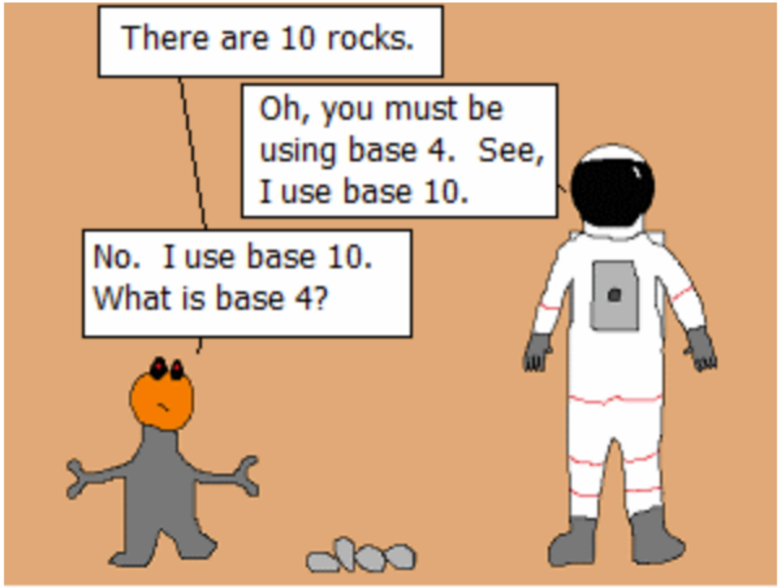
面向科学问题求解的编程实践

- 期末报告的截止时间是6月26日周三16:00。
- 提交方式：https://www.bb.ustc.edu.cn/

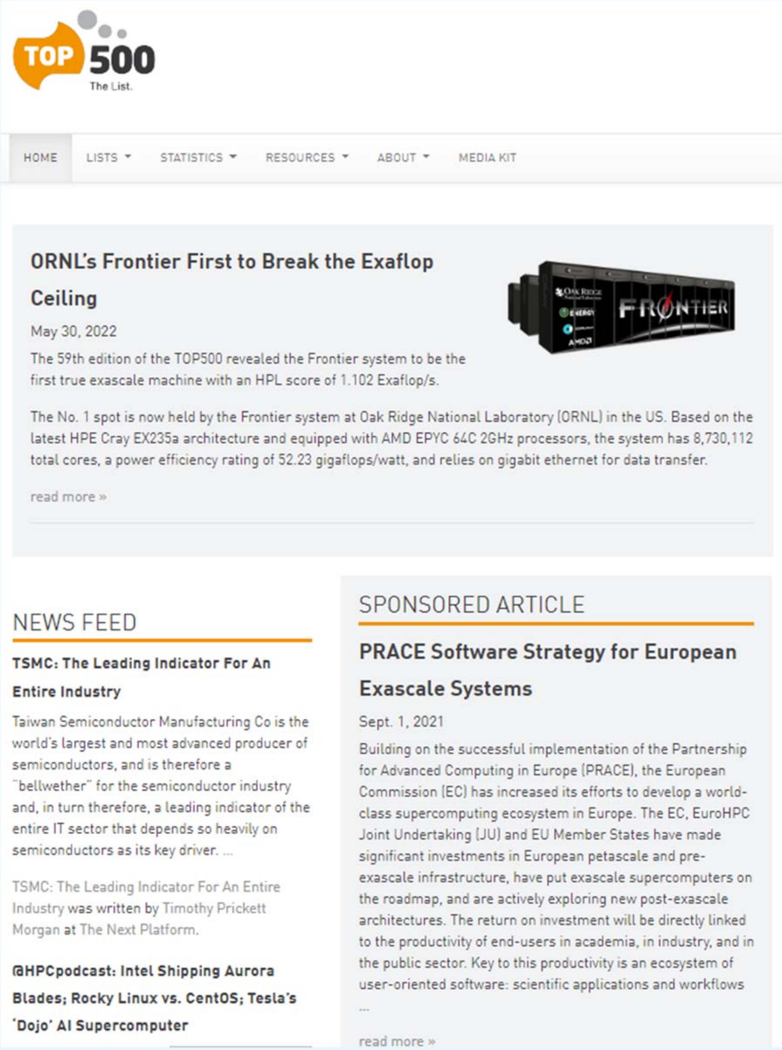

There are 10 rocks.
Oh, you must be using base 4. See, I use base 10.
No. I use base 10. What is base 4?

Every base is base 10.

内容提要

- 摩尔定律
- 并行计算与高性能计算
- 并行加速比定律

http://www.top500.org



ORNL's Frontier First to Break the Exaflop Ceiling

May 30, 2022

The 59th edition of the TOP500 revealed the Frontier system to be the first true exascale machine with an HPL score of 1.102 Exaflop/s.

The No. 1 spot is now held by the Frontier system at Oak Ridge National Laboratory (ORNL) in the US. Based on the latest HPE Cray EX235a architecture and equipped with AMD EPYC 64C 2GHz processors, the system has 8,730,112 total cores, a power efficiency rating of 52.23 gigaflops/watt, and relies on gigabit ethernet for data transfer.

read more »

NEWS FEED

TSMC: The Leading Indicator For An Entire Industry

Taiwan Semiconductor Manufacturing Co is the world's largest and most advanced producer of semiconductors, and is therefore a "bellwether" for the semiconductor industry and, in turn therefore, a leading indicator of the entire IT sector that depends so heavily on semiconductors as its key driver. ...

TSMC: The Leading Indicator For An Entire Industry was written by Timothy Prickett Morgan at The Next Platform.

@HPCpodcast: Intel Shipping Aurora Blades; Rocky Linux vs. CentOS; Tesla's 'Dojo' AI Supercomputer

SPONSORED ARTICLE

PRACE Software Strategy for European Exascale Systems

Sept. 1, 2021

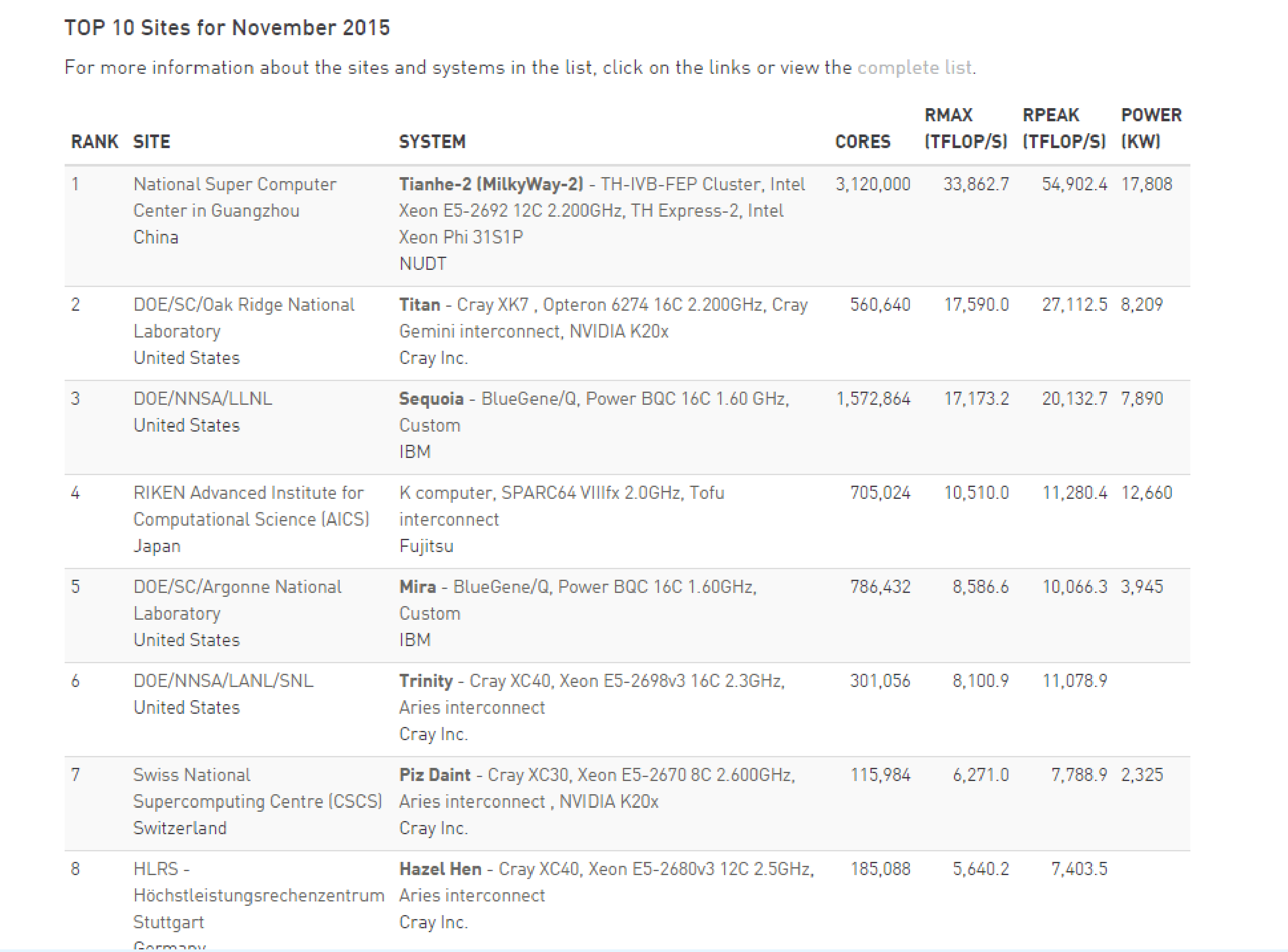
Building on the successful implementation of the Partnership for Advanced Computing in Europe (PRACE), the European Commission (EC) has increased its efforts to develop a world-class supercomputing ecosystem in Europe. The EC, EuroHPC Joint Undertaking (JU) and EU Member States have made significant investments in European petascale and pre-exascale infrastructure, have put exascale supercomputers on the roadmap, and are actively exploring new post-exascale architectures. The return on investment will be directly linked to the productivity of end-users in academia, in industry, and in the public sector. Key to this productivity is an ecosystem of user-oriented software: scientific applications and workflows ...

read more »

每年更新两次

- 十一月：SC（美国）
- 六月：ISC（欧洲）

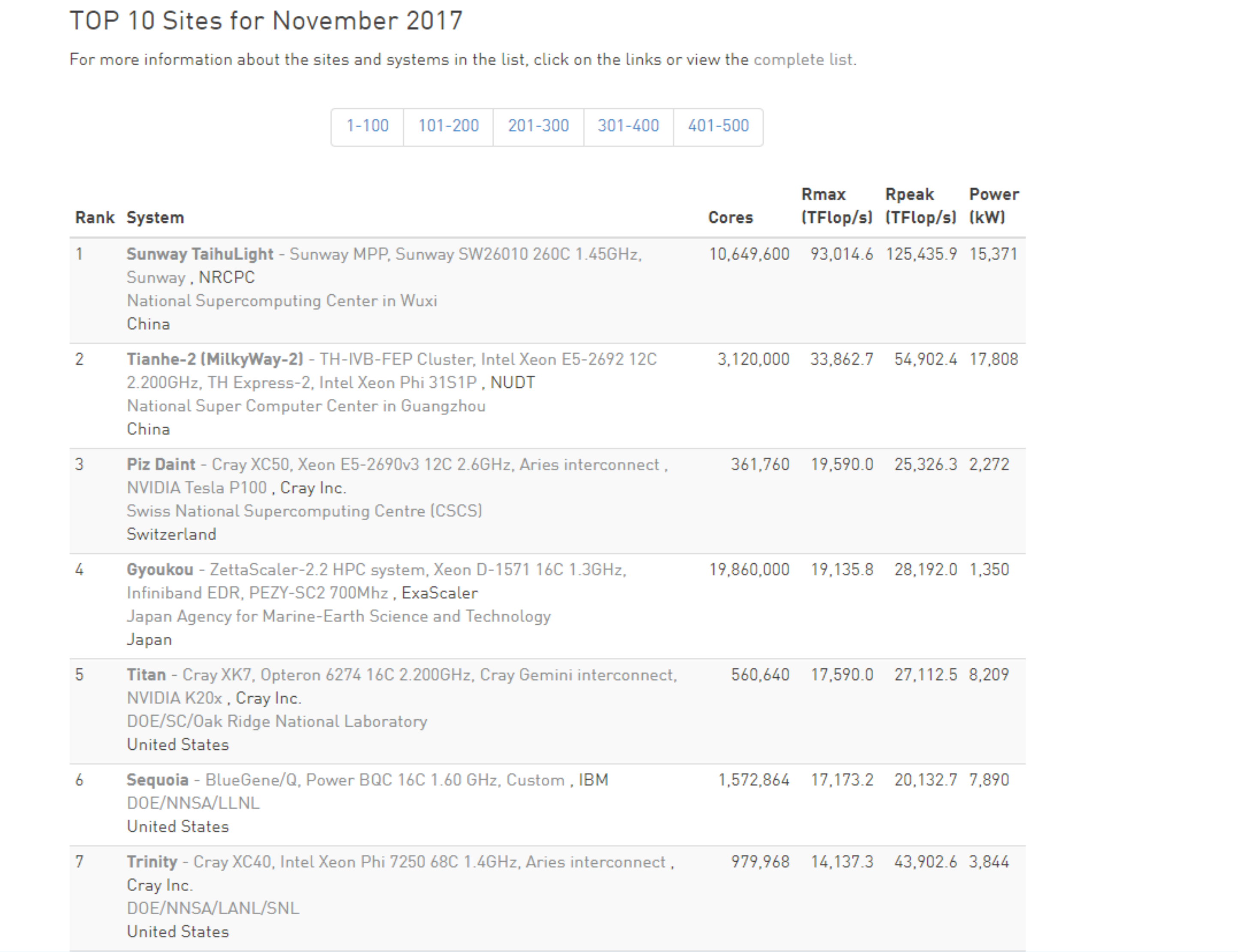
自1993年至今，2023年11月是第62版



TOP 10 Sites for November 2015

For more information about the sites and systems in the list, click on the links or view the complete list.

				RMAX	RPEAK	POWER
RANK	SITE	SYSTEM	CORES	(TFLOP/S)	(TFLOP/S)	(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	DOE/NNSA/LANL/SNL United States	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	301,056	8,100.9	11,078.9	
7	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
8	HLRS - Höchstleistungsrechenzentrum Stuttgart Germany	**Hazel Hen** - Cray XC40, Xeon E5-2680v3 12C 2.5GHz, Aries interconnect Cray Inc.	185,088	5,640.2	7,403.5	

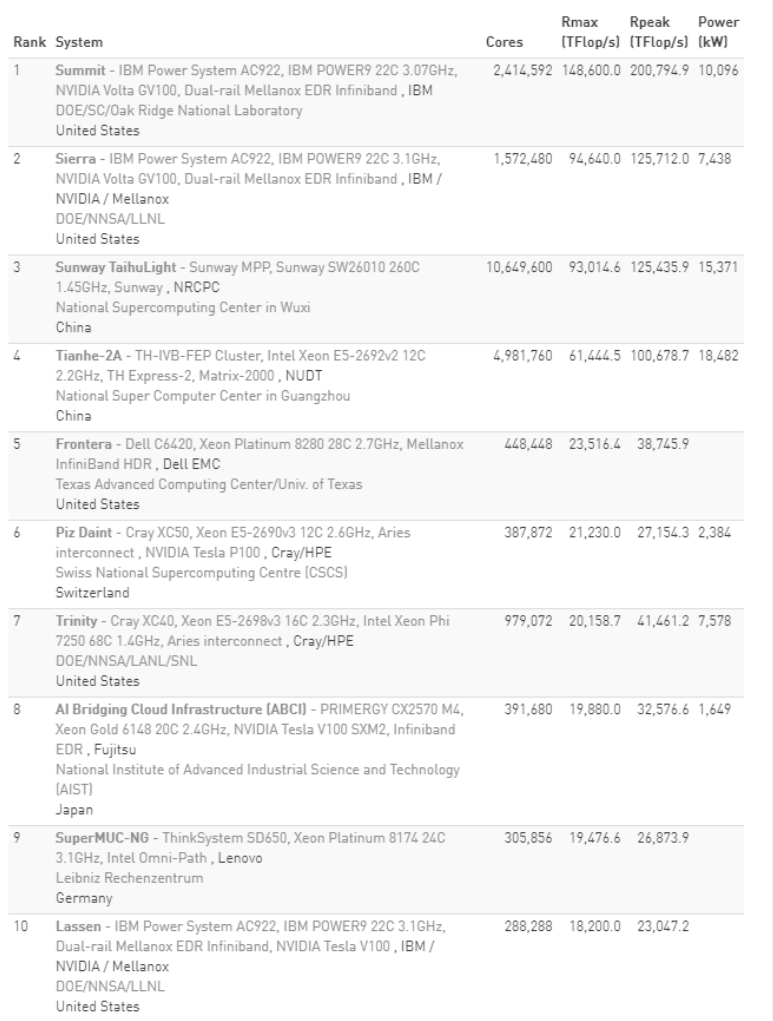


TOP 10 Sites for November 2017

For more information about the sites and systems in the list, click on the links or view the complete list.

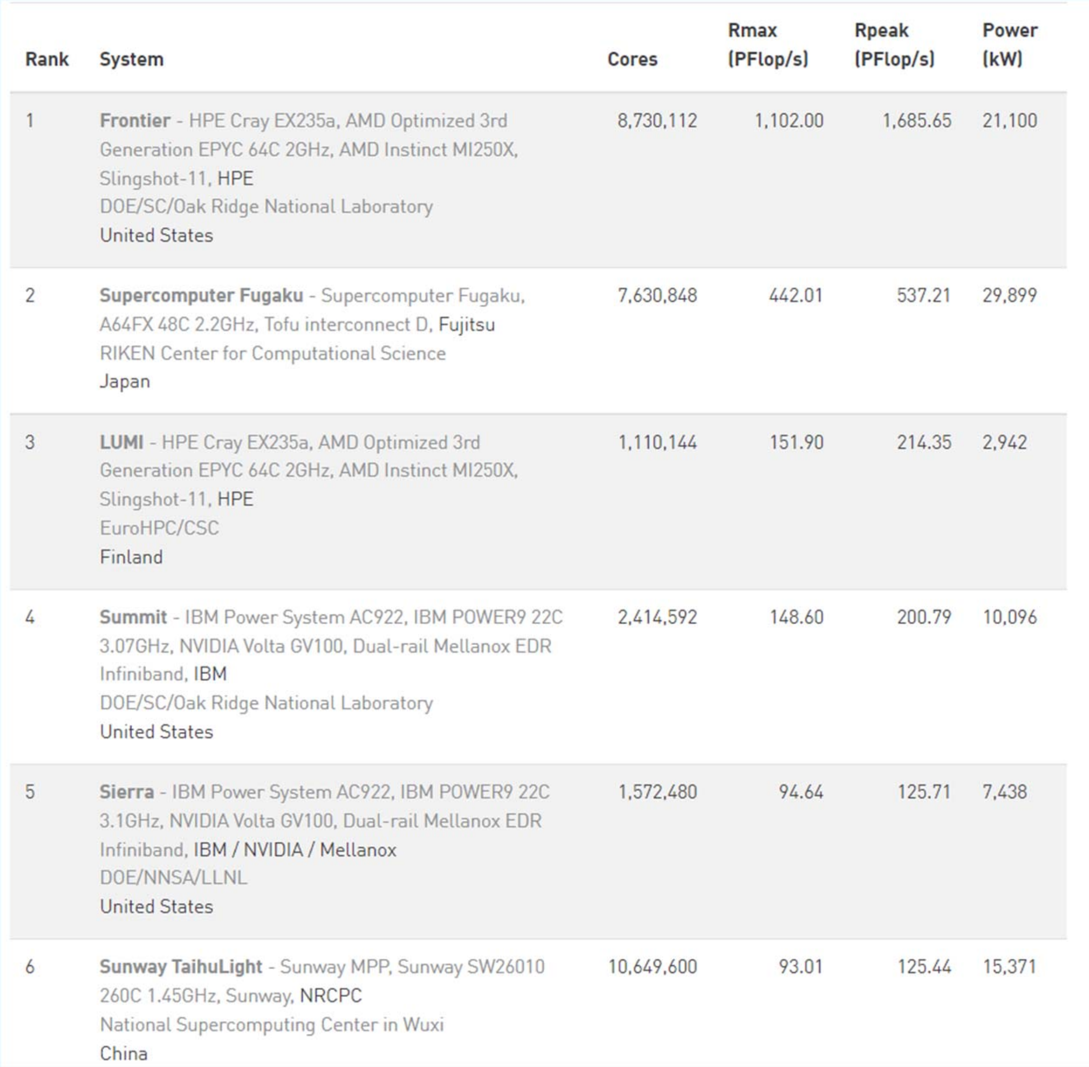
1-100 | 101-200 | 201-300 | 301-400 | 401-500

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	**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	28,192.0	1,350
5	**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
6	**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
7	**Trinity** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,968	14,137.3	43,902.6	3,844

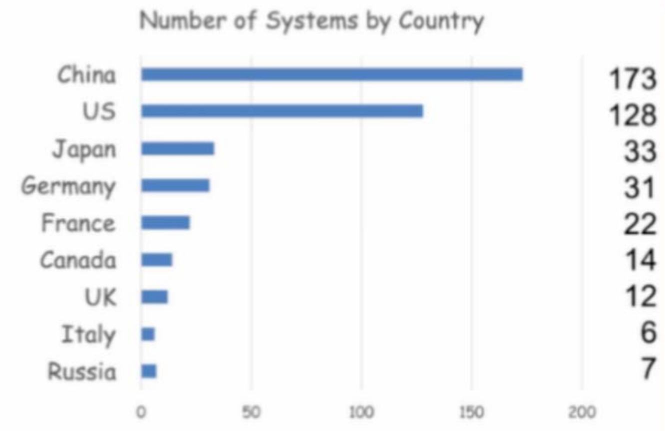


Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
2	**Sierra** - IBM Power System 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	

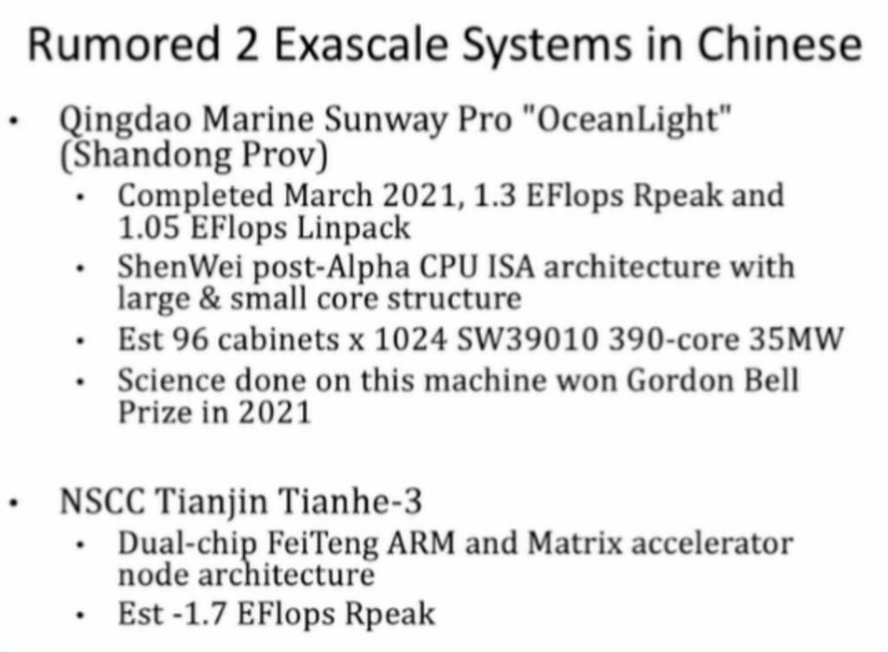
2024/6/5



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	1,110,144	151.90	214.35	2,942
4	**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
5	**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
6	**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



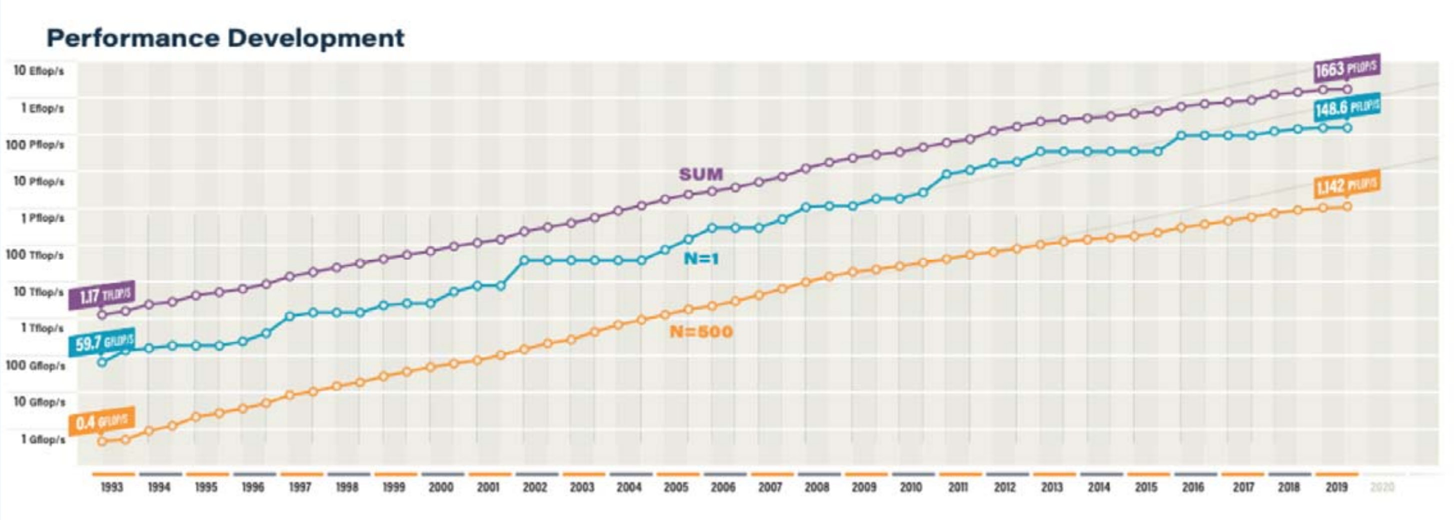
China: Top consumer and producer overall.

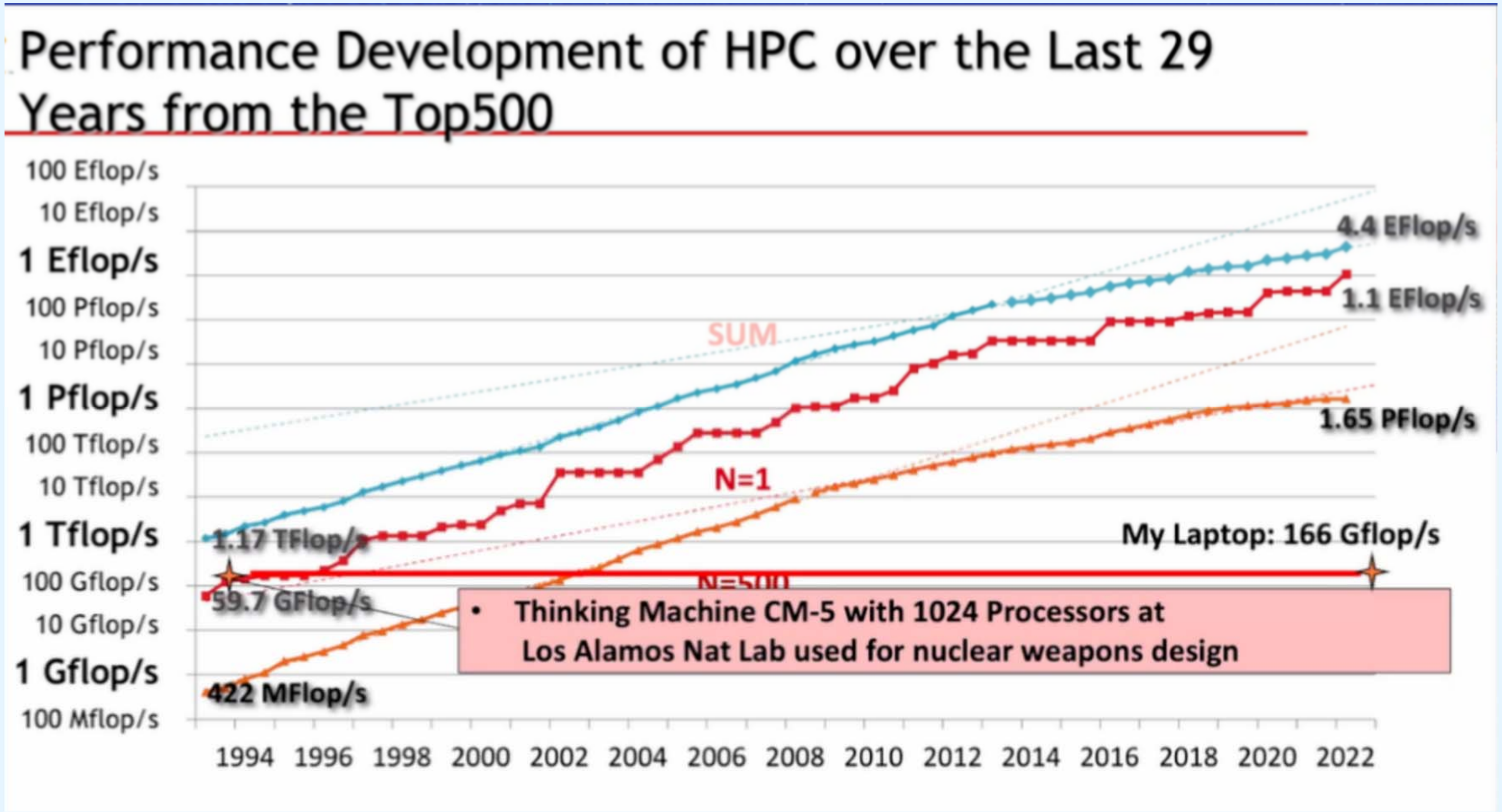


Rumored 2 Exascale Systems in Chinese

- Qingdao Marine Sunway Pro "OceanLight" (Shandong Prov)
  - Completed March 2021, 1.3 EFlops Rpeak and 1.05 EFlops Linpack
  - ShenWei post-Alpha CPU ISA architecture with large & small core structure
  - Est 96 cabinets x 1024 SW39010 390-core 35MW
  - Science done on this machine won Gordon Bell Prize in 2021
- NSCC Tianjin Tianhe-3
  - Dual-chip FeiTeng ARM and Matrix accelerator node architecture
  - Est -1.7 EFlops Rpeak

From Jack Dongarra 2022


Performance Development
10 Eflop/s
1 Eflop/s
100 Pflop/s
10 Pflop/s
1 Pflop/s
100 Tflop/s
10 Tflop/s
1 Tflop/s
100 Gflop/s
10 Gflop/s
1 Gflop/s
SUM
N=1
N=500
1663 PFLOP/S
148.6 PFLOP/S
1.142 PFLOP/S
1.17 TFLOP/S
59.7 GFLOP/S
0.4 GFLOP/S
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

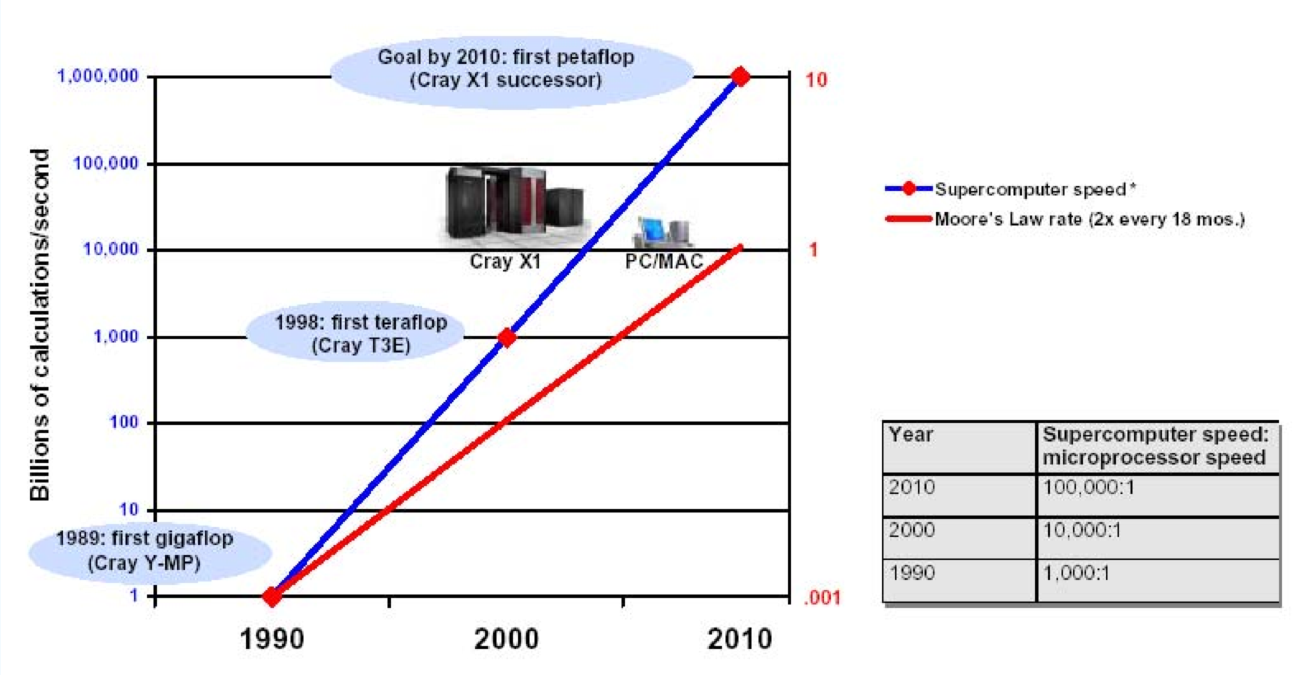


Performance Development of HPC over the Last 29 Years from the Top500

- Thinking Machine CM-5 with 1024 Processors at Los Alamos Nat Lab used for nuclear weapons design

From Jack Dongarra 2022

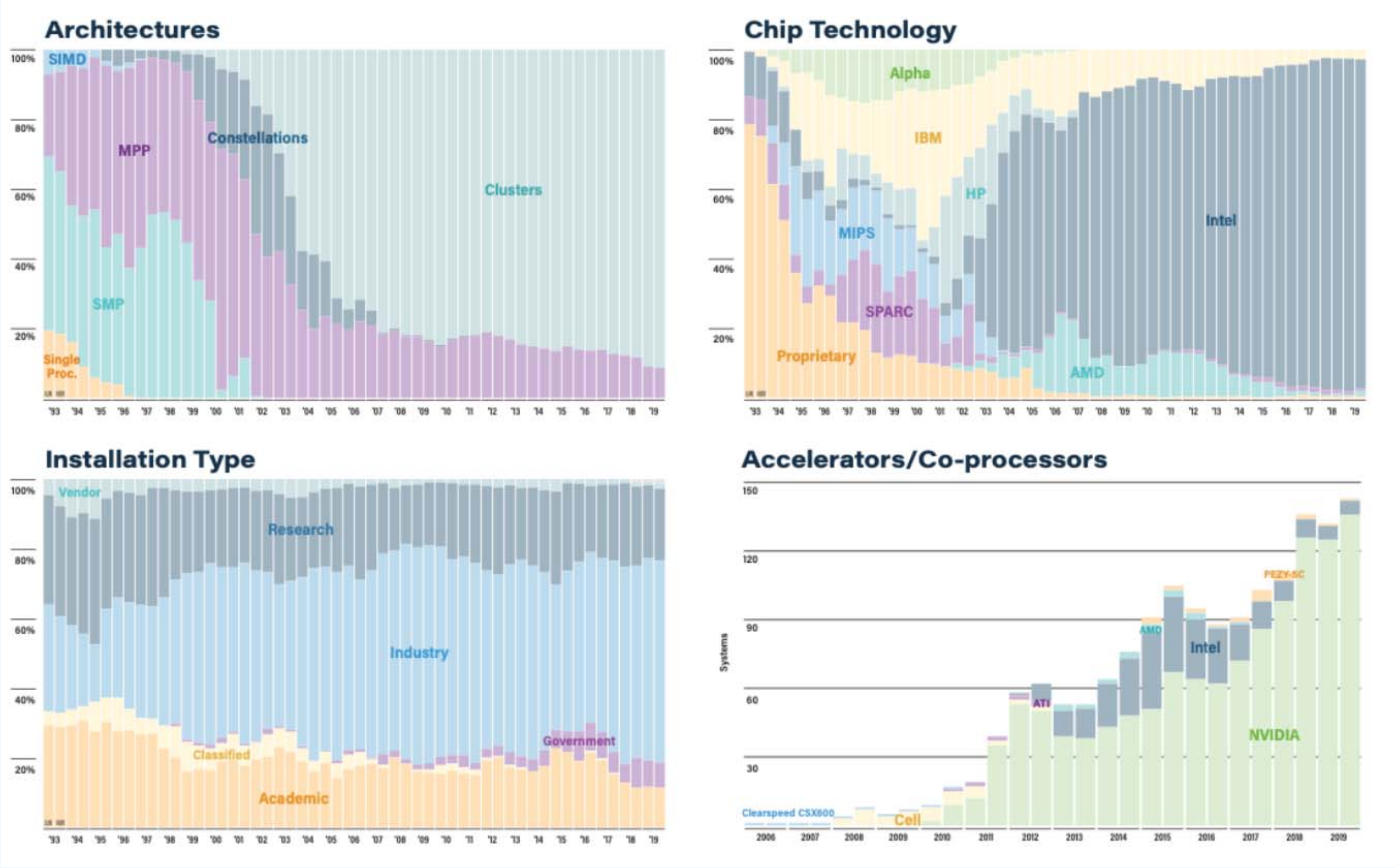
超级计算机的增长速度——超过摩尔定律100倍



Year	Supercomputer speed: microprocessor speed
2010	100,000:1
2000	10,000:1
1990	1,000:1

练习：2022年超算与微机的速度比例如何?

Cite from CRAY Inc.



Architectures

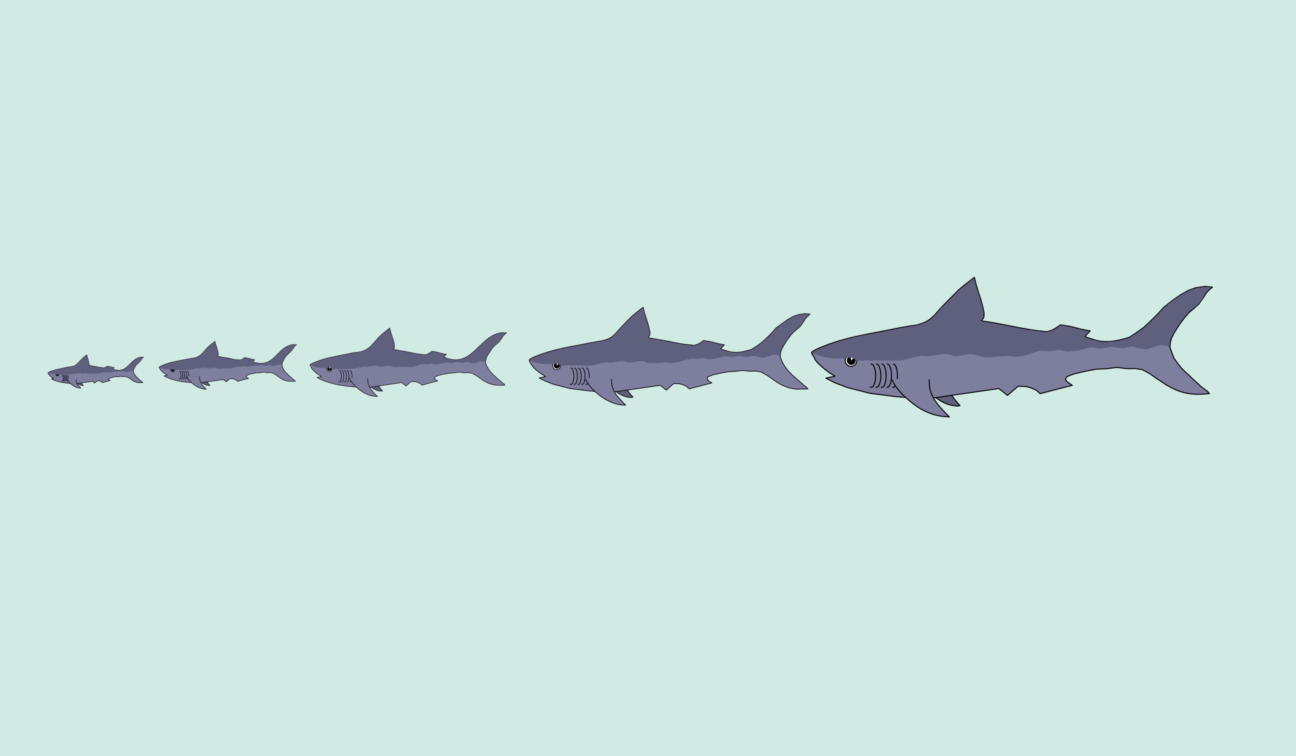
Chip Technology

Installation Type

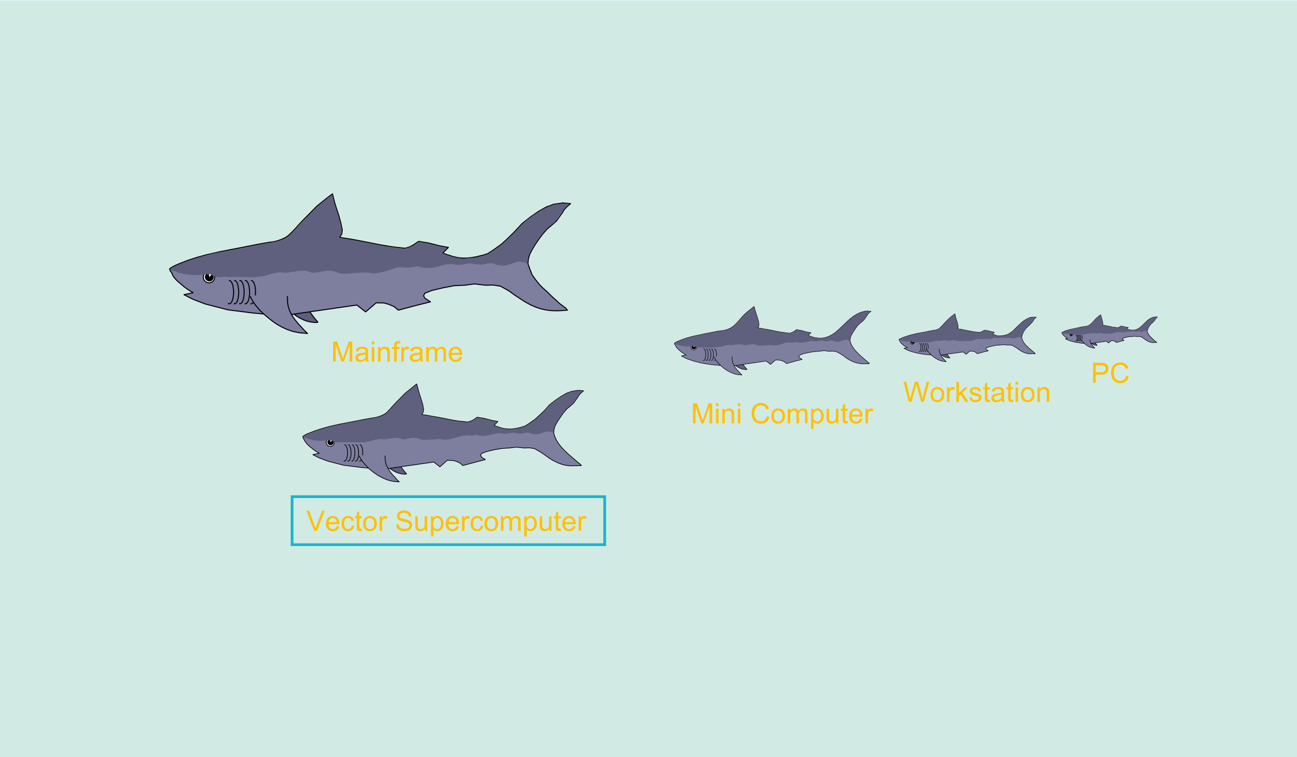
Accelerators/Co-processors

http://now.cs.berkeley.edu

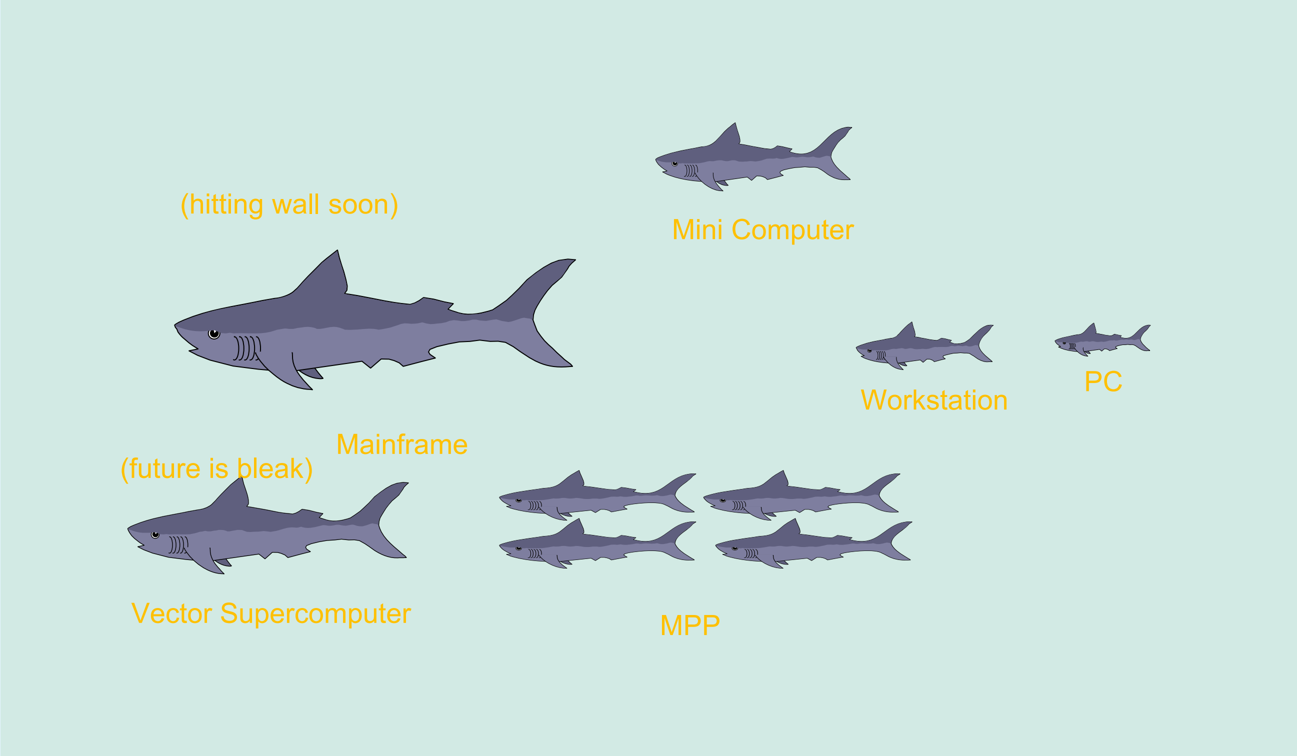
Trends? - 动物食物链



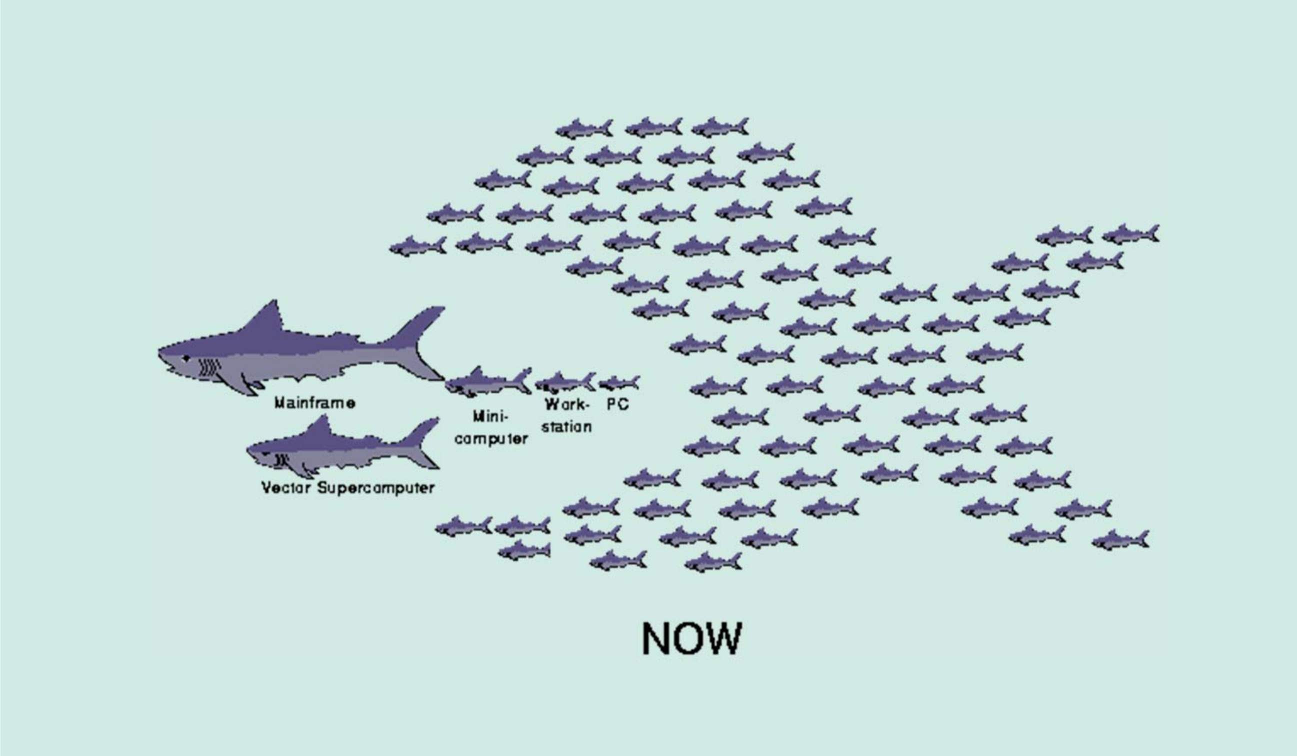
Trends? - 1984计算机食物链



Trends? - 1994计算机食物链



Trends？ - 计算机食物链(现在和将来)

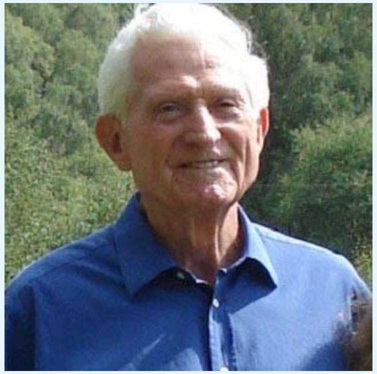


http://now.cs.berkeley.edu

并行加速比定律

- Amdahl定律
- Gustafson定律
- Sun和Ni定律

Gene Amdahl (1922 — 2015)



Famous for formulating a computer science concept known as Amdahl's Law (1967) and for establishing a major IT company called the Amdahl Corporation, Amdahl is also notable for his work with IBM.

Amdahl 定律（1）

- P：处理器数；
- W：问题规模（计算负载、工作负载，给定问题的总计算量）；
  - W_s：应用程序中的串行分量，f是串行分量比例（f = W_s/W，W_s=W_1）；
  - W_P：应用程序中可并行化部分，1-f为并行分量比例；
  - W_s +W $_p$ =W；
- T_s=T_1：串行执行时间，T $_p$：并行执行时间；
- S：加速比，E：效率；
- 出发点：Base on Fixed Problem Size
  - 固定不变的计算负载；
  - 固定的计算负载分布在多个处理器上的，
  - 增加处理器加快执行速度，从而达到了加速的目的。

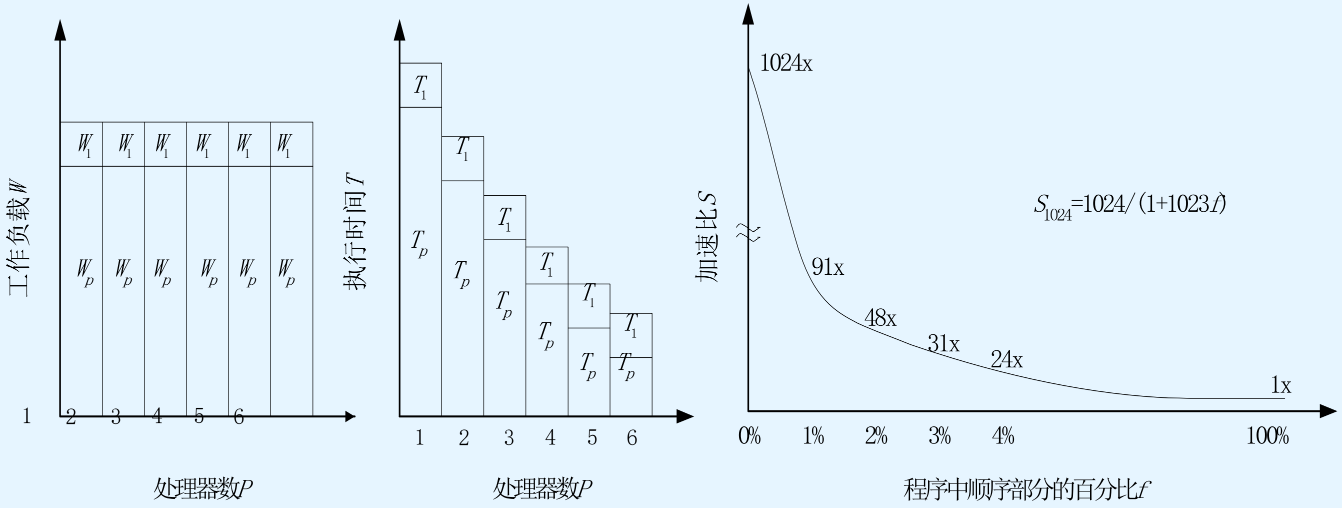
Amdahl定律（2）

- Amdahl’s Law 表明:
  - 适用于实时应用问题。当问题的计算负载或规模固定时，我们必须通过增加处理器数目来降低计算时间;
  - 加速比受到算法中串行工作量的限制。
  - Amdahl’s law: argument against massively parallel systems
- 公式推导

$$T_s = fW + (1-f)W \qquad T_p = fW + \frac{(1-f)W}{p}$$

$$S_p = \frac{W}{fW + \frac{(1-f)W}{p}} = \frac{p}{pf + 1 - f} = \frac{1}{\frac{(p-1)f+1}{p}} \xrightarrow[p\to\infty]{} \frac{1}{f}$$

Amdahl定律（3）

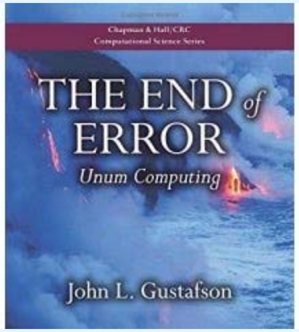


(a)

(b)

(c)

John Gustafson (1955—)



Dr. Gustafson is an American computer scientist and businessman, chiefly known for his work in High Performance Computing (HPC) such as the invention of Gustafson's Law, introducing the first commercial computer cluster, etc.

Gustafson定律（1）

- 出发点：Base on Fixed Execution Time
  - 对于很多大型计算，精度要求很高，即在此类应用中精度是个关键因素，而计算时间是固定不变的。此时为了提高精度，必须加大计算量，相应地亦必须增多处理器数才能维持时间不变；
  - 表明：随着处理器数目的增加，串行执行部分*f*不再是并行算法的瓶颈。
- Gustafson加速定律：

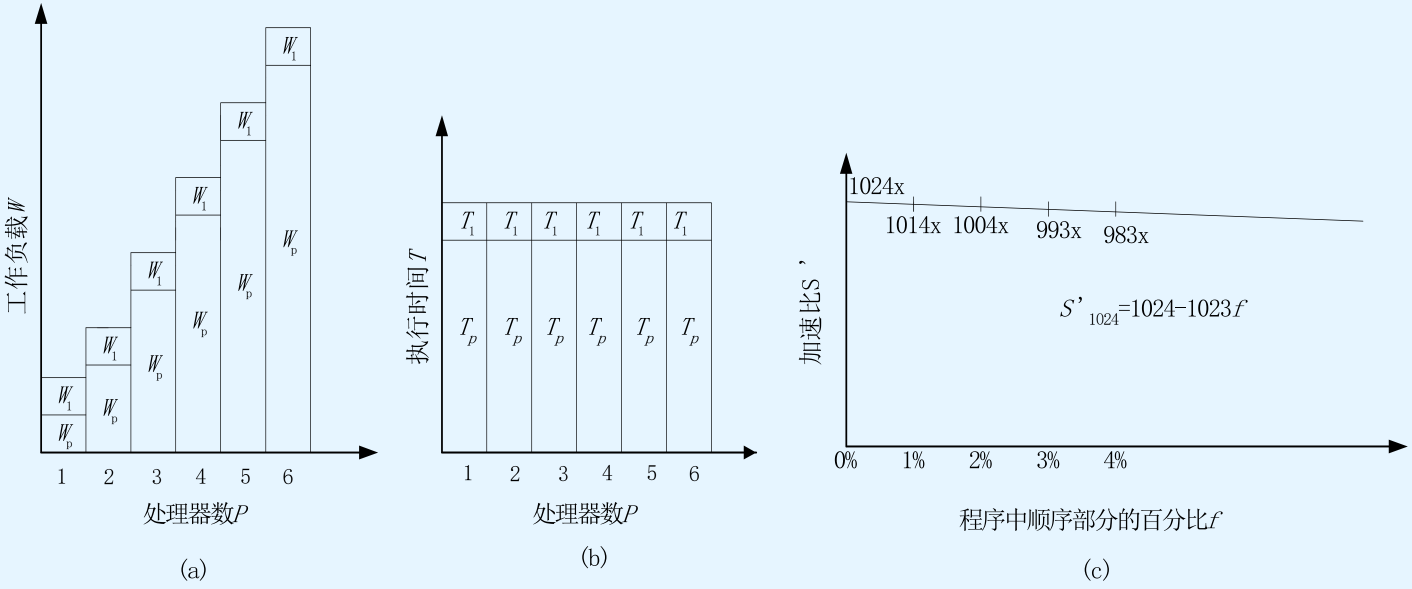
$$S' = \frac{W_S + pWp}{W_S + p \cdot Wp / p} = \frac{W_S + pWp}{W_S + W_P}$$

$$S' = f + p(1\text{-}f) = p + f(1\text{-}p) = p\text{-}f(p\text{-}1)$$

- 并行开销W$_O$：

$$S' = \frac{W_S + pW_P}{W_S + W_P + W_O} = \frac{f + p(1 - f)}{1 + W_O / W}$$

Gustafson定律（2）



孙贤和（Xian-He Sun）、倪明选（Lionel M. Ni）

Sun 和 Ni定律 （1）

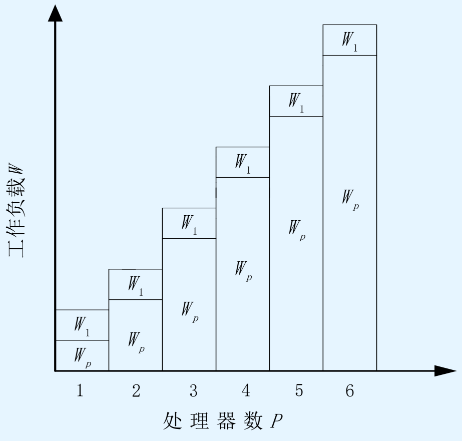
- 出发点： Base on Memory Bounding
  - 充分利用存储空间等计算资源，尽量增大问题规模以产生更好/更精确的解。是Amdahl定律和Gustafson定律的推广。
- 公式推导：
  - 设单机上的存储器容量为M，其工作负载W=fW+(1-f)W
    当并行系统有*p*个结点时，存储容量扩大了pM，用G(p)表示系统的存储容量增加p倍时工作负载的增加量。则存储容量扩大后的工作负载为W=fW+(1-f)G(p)W，所以存储受限的加速为

$$S'' = \frac{fW+(1-f)G(p)W}{fW+(1-f)G(p)W/p} = \frac{f+(1-f)G(p)}{f+(1-f)G(p)/p}$$

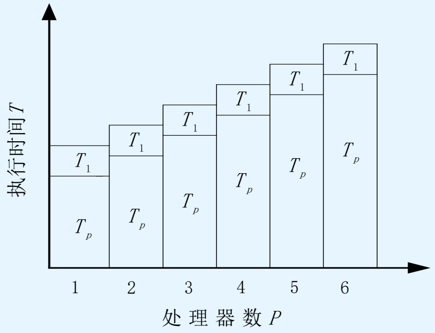
- 并行开销W_o:

$$S'' = \frac{fW+(1-f)WG(p)}{fW+(1-f)G(p)W/p+W_O} = \frac{f+(1-f)G(p)}{f+(1-f)G(p)/p+W_O/W}$$

Sun 和 Ni定律（2）



(a)



(b)

- G（p）=1时就是Amdahl加速定律；
- G（p）=p 变为 f + p（1-f），就是Gustafson加速定律
- G（p）>p时，相应于计算机负载比存储要求增加得快，此时Sun和 N i 加速均比 Amdahl 加速和 Gustafson 加速为高。

加速比讨论

- 参考的加速经验公式： $p/\log p \le S \le P$
  - 线性加速比：很少通信开销的矩阵相加、内积运算等
  - p/log p的加速比：分治类的应用问题

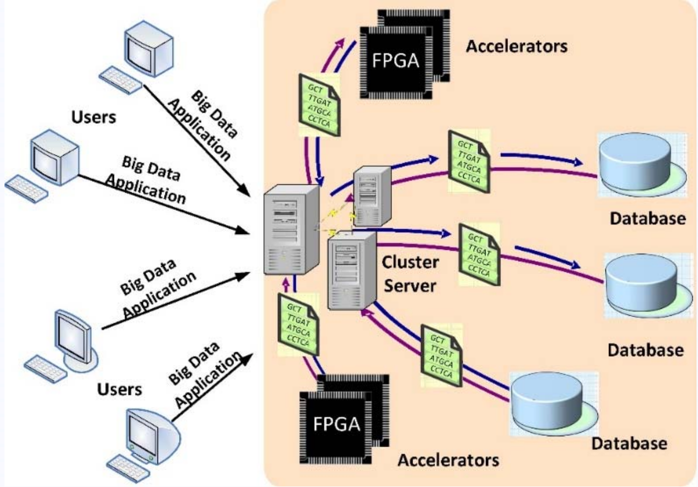
- 通信密集类的应用问题 ： $S = 1 / C(p)$

  这里C(p)是p个处理器的某一通信函数
- 超线性加速

- 绝对加速：最佳串行算法与并行算法
- 相对加速：同一算法在单机和并行机的运行时间

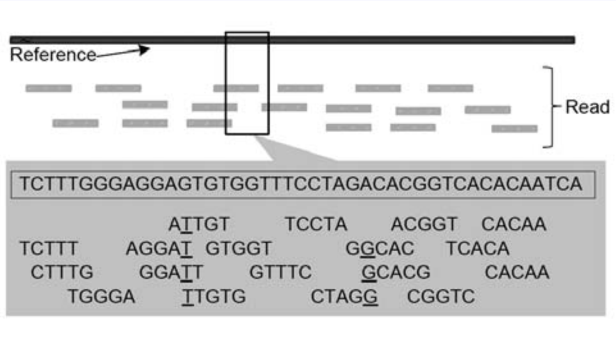
研究工作示例

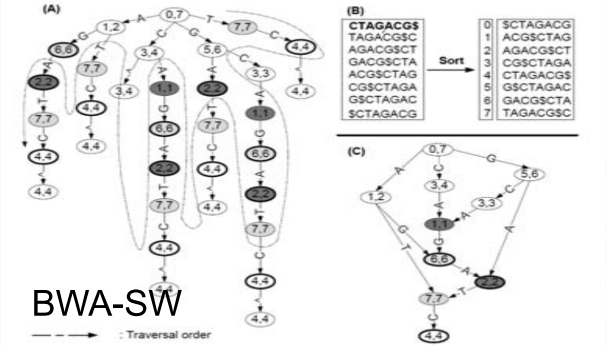
- 面向大数据应用的加速器设计



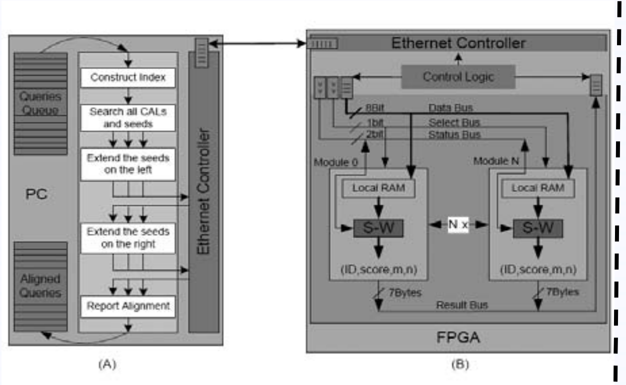
- 基因测序加速
- 深度学习预测过程加速
- 机器学习算法加速
- 大规模图(>10^7节点, 10^9边)计算的加速器
  1. 单加速器针对GPU 1~2倍；
  2. 功耗为GPU的5%左右。

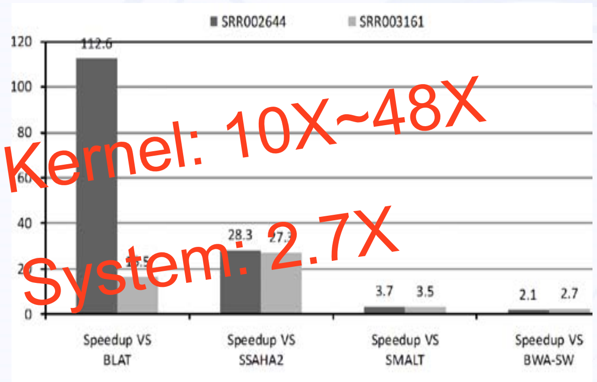
典型的大数据应用加速


Reference
Read
TCTTTGGGAGGAGTGTGGTTTCCTAGACACGGTCACACAATCA
ATTGT TCCTA ACGGT CACAA
TCTTT AGGAT GTGGT GGCAC TCACA
CTTTG GGATT GTTTC GCACG CACAA
TGGGA TTGTG CTAGG CGGTC



BWA-SW





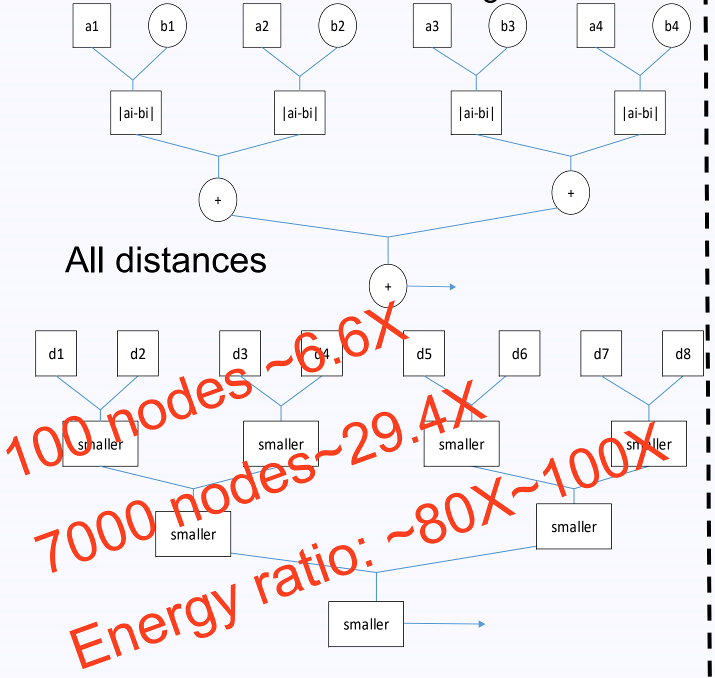
Kernel: 10X~48X

System: 2.7X

Chao Wang, Xi Li, Peng Chen, Xuehai Zhou,, “Heterogeneous Cloud Framework for Big Data Genome Sequencing”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.{封面亮点文章}
Peng Chen, Chao Wang, Xi Li, Xuehai Zhou, “Accelerating the Next Generation long read mapping with the FPGA-based system”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics.*

典型的大数据应用加速

K-Means in Machine Learning



All distances

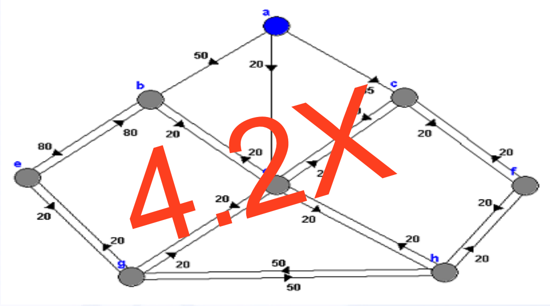
100 nodes ~6.6X

7000 nodes~29.4X

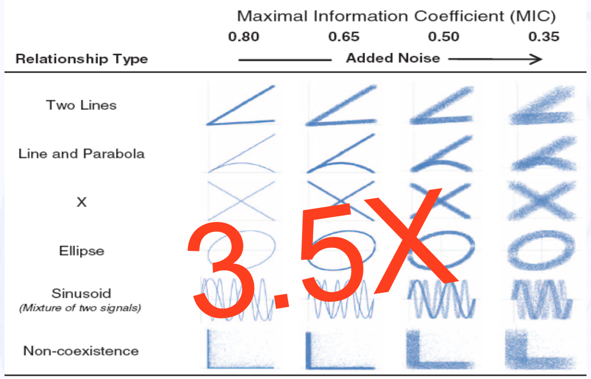
Energy ratio: ~80X~100X

the shortest distance

CSPF in Software Defined Network



4.2X

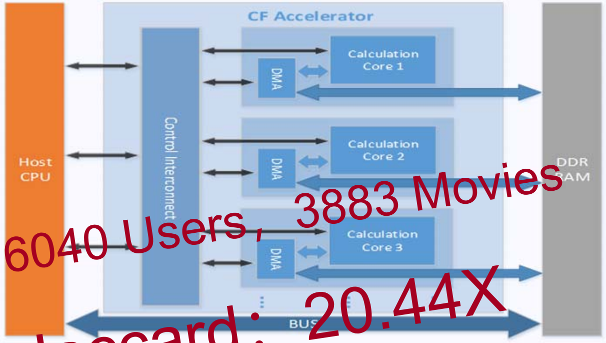


3.5X

Maximal Information Coefficient (MIC)

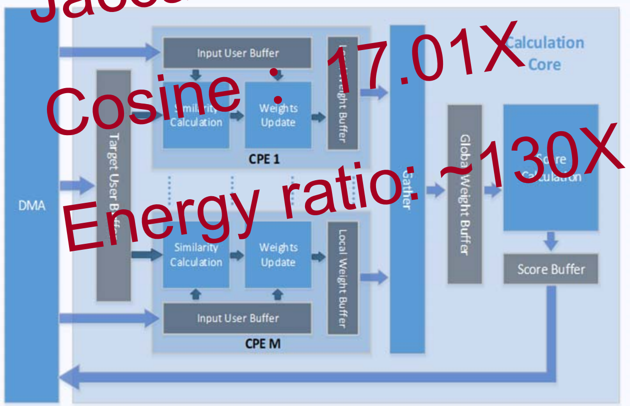
典型的大数据应用加速

Recommendation System @Xilinx Zynq



6040 Users，3883 Movies

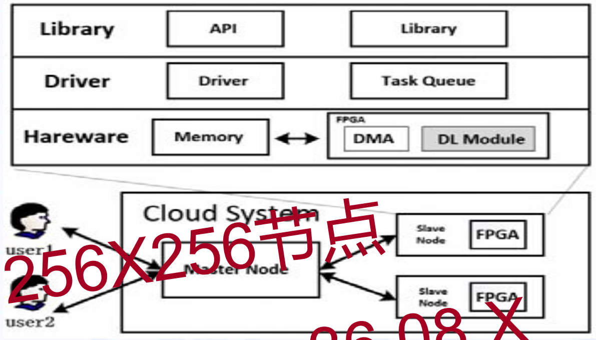
Jaccard：20.44X



Cosine ： 17.01X

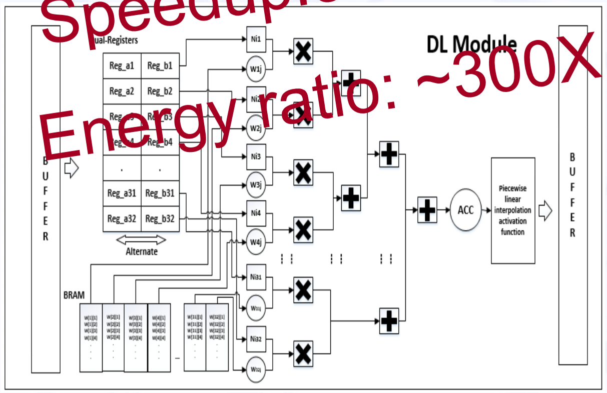
Energy ratio: ~130X

Deep Learning Engine @Xilinx Zynq



256X256节点

Speedup:36.08 X



Energy ratio: ~300X