



National Astronomical Observatories, CAS



RECRUITMENT  
PROGRAM OF GLOBAL EXPERTS

UNIVERSITÄT  
HEIDELBERG  
Zukunft. Seit 1386.



# Introduction to GPU

## Accelerated Computing:

### 1. History of Computer Architecture

### Many-Core, GPU, and other ideas...

University

Rainer Spurzem

Astronomisches Rechen-Inst., ZAH, Univ. of Heidelberg, Germany

National Astronomical Observatories (NAOC), Chinese Academy of Sciences

Kavli Institute for Astronomy and Astrophysics (KIAA), Peking University

The SILK ROAD PROJECT at NAOC/KIAA

丝绸之路 计划

[spurzem@ari.uni-heidelberg.de](mailto:spurzem@ari.uni-heidelberg.de)

<http://silkroad.bao.ac.cn>



北京大学  
PEKING UNIVERSITY

# Introduction to GPU Accelerated Computing

## February 10 – 13, 2020

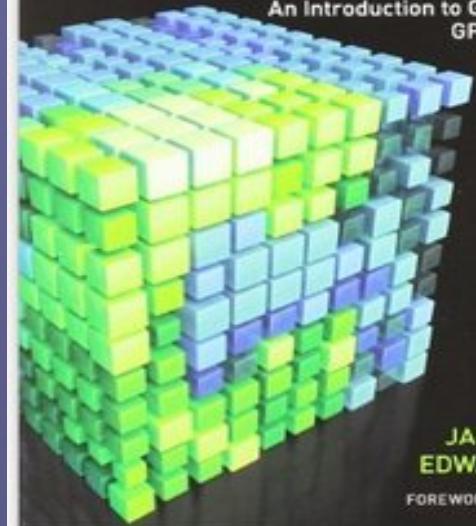
### Table of Contents (subject to adjustment/change):

1. Monday morning: General Introduction Computer Architecture, Many-Core, GPU and others..., Access...
2. Monday afternoon: Access to kepler, CUDA Hello, GPU Properties, Simple Add, Vector Add
3. Tuesday morning: More on GPU Software and Hardware
4. Tuesday afternoon: CUDA More Vector Add, Scalar Products, Using Blocks and Threads
5. Wednesday morning: Parallelization and Amdahl's Law, GPU Acceleration, Future Architecture
6. Wednesday Afternoon: Events, Histograms, Matrix Multiplication
7. Thursday Morning: Astrophysical N-Body Code
8. Thursday Afternoon: Astrophysical Parallel N-Body Code Using MPI and GPU
9. Access: Use **ssh-keygen -t rsa** (give passphrase)  
Send **id\_rsa.pub** to **spurzem@ari.uni-heidelberg.de**



# CUDA BY EXAMPLE

An Introduction to General-Purpose GPU Programming



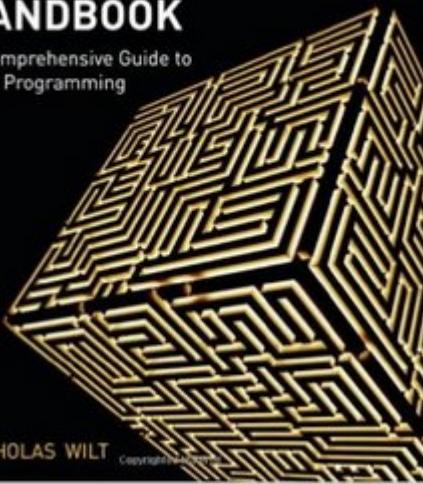
JASON SANDERS  
EDWARD KANDROT

FOREWORD BY JACK DONGARRA

# Literature

## THE CUDA HANDBOOK

A Comprehensive Guide to GPU Programming



NICHOLAS WILT

David B. Kirk  
Wen-mei W. Hwu

## SECOND EDITION Programming Massively Parallel Processors

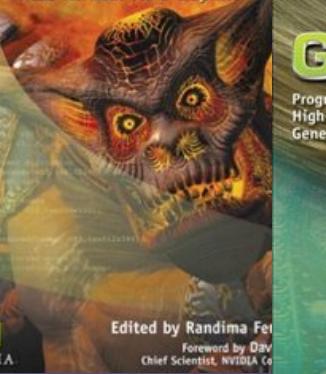
A Hands-on Approach



Copyright Material

## GPU Gems

Programming Techniques, Tips, and Tricks for Real-Time Graphics



Edited by Randima Fernando  
Foreword by David S. Ebert, NVIDIA CTO

## GPU Gems 2

Programming Techniques for High-Performance Graphics and General-Purpose Computation



Edited by Matt Pharr

Foreword by Tim Sweeney, Epic Games

Randima Fernando, Series Editor

## GPU Gems 3



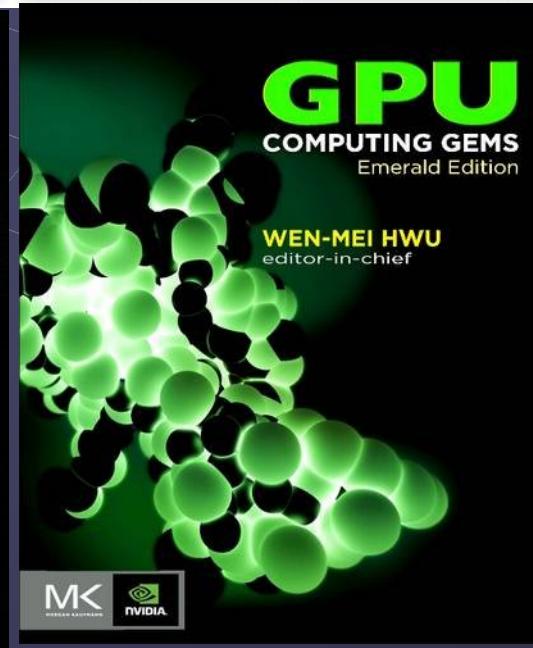
Edited by Hubert Nguyen

Foreword by Kurt Akeley, Microsoft Research

## GPU COMPUTING GEMS

Emerald Edition

WEN-MEI HWU  
editor-in-chief





Observations (Experiment)



Theory



Computational Physics



# GPU Computing

## History

# History

Erik Holmberg (1908-2000)

Dissertation Univ. Lund (Schweden) (1937):

“A study of double and multiple galaxies”

Galaxies often in Groups and Pairs

Irregular Distribution of Satellite Galaxies  
(Holmberg-Effect)

**Father of numerical astrophysics?**

» ...with 200 light bulbs





# History

<http://cdsads.u-strasbg.fr/abs/1941ApJ....94..385H>

The Astrophysical Journal, Nov. 1941

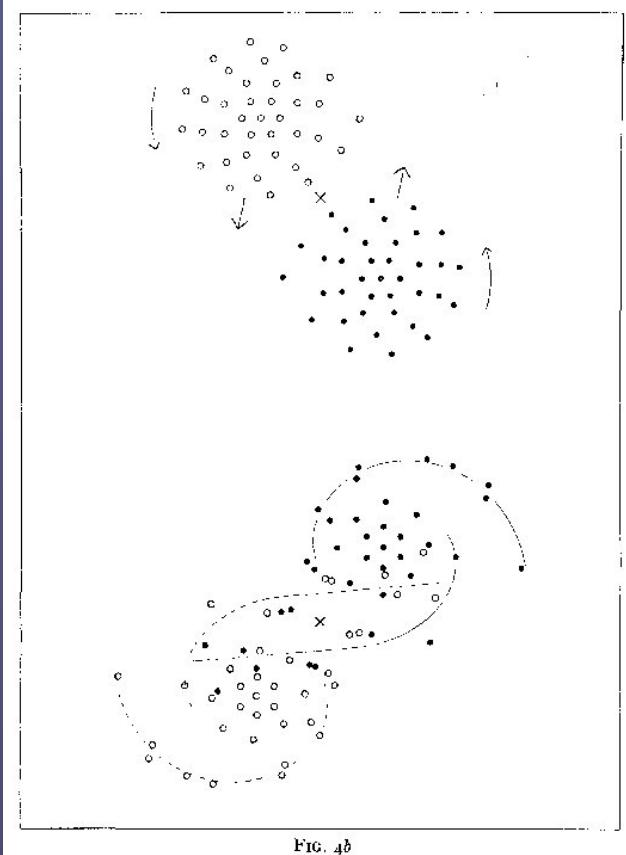


FIG. 4b

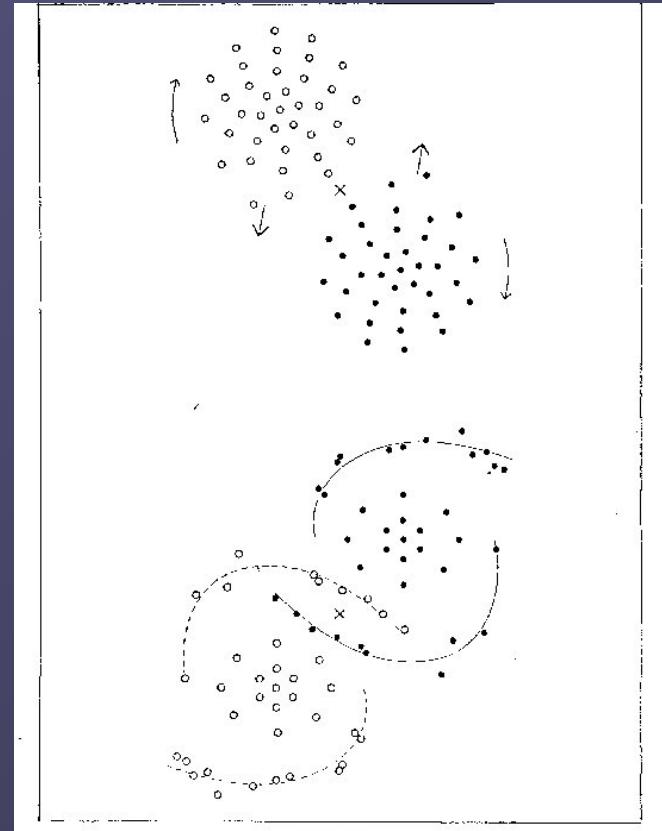


FIG. 4a

# HARDWARE

...before von Neumann...

● Konrad Zuse (1910-1995) Berlin



Invented freely programmable Computer

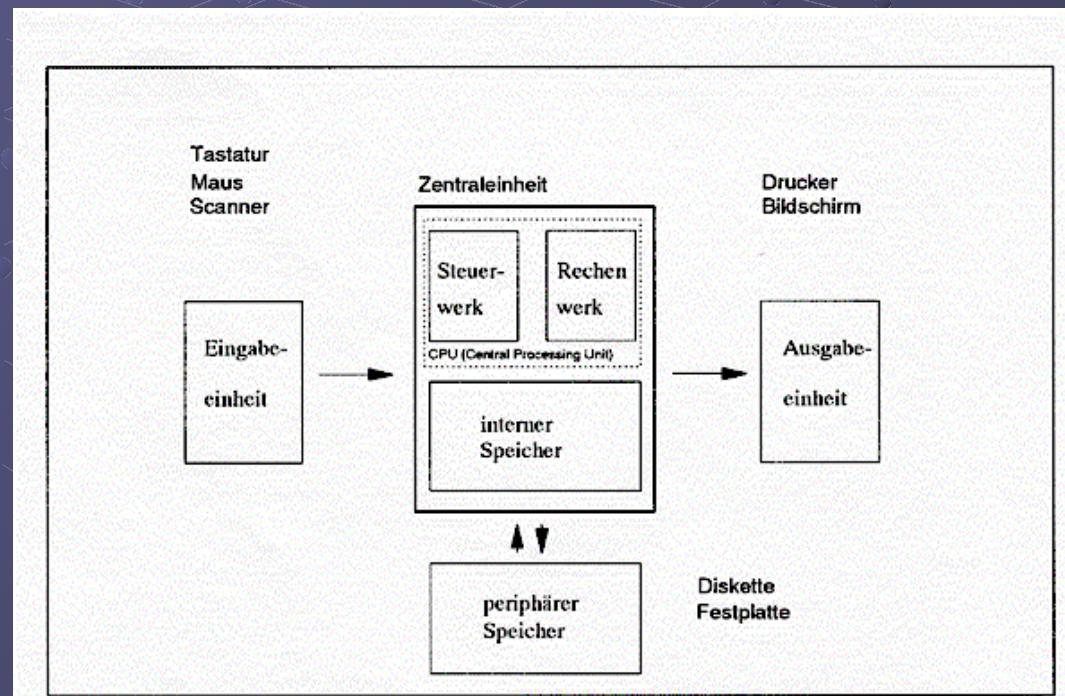


Z1 in parental flat 1936

# HARDWARE

- John von Neumann (1903-1957)

Born Budapest, Lecturer Berlin, since 1930 Princeton Univ.  
Requirements for the Construction of an electronic computing device(1946)

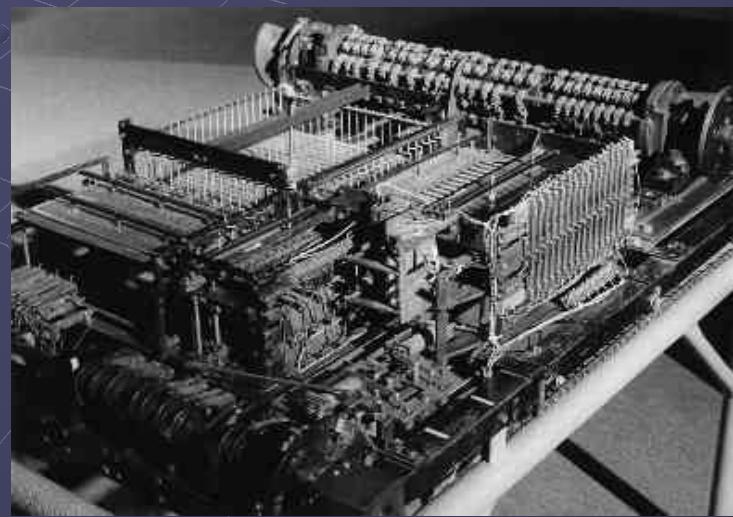


# History

Zuse Z4: 1944 Berlin, 1950 Zürich, 1954 Frankreich  
1959 Deutsches Museum München



Computing Speed 0.03 MHz



Memory 256 byte





Astronomisches  
Rechen-Institut (ARI)  
at Univ. of  
Heidelberg, Germany



**Siemens 2002  
Computer in 1964  
At ARI**



# History

<http://cdsads.u-strasbg.fr/abs/1960ZA.....50..184V>

Astronomisches Rechen-Institut in Heidelberg  
Mitteilungen Serie A Nr. 14

## Die numerische Integration des *n*-Körper-Problemes für Sternhaufen I

Von

**SEBASTIAN VON HOERNER**

Mit 3 Textabbildungen

(Eingegangen am 10. Mai 1960)

Tabelle 5. Zahl der gegenseitigen Umläufe,  
Häufigkeit des Auftretens und kleinster  
gegenseitiger Abstand  $D_m$  der engsten Paare.  
(Alle engsten Paare mit mehr als zwei  
vollen Umläufen wurden notiert)

Umläufe	Häufigkeit	$D_m$
2—3	11	0.0102
3—5	9	0.0177
5—10	5	0.0070
10—20	2	0.0141
20—50	1	0.0007
50—100	1	0.0035
100—200	1	0.0039

Astronomisches Rechen-Institut in Heidelberg  
Mitteilungen Serie A Nr. 19

## Die numerische Integration des *n*-Körper-Problems für Sternhaufen, II.

Von

**SEBASTIAN VON HOERNER**

Mit 10 Textabbildungen

(Eingegangen am 19. November 1962)

S.v. Hoerner,  
Z.f.Astroph. 1960, 63

Siemens 2002  
N=4,8,12,16 (4 Trx)

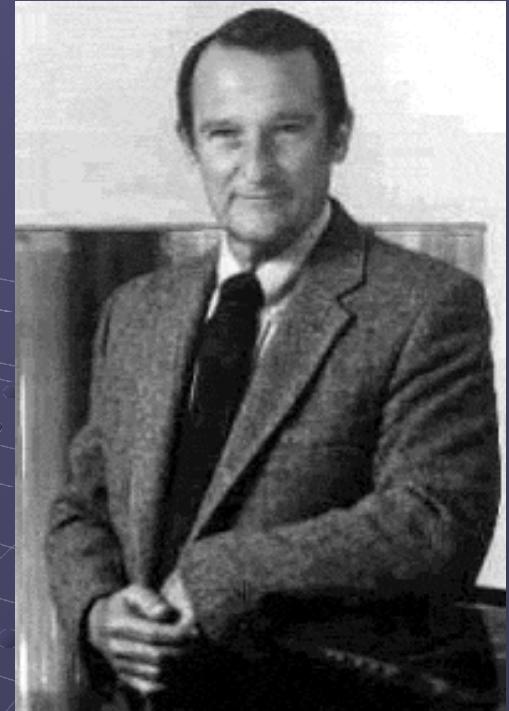
N=16,25 (40 Trx)

<http://cdsads.u-strasbg.fr/abs/1963ZA.....57...47V>

# History

## ● Seymour Cray (1925-1996)

“father of supercomputing”



- **CRAY1: Vectorregisters (1976)**  
**160 Mflop, 80 MHz, 8 MByte RAM**
- **CRAY2: (1984)**  
**1Gflop, 120MHz, 2GByte RAM**

# History

*Supercomputer  
JUGENE  
IBM Blue Gene  
At FZ Jülich,  
Germany*



*Opening Ceremony June 2008*



# Computational Science...

Exaflop/s?

...after von Neumann...

Petaflop/s

Teraflop/s

GigaFlop/s

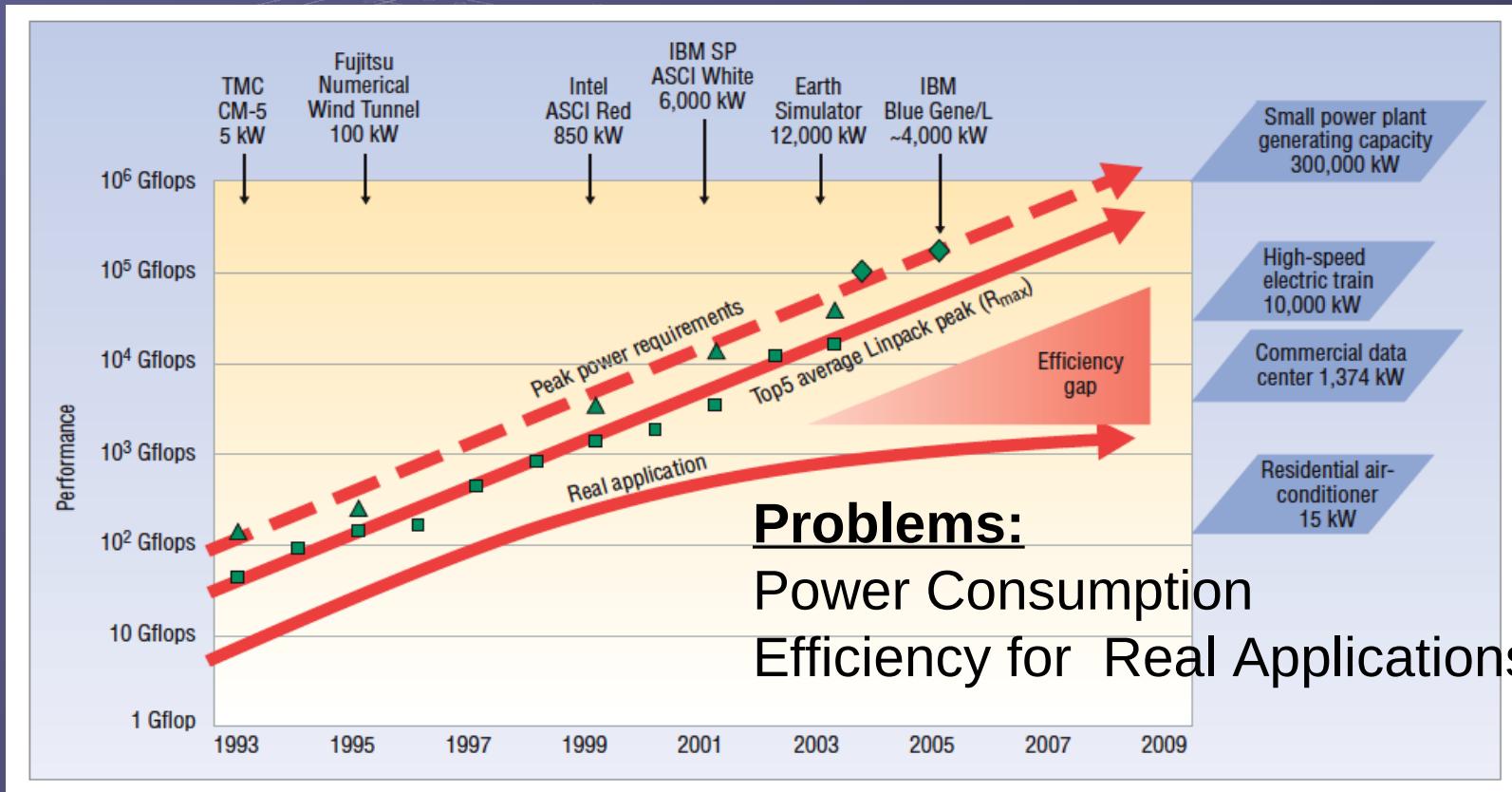


Figure 1. Rising power requirements. Peak power consumption of the top supercomputers has steadily increased over the past 15 years.

Thanks to Horst Simon, LBNL/NERSC for this diagram.

# GPU Computing

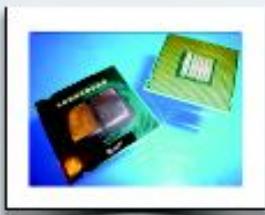
# Special Hardware

# Accelerators

# SPECIAL HARDWARE

## CPUs

Central Processing Units



General Purpose oriented

1-12 Cores

Up to 4 pipes per core using Vector Units

Fully Programmable, many languages available

Very well studied

Max. 125W per processor

## GPUs

Graphic Processing Units



Graphics oriented

16-512 Cores

Massively Parallel Architecture, specialized instructions for parallel processing

Fully programmable, but limited languages

Algorithms not fully explored

Max. 400W per card

## FPGAs

Field Programmable Gate Arrays



Custom designs, best for processing streaming data

Programmable Logic, Architecture is custom-built for the required application

Requires extensive knowledge to program, development time is longer than CPUs and GPUs

Application interface is custom built on each case

Max. 60W per FPGA

## ASICs

Application Specific Integrated Circuits



Fully custom designs, built for a specific application

Not flexible, cannot be changed once it is built

Development is even more specialized than FPGAs

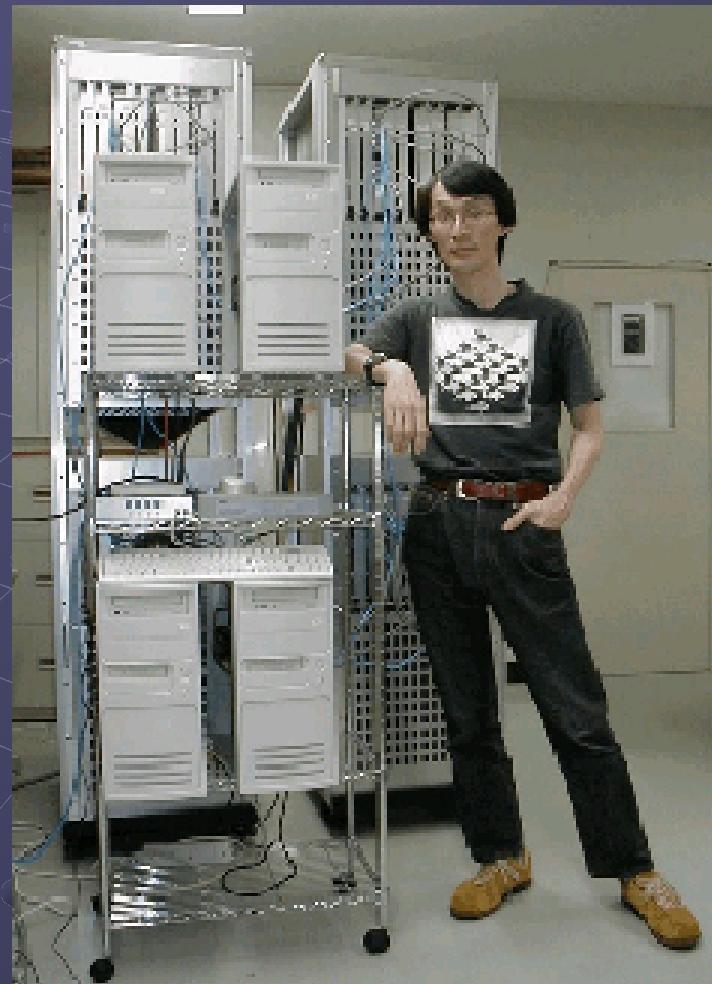
Power consumption varies with the application, usually best performance per Watt

Slide: Guillermo Marcus

# HARDWARE

## GRAPE-6 Gravity/Coulomb Part

- G6 Chip:  $0.25\mu$  2MGate ASIC, 6 Pipelines
- at 90MHz, 31Gflops/chip
- 48Tflops full system (March 2002)
- Plan up to 72Tflops full system (in 2002)
- Installed in Cambridge, Marseille, Drexel, Amsterdam, New York (AMNH), Mitaka (NAO), Tokyo, etc.. New Jersey, Indiana, Heidelberg



## GRAPE-6



1998, 120 Gflops

Developers: Junichiro Makino, Toshiyuki Fukushige, Hiroshi Daisaka, Eiichiro Kokubo, Masaki Koga, Makoto Taiji, Ken Namura

[GRAPE-6: Massively-Parallel Special-Purpose Computer for Astrophysical Particle Simulations](#)

[Sales information](#)

## The Green500 List - November 2010

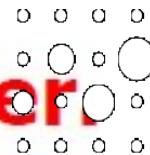
Listed below are the November 2010 The Green500's energy-efficient supercomputers ranked from 1 to 100.

<http://www.green500.org>

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	1684.20	IBM Thomas J. Watson Research Center	NNSA/SC Blue Gene/Q Prototype	38.80
2+	1448.03	National Astronomical Observatory of Japan	GRAPE-DR accelerator Cluster, Infiniband	24.59
2	958.35	GSIC Center, Tokyo Institute of Technology	HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows	1243.80
3	933.06	NCSA	Hybrid Cluster Core i3 2.93Ghz Dual Core, NVIDIA C2050, Infiniband	36.00

# GPU: NAOC laohu cluster Beijing, China





### Kepler GPU cluster

**12 nodes = 12 x 16 = 192 CPU cores (@ 2 GHz)**

**12 x 64 GB = 768 GB RAM CPU memory**

**12 GPUs K20m = 12 x 2496 ~ 30k GPU threads**

**12 x 4.8 GB ~ 57 GB GPU device memory**

**4 x Xilinx Virtex-6 FPGA (ML 605)**

**since beg. 2013 operated.**



# NVIDIA Volta V100 GPU, 21 billion transistors, 5120 cores



With NVLINK

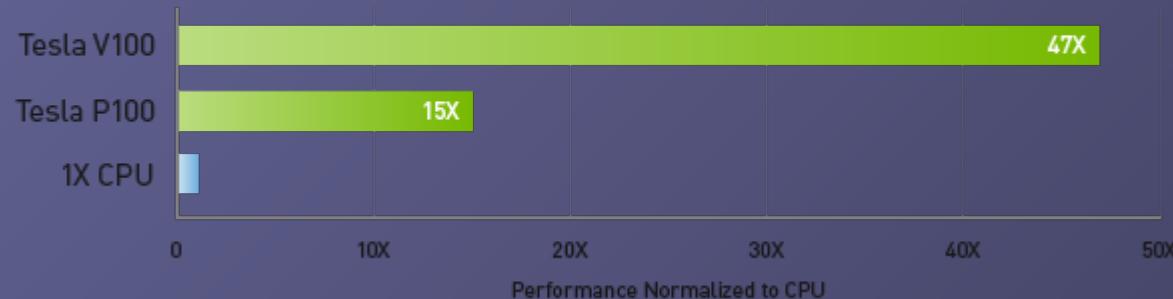


Without NVLINK

PERFORMANCE with NVIDIA GPU Boost™		DOUBLE-PRECISION	DOUBLE-PRECISION
		7.8 teraFLOPS	7 teraFLOPS
		SINGLE-PRECISION	SINGLE-PRECISION
		15.7 teraFLOPS	14 teraFLOPS
		DEEP LEARNING	DEEP LEARNING
		125 teraFLOPS	112 teraFLOPS
INTERCONNECT BANDWIDTH Bi-Directional		NVLINK	PCIE
		300 GB/s	32 GB/s
MEMORY CoWoS Stacked HBM2		CAPACITY	
		32/16 GB HBM2	
POWER Max Consumption		BANDWIDTH	
		900 GB/s	
POWER Max Consumption		300 WATTS	250 WATTS

# NVIDIA Volta V100 GPU, 21 billion transistors, 5120 cores

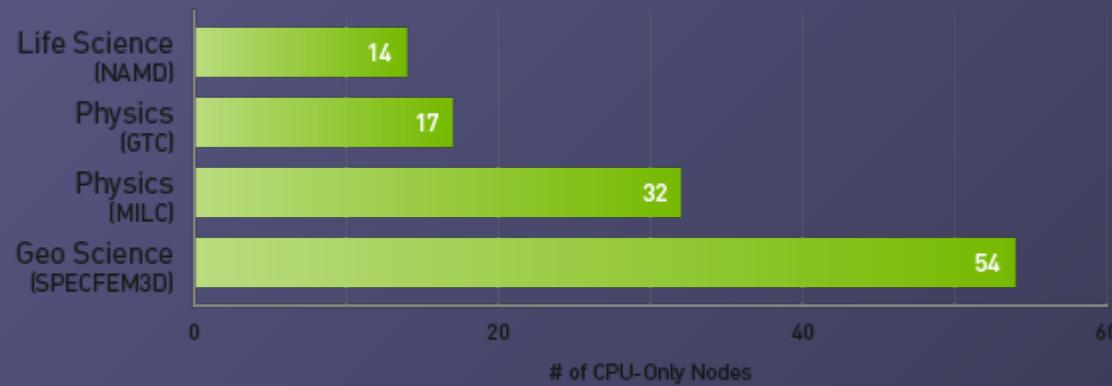
47X Higher Throughput Than CPU Server on Deep Learning Inference



Workload: ResNet-50 | CPU: 1X Xeon E5-2690v4 @ 2.6 GHz | GPU: Add 1X Tesla P100 or V100

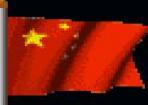
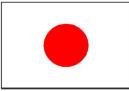
1 GPU Node Replaces Up To 54 CPU Nodes

Node Replacement: HPC Mixed Workload



CPU Server: Dual Xeon Gold 6140@2.30GHz, GPU Servers: same CPU server w/ 4x V100 PCIe | CUDA Version: CUDA 9.x | Dataset: NAMD (STMV), GTC (mpi#proc.in), MILC (APEX Medium), SPECFEM3D (four\_material\_simple\_model) | To arrive at CPU node equivalence, we use measured benchmark with up to 8 CPU nodes. Then we use linear scaling to scale beyond 8 nodes.

# Top 10 List November 2010

1	National Supercomputing Center in Tianjin China		Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C NUDT	<b>GPU</b>
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	<b>GPU</b>
4	GSIC Center, Tokyo Institute of Technology Japan		TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows NEC/HP	<b>GPU</b>
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 bullex 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 Jülich (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.	

From [www.top500.org](http://www.top500.org) - list of fastest supercomputers in the world...  
... last year Nov. 2010:

## ► China Grabs Supercomputing Leadership Spot in Latest Ranking of World's Top 500 Supercomputers

Thu, 2010-11-11 22:42

MANNHEIM, Germany; BERKELEY, Calif.; and KNOXVILLE, Tenn.—The 36<sup>th</sup> edition of the closely watched TOP500 list of the world's most powerful supercomputers confirms the rumored takeover of the top spot by the Chinese Tianhe-1A system at the National Supercomputer Center in Tianjin, achieving a performance level of 2.57 petaflop/s (quadrillions of calculations per second).

# NCSA director: GPU is future of supercomputing

by Brooke Crothers

A

A

Font size



Print



E-mail



Share



6 comments

Tweet

99

share

25

2

Digg

The director of the National Center for Supercomputing Applications has seen the future of supercomputing and it can be summed up in three letters: GPU.

Thom Dunning, who directs the NCSA and the Institute for Advanced Computing Applications and Technologies at the famed supercomputing facilities on the campus of University of Illinois at Urbana-Champaign, says high-performance computing will begin to move toward graphics processing units or GPUs. Not coincidentally, **this is exactly what China has done to achieve the world's fastest speeds with its "Tianhe-1A"** supercomputer. That computer combines about 7,000 Nvidia GPUs with 14,000 Intel CPUs: the only hybrid CPU-GPU system in the world of that scale.

"What we're really seeing in the efforts in China as well as the ones we have in the U.S. is that GPUs are what the future will look like," said Dunning in a phone interview Thursday. "What we're seeing is the beginning of something that's going to be happening all over the world."

NCSA already has a small CPU-GPU hybrid system. "It's something we have been working on for a number of years. We have a CPU-GPU cluster for the NCSA academic community. Made up of Intel CPUs and Nvidia GPUs. A 50 teraflop machine," he said. (Note that **Oak Ridge National Laboratories is also installing a hybrid system now.**)



Thom Dunning directs the Institute for Advanced Computing Applications and Technologies and the NCSA.

# Intel MIC Hardware

## INSPUR, NAOC - 2013.XI.26



**icpc ... "-mmic" ...  $61 \times 4 = 244$  x 1.1 GHz omp cores !!!  
Full fp64 !!!**

# Intel MIC Hardware

## Intel® Xeon Phi™ Coprocessor Family Reference Table

SKU #	Form Factor, Thermal	Peak Double Precision	Max # of Cores	Clock Speed (GHz)	GDDR5 Memory Speeds (GT/s)	Peak Memory BW	Memory Capacity (GB)	Total Cache (MB)	Board TDP (Watts)	Process
SE10P <small>(special edition)</small>	PCIe Card, Passively Cooled	1073 GF	61	1.1	5.5	352	8	30.5	300	22nm
SE10X <small>(special edition)</small>	PCIe Card, No Thermal Solution	1073 GF	61	1.1	5.5	352	8	30.5	300	
S110P	PCIe Card, Passively Cooled	1011 GF	60	1.053	5.0	320	8	30	225	
3100 Series	PCIe Card, Actively Cooled	>1 TF	Disclosed at 3100 series launch (H1'13)	5.0	240	6	28.5	300	22nm	
	PCIe Card, Passively Cooled	> 1 TF		5.0	240	6	28.5	300		



PCIe Card, Actively Cooled



PCIe Card, Passively Cooled

Current Generation:  
Knights Landing  
14nm

## Intel MIC hardware / Recent Processors



### Intel® Xeon Phi™ Processor 7290

- 36 MB L2 Cache
- 72 Cores
- 72 Threads
- 1.70 GHz Max Turbo Frequency



### Intel® Xeon Phi™ Processor 7290F

- 36 MB L2 Cache
- 72 Cores
- 72 Threads
- 1.70 GHz Max Turbo Frequency



Supercomputer from China: 96/33 Pflop/s Linpack  
Wuxi/Guangzhou/Tianjin National Supercomputing Center  
Taihu 10 mill. cores



Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon  
E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi  
31S1P



32000 Intel Xeon 12 core,  
48000 Intel Phi Accelerators 57 Core,  
now Chinese processor



Test of Taihu planned;  
But:  
Local cluster with new  
GPUs at NAOC gives  
much more resources.

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	<b>Summit</b> - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	2,414,592	148,600.0	200,794.9	10,096
2	DOE/NNSA/LLNL United States	<b>Sierra</b> - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM / NVIDIA / Mellanox	1,572,480	94,640.0	125,712.0	7,438
3	National Supercomputing Center in Wuxi China	<b>Sunway TaihuLight</b> - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
4	National Super Computer Center in Guangzhou China	<b>Tianhe-2A</b> - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 NUDT	4,981,760	61,444.5	100,678.7	18,482
5	Texas Advanced Computing Center/Univ. of Texas United States	<b>Frontera</b> - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR Dell EMC	448,448	23,516.4	38,745.9	

**GPU Volta**

**USA**

**GPU Volta**

**USA**

**Chinese Processor**



**Xeon Φ**



**Xeon Platinum**

**USA**

6	Swiss National Supercomputing Centre (CSCS) Switzerland	<b>Piz Daint</b> - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect, NVIDIA Tesla P100 Cray/HPE	387,872	21,230.0	27,154.3	2,384
---	--	---	---------	----------	----------	-------

**Swiss**

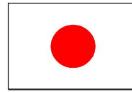
**GPU Pascal**

7	DOE/NNSA/LANL/SNL United States	<b>Trinity</b> - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect Cray/HPE	979,072	20,158.7	41,461.2	7,578
---	------------------------------------	---	---------	----------	----------	-------

**USA**

**XeonΦ**

8	National Institute of Advanced Industrial Science and Technology (AIST) Japan	<b>AI Bridging Cloud Infrastructure (ABCI)</b> - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR Fujitsu	391,680	19,880.0	32,576.6	1,649
---	--	--	---------	----------	----------	-------



**GPU Volta**

9	Leibniz Rechenzentrum Germany	<b>SuperMUC-NG</b> - ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path Lenovo	305,856	19,476.6	26,873.9	
---	----------------------------------	--	---------	----------	----------	--



**Xeon Platinum**

10	DOE/NNSA/LLNL United States	<b>Lassen</b> - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 IBM / NVIDIA / Mellanox	288,288	18,200.0	23,047.2	
----	--------------------------------	---	---------	----------	----------	--

**USA**

**GPU Volta**

# TOP500 List Refreshed, US Edged Out of Third Place



TOP500 Team | June 19, 2017 00:22 CEST

FRANKFURT, Germany; BERKELEY, Calif.; and KNOXVILLE, Tenn.— The 49th edition of the TOP500 list was released today in conjunction with the opening session of the ISC High Performance conference, which is taking place this week in Frankfurt, Germany. The list ranks the world's most powerful supercomputers based on the Linpack benchmark and is released twice per year.

[Read more](#)

## : System

Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45G

Sunway , NRCPC

National Supercomputing Center in Wuxi

China

Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-265

2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT

National Super Computer Center in Guangzhou

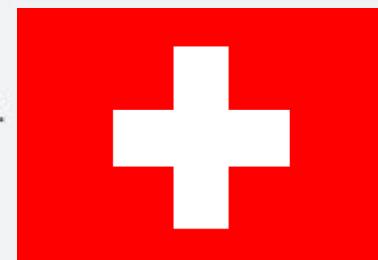
China

Piz Daint - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interci

NVIDIA Tesla P100 , Cray Inc.

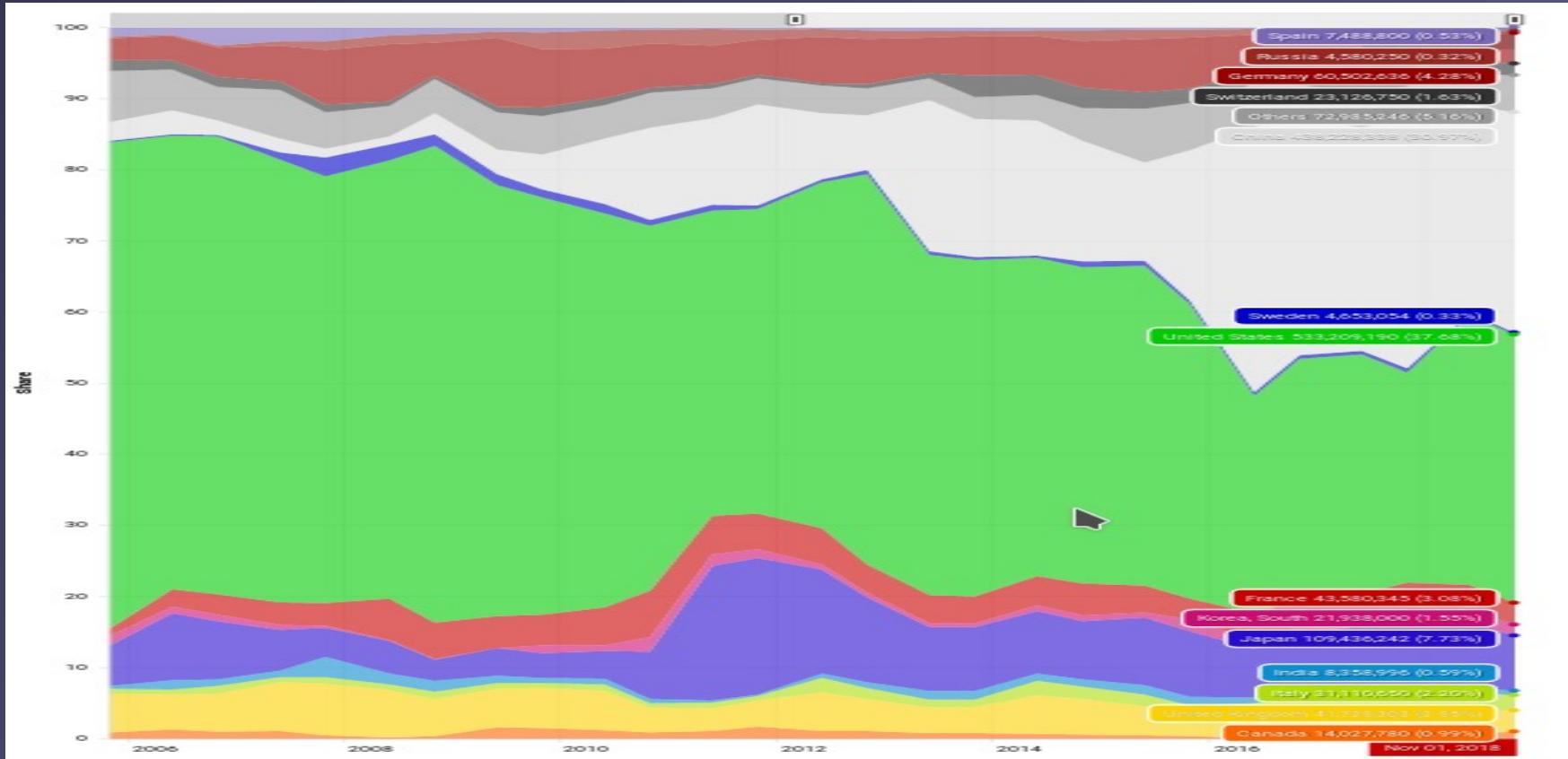
Swiss National Supercomputing Centre (CSCS)

Switzerland



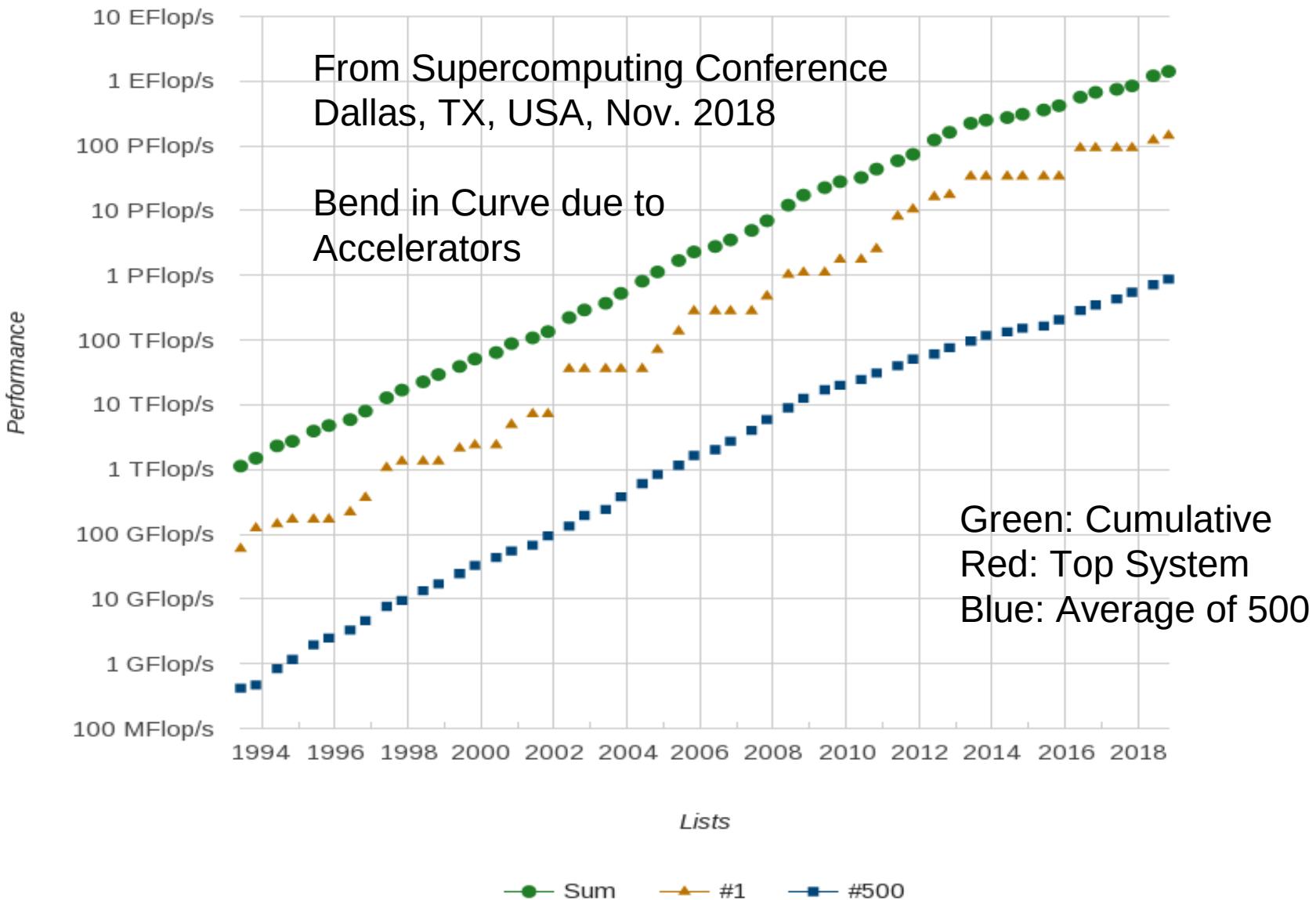
By  
Switzerland

# Top 500 List November 2018 – Performance Share of Countries



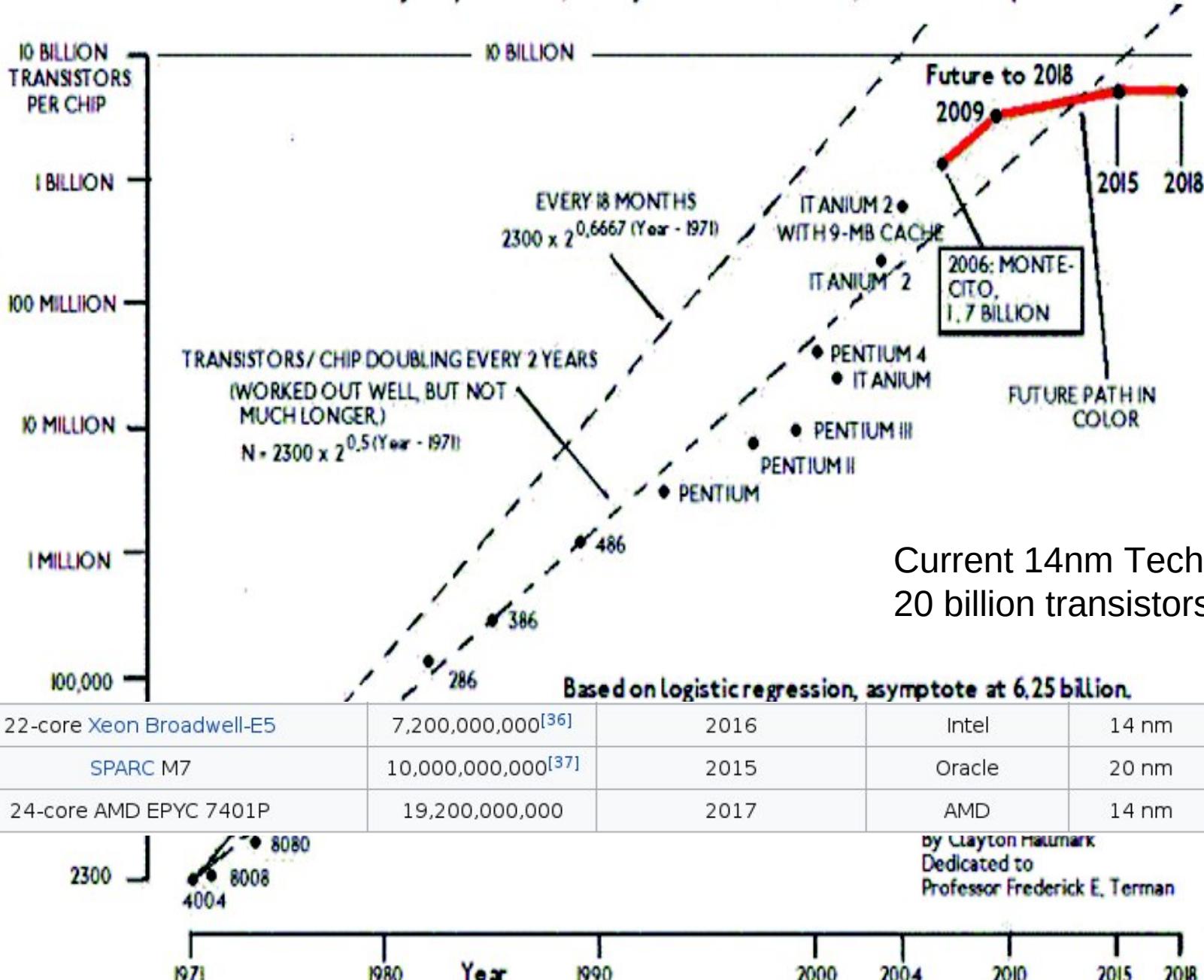
## Performance Development

## Moore's Law?



Moore's Law Ending (Red Line):  
Delayed products, Delayed 45nm / 32 nm, Reduced Capex

Number of transistors on



# GREEN 500 list 2018 Power Efficiency (Gflops/Watts), see also <http://www.top500.org/green500> - 2019 similar.

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	375	<b>Shoubu system B</b> - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	953,280	1,063.3	60	17.604
2	374	<b>DGX SaturnV Volta</b> - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100 , Nvidia NVIDIA Corporation United States	22,440	1,070.0	97	15.113
3	1	<b>Summit</b> - 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	2,397,824	143,500.0	9,783	14.668
4	7	<b>AI Bridging Cloud Infrastructure (ABCi)</b> - 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	1,649	14.423
5	22	<b>TSUBAME3.0</b> - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 , HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	792	13.704
6	2	<b>Sierra</b> - IBM Power System S922LC, 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	7,438	12.723
7	446	<b>AIST AI Cloud</b> - NEC 4U-8GPU Server, Xeon E5-2630Lv4 10C 1.8GHz, Infiniband EDR, NVIDIA Tesla P100 SXM2 , NEC National Institute of Advanced Industrial Science and Technology Japan	23,400	961.0	76	12.681
8	411	<b>MareNostrum P9 CTE</b> - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM Barcelona Supercomputing Center Spain	19,440	1,018.0	86	11.865
9	38	<b>Advanced Computing System(PreE)</b> - Sugon TC8600, Hygon Dhanya 32C 2GHz, Deep Computing Processor, 200Gb 6D-Torus , Sugon Sugon China	163,840	4,325.0	380	11.382
10	20	<b>Taiwania 2</b> - QCT QuantaGrid D52G-4U/LC, Xeon Gold 6154 18C 3GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100 SXM2 , Quanta Computer / Taiwan Fixed Network / ASUS Cloud National Center for High Performance Computing Taiwan	170,352	9,000.0	798	11.285