsystems	62976	326.00	503.81	2000.00
5	DOE/Oak Ridge National Laboratory United States	Jaguar - Cray XT4 QuadCore 2.1 GHz / 2008 Cray Inc.	30976	205.00	260.20	1580.71
6	Forschungszentrum Juelich (FZJ) Germany	JUGENE - Blue Gene/P Solution / 2007 IBM	65536	180.00	222.82	504.00
7	New Mexico Computing Applications Center (NMCAC) United States	Encanto - SGI Altix ICE 8200, Xeon quad core 3.0 GHz / 2007 SGI	14336	133.20	172.03	861.63
8	Computational Research Laboratories, TATA SONS India	EKA - Cluster Platform 3000 BL460c, Xeon 53xx 3GHz, Infiniband / 2008 Hewlett-Packard	14384	132.80	172.61	786.00
9	IDRIS France	Blue Gene/P Solution / 2008 IBM	40960	112.50	139.26	315.00
10	Total Exploration Production France	SGI Altix ICE 8200EX, Xeon quad core 3.0 GHz / 2008 SGI	10240	106.10	122.88	442.00

November 2010

Rank	Site	Computer
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C NUDT
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU Dawning
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows NEC/HP
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz Cray Inc.
6	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 Bull SA
7	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband IBM
8	National Institute for Computational Sciences/University of Tennessee United States	Kraken XT5 - Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.
9	Forschungszentrum Juelich (FZJ) Germany	JUGENE - Blue Gene/P Solution IBM
10	DOE/NNSA/LANL/SNL United States	Cielo - Cray XE6 8-core 2.4 GHz Cray Inc.

June 2012

Rank	Site	Computer
1	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu
3	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
4	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR IBM
5	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 NUDT
6	DOE/SC/Oak Ridge National Laboratory United States	**Jaguar** - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 Cray Inc.
7	CINECA Italy	**Fermi** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
8	Forschungszentrum Juelich (FZJ) Germany	**JuQUEEN** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
9	CEA/TGCC-GENCI France	**Curie thin nodes** - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR Bull
10	National Supercomputing Centre in Shenzhen (NSCS) China	**Nebulae** - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 Dawning

June 2015

RANK	SITE	SYSTEM	CORES	RMAX (TFLOP/S)	RPEAK (TFLOP/S)	POWER (KW)
1	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945
6	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint** - Cray XC30, Xeon E5-2670 8C 2.600GHz, Aries interconnect , NVIDIA K20x Cray Inc.	115,984	6,271.0	7,788.9	2,325
7	King Abdullah University of Science and Technology Saudi Arabia	**Shaheen II** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	196,608	5,537.0	7,235.2	2,834
8	Texas Advanced Computing Center/Univ. of Texas United States	**Stampede** - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell	462,462	5,168.1	8,520.1	4,510
9	Forschungszentrum Juelich (FZJ) Germany	**JUQUEEN** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	458,752	5,008.9	5,872.0	2,301
10	DOE/NNSA/LLNL United States	**Vulcan** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	393,216	4,293.3	5,033.2	1,972

June 2017

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

November 2019

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
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	200,794.9	10,096
2	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
5	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR , Dell EMC Texas Advanced Computing Center/Univ. of Texas United States	448,448	23,516.4	38,745.9	
6	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray/HPE Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	27,154.3	2,384
7	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray/HPE DOE/NNSA/LANL/SNL United States	979,072	20,158.7	41,461.2	7,578
8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	32,576.6	1,649
9	**SuperMUC-NG** - ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , Lenovo Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	
10	**Lassen** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	288,288	18,200.0	23,047.2	

November 2022

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,463,616	174.70	255.75	5,610
5	**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	148.60	200.79	10,096
6	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438
7	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44	15,371
8	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE	761,856	70.87	93.75	2,589

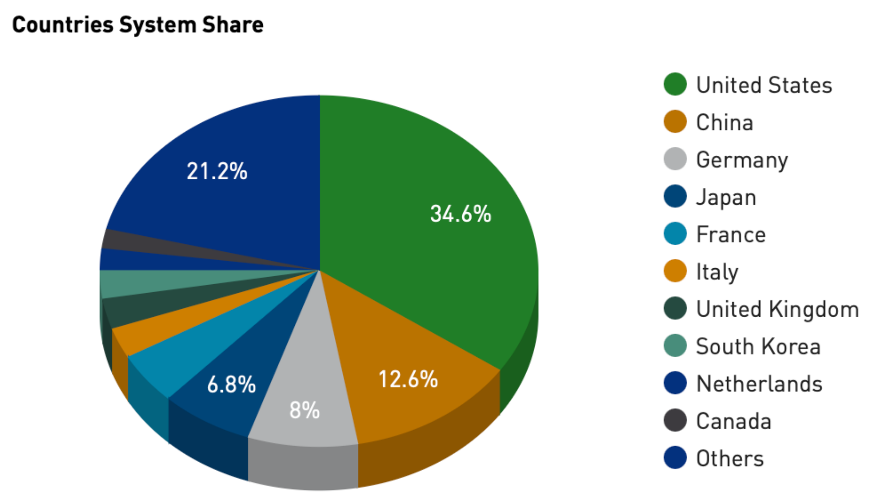
June 2023

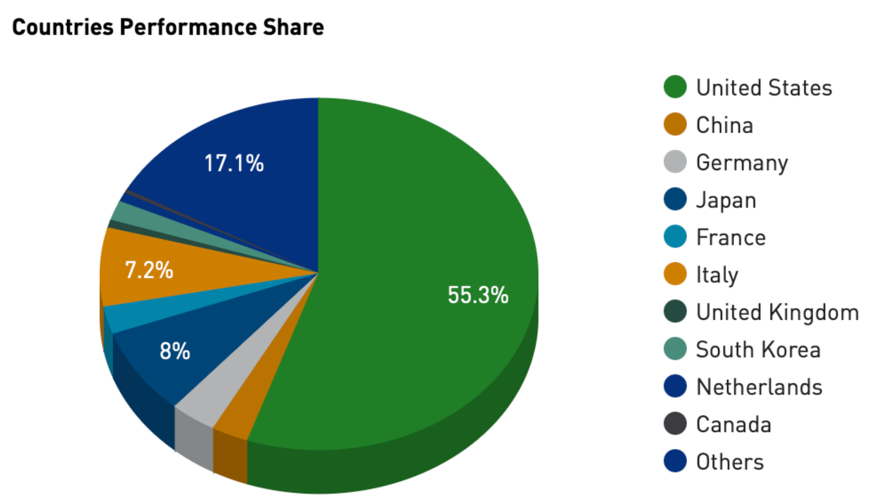
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	1,679.82	22,703
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,824,768	238.70	304.47	7,404
5	**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	148.60	200.79	10,096
6	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438
7	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44	15,371
8	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75	2,589

November 2024

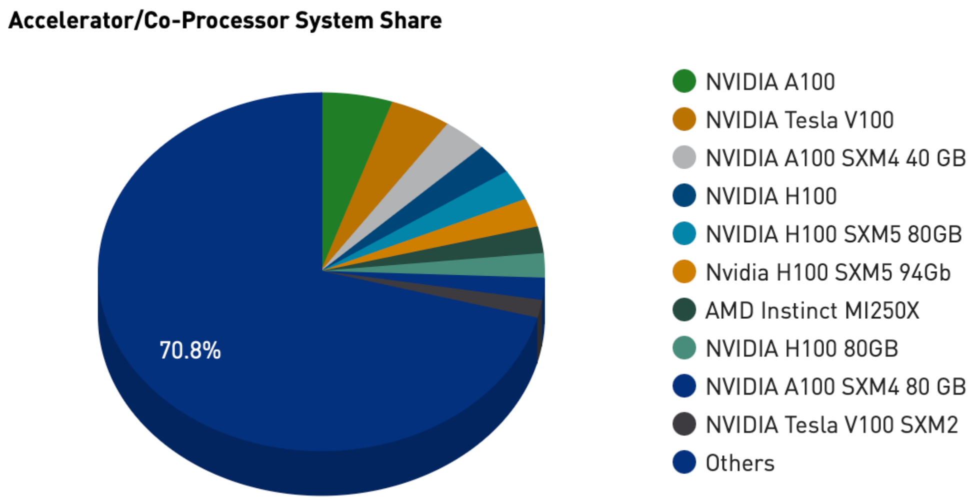
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**El Capitan** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581
2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
4	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84	
5	**HPC6** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A. Italy	3,143,520	477.90	606.97	8,461
6	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	574.84	7,124
8	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107
9	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN EuroHPC/CINECA Italy	1,824,768	241.20	306.31	7,494
10	**Tuolumne** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	1,161,216	208.10	288.88	3,387

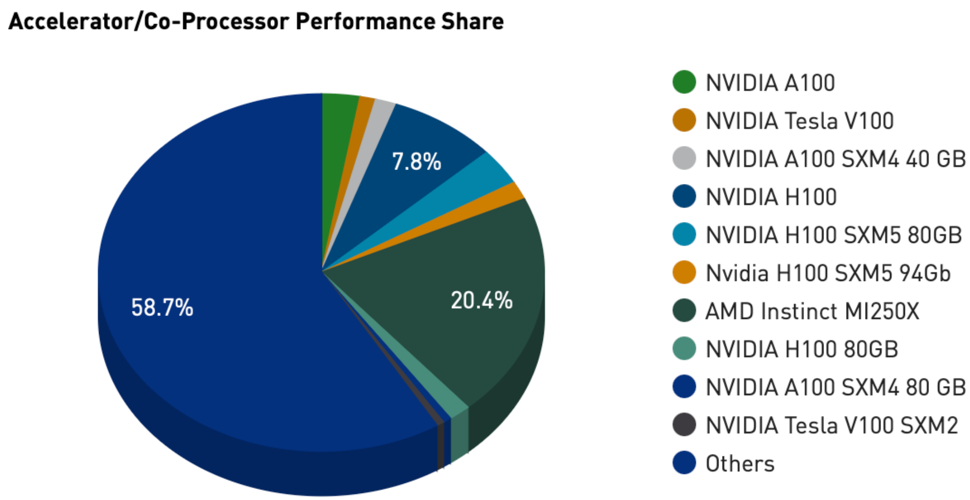
Statistics: Countries





Statistics: Accelerator/Co-Processor

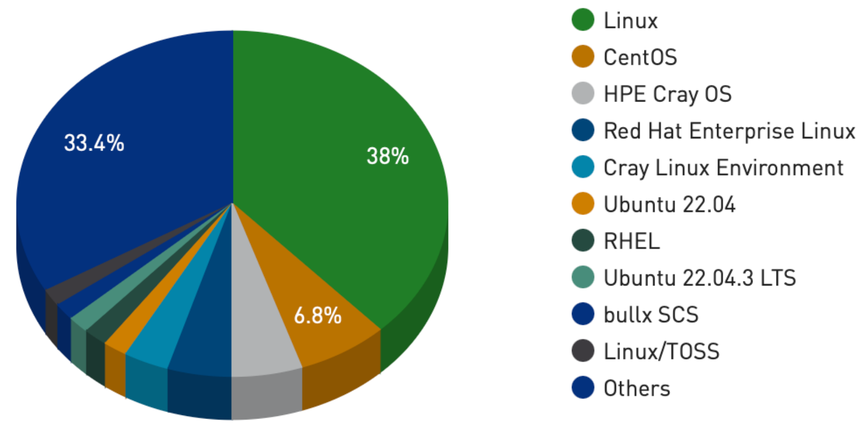




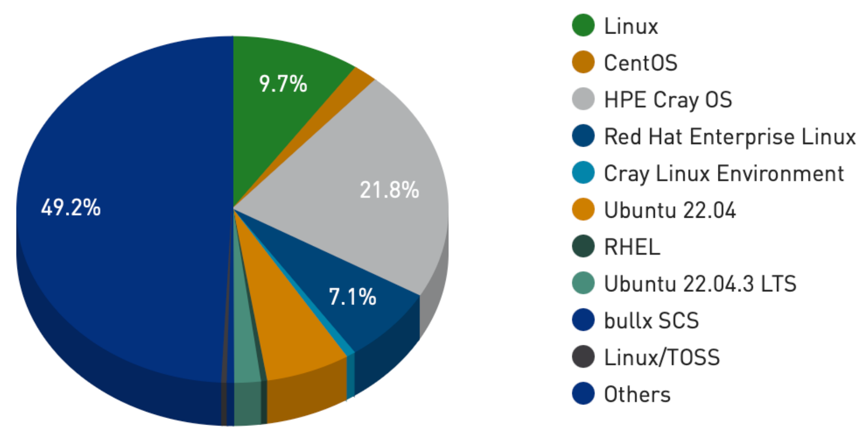
Operating Systems

- https://www.top500.org

Operating System System Share

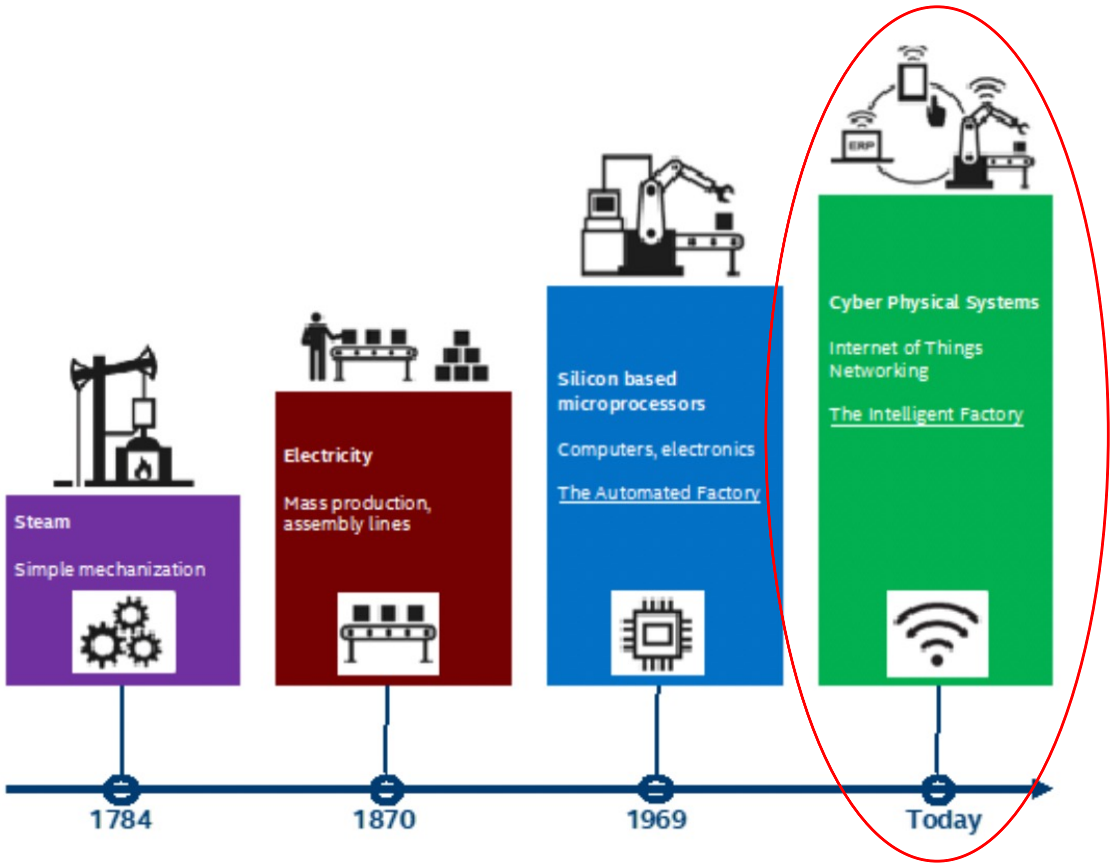


Operating System Performance Share

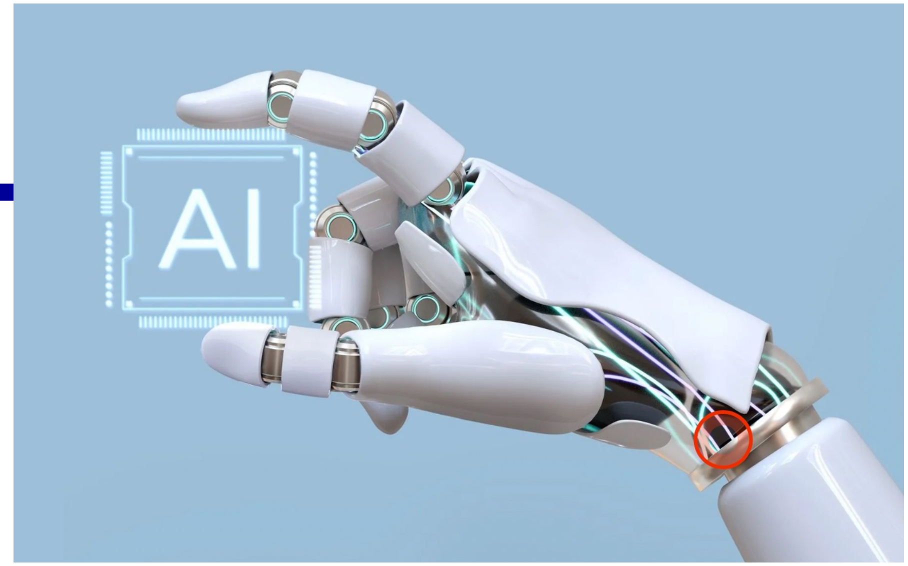


Fourth Industrial Revolution

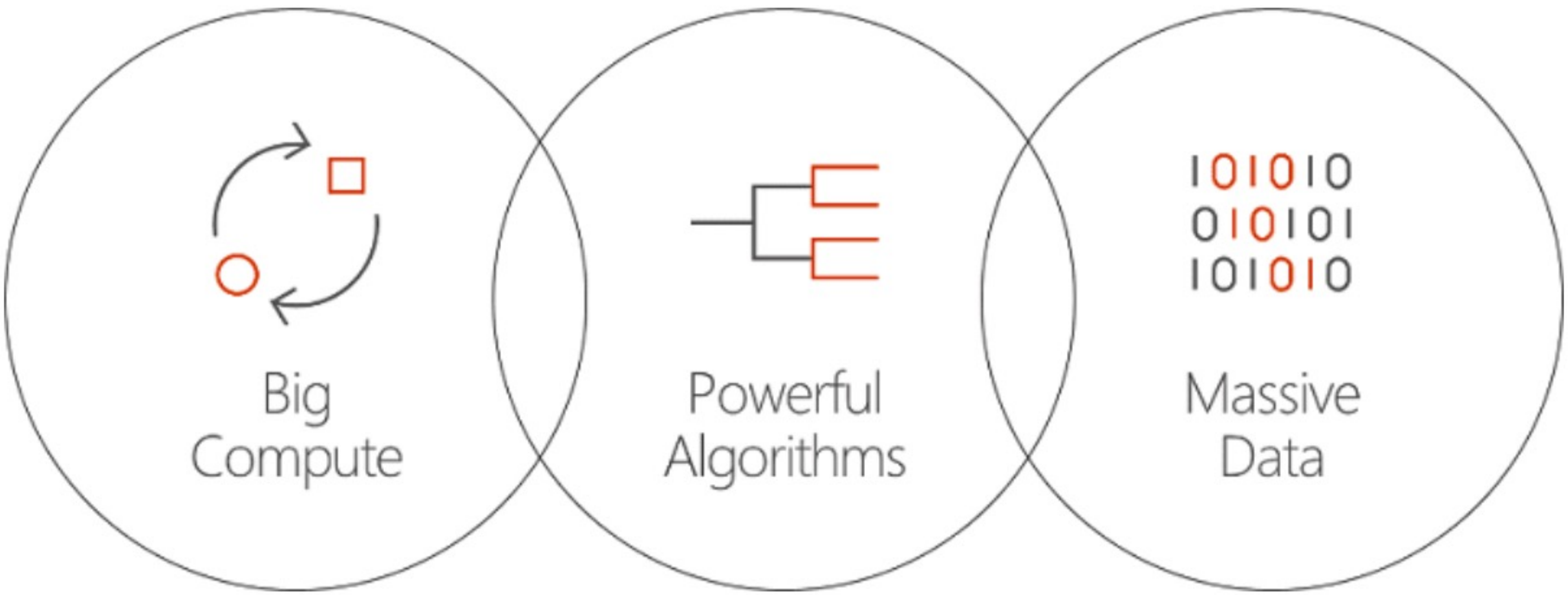
- **In essence, the Fourth Industrial Revolution is the trend towards automation and data exchange in manufacturing technologies and processes**
  - Including cyber-physical systems (CPS), IoT, industrial internet of things, cloud computing, cognitive computing, and artificial intelligence



AI Factors



- **Three factors:**
  - Algorithm innovation
  - Big Data
  - Compute power
    - The driving force for large-scale AI models



August, 1956, Dartmouth