

中国科学院国家天文台

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@nao.cas.cn](mailto:spurzem@nao.cas.cn)  
<http://silkroad.bao.ac.cn>



北京大學  
PEKING UNIVERSITY

# Introduction to GPU Accelerated Computing

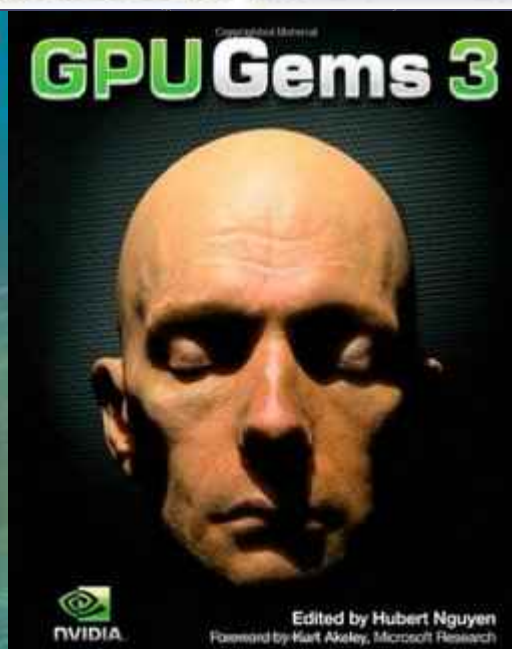
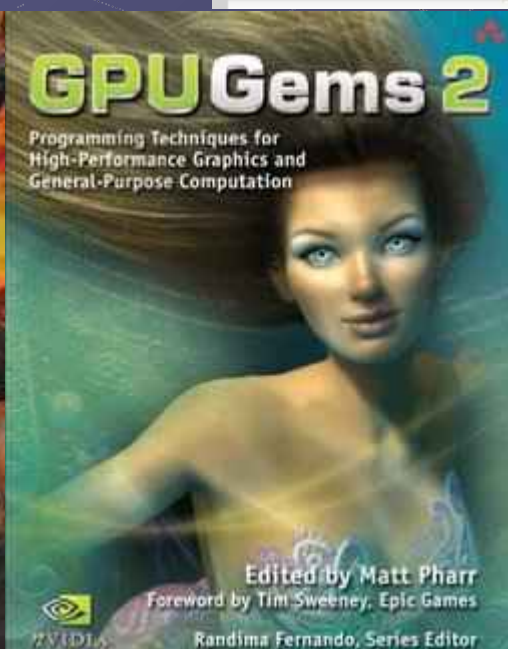
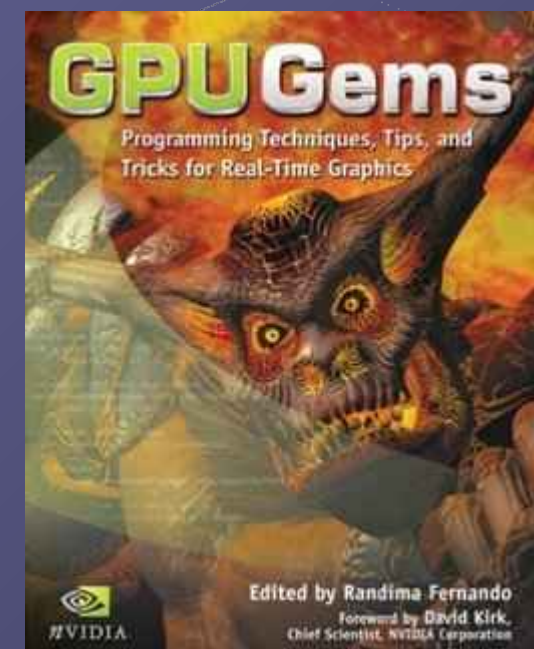
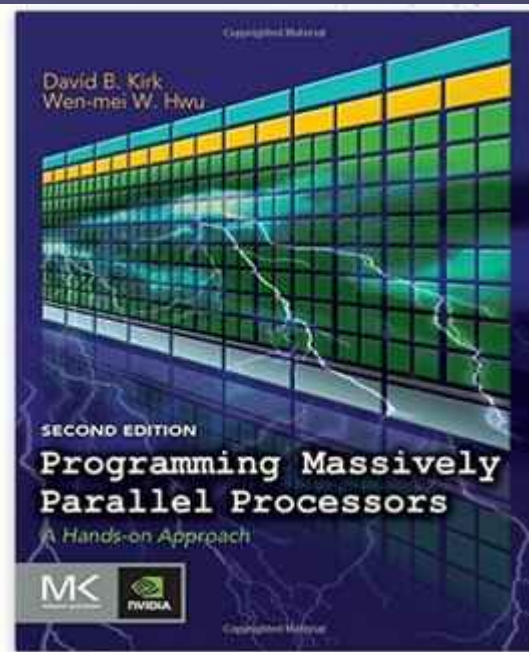
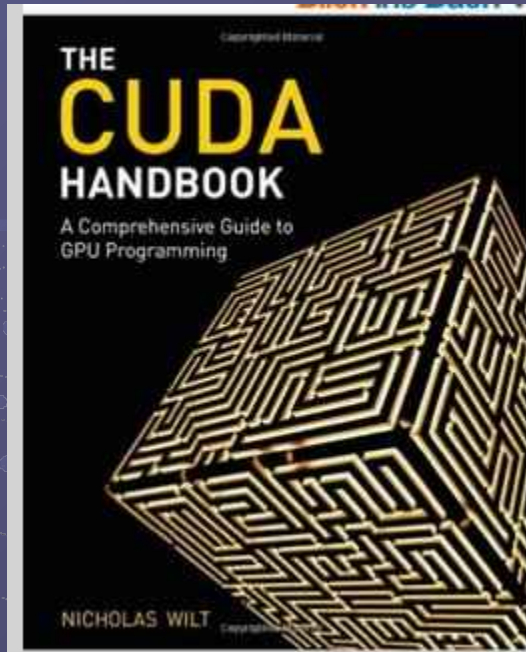
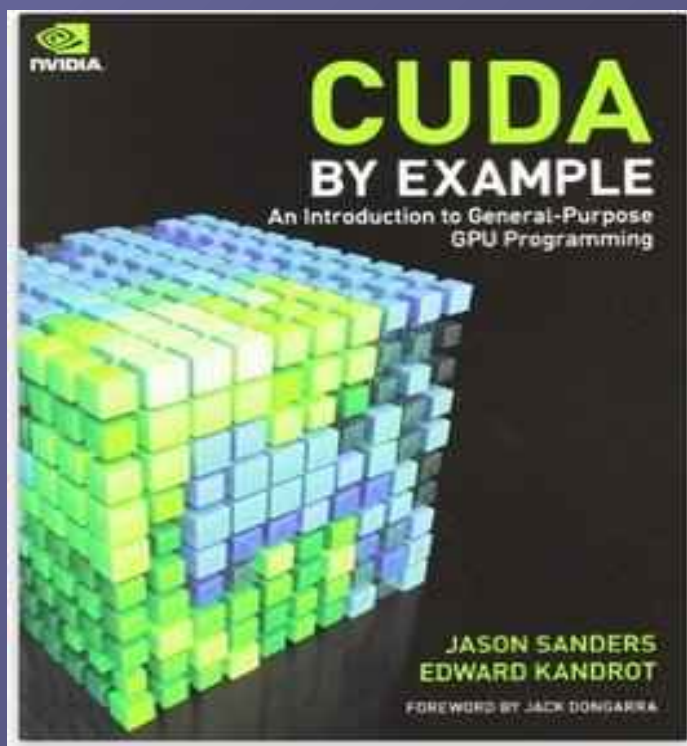
## July 25-28, 2016

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

1. Monday morning: General Introduction Computer Architecture, Many-Core, GPU and others...
2. Monday afternoon: 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 Product, Histograms, Events
5. Wednesday morning: New Features of Kepler Architecture, Astrophysical N-Body Code
6. Wednesday Afternoon: Astrophysical Parallel N-Body Code Using MPI and GPU
7. Thursday Morning: Parallelization and Amdahl's Law
8. Thursday Afternoon: Amdahl's Law and GPU Acceleration



# Literature







Observations (Experiment)



Theory



Computational Physics



# 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?**

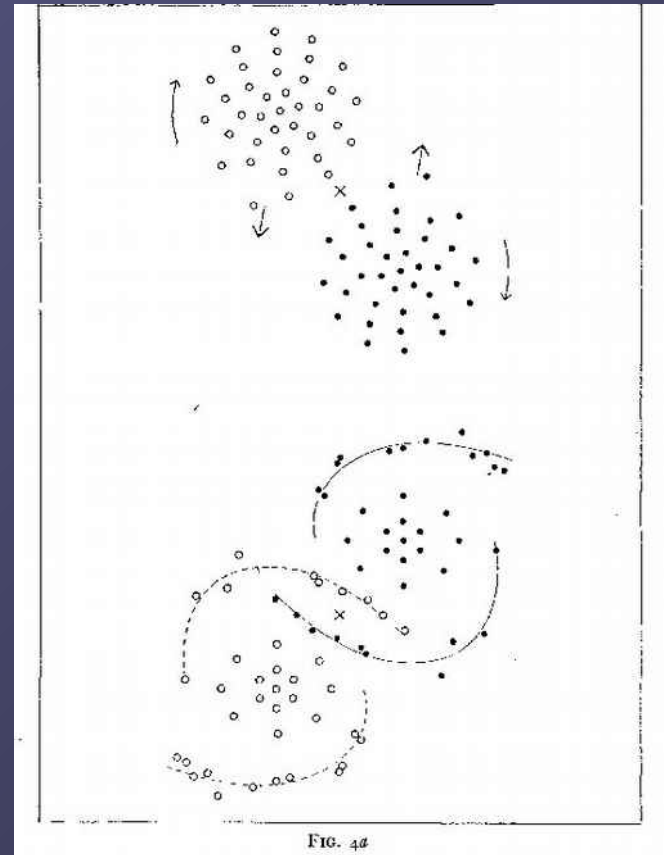
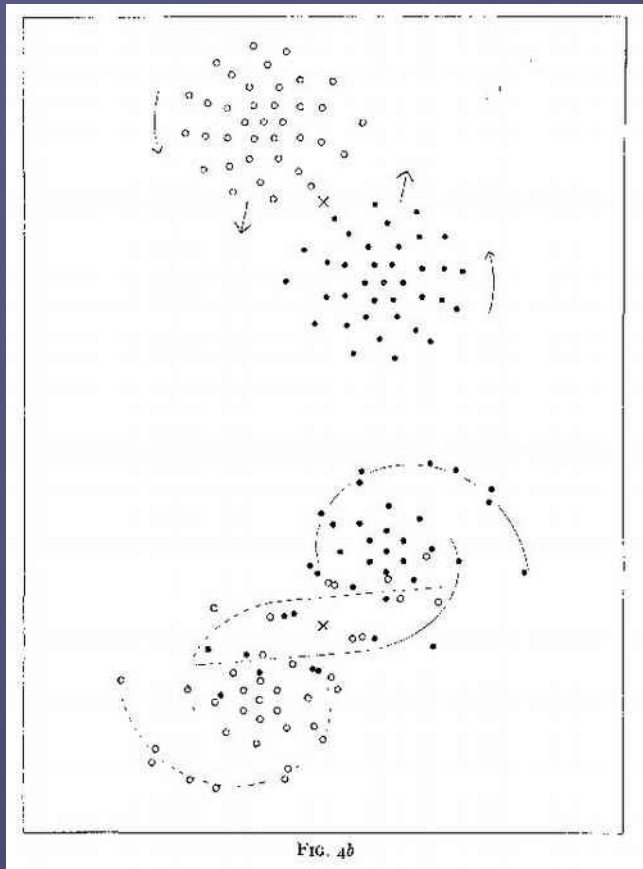
**» ...with 200 light bulbs**

# Geschichte



The Astrophysical Journal, Nov. 194

**LUMA METALL**



# HARDWARE

...before von Neumann...

● Konrad Zuse (1910-1995) Berlin



Invented freely programmable Computer



Z1 in parental flat 1936

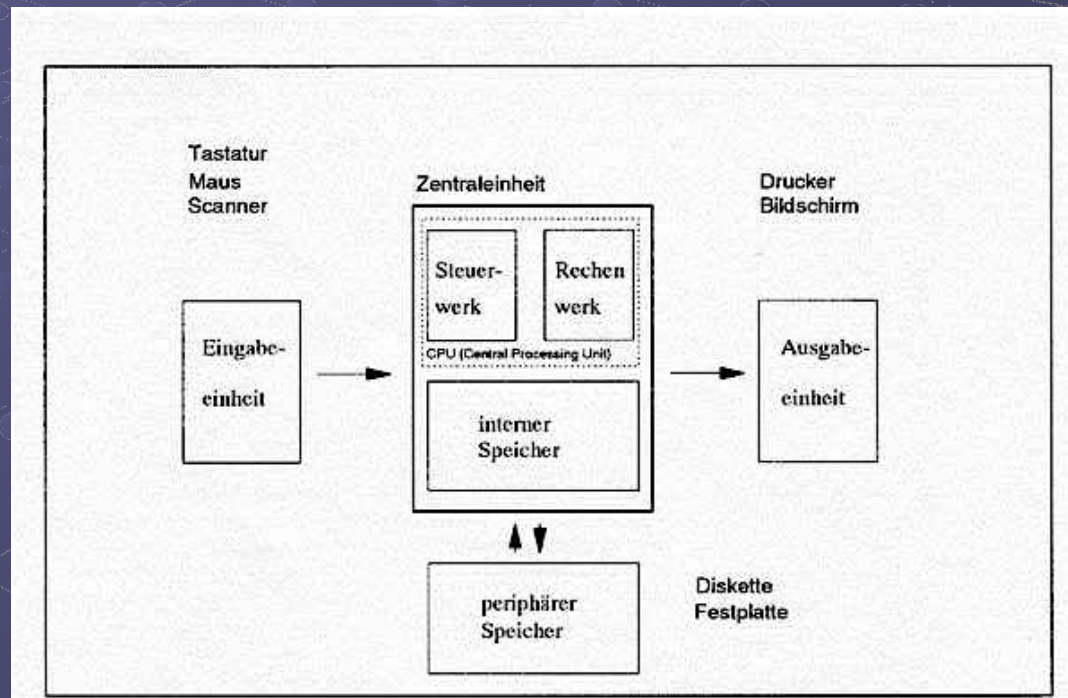


# HARDWARE

- John von Neumann (1903-1957)

Born Budapest, Lecturer Berlin, since 1930 Princeton Univ.

Requirements for the Construction of an electronic computing device(1946)

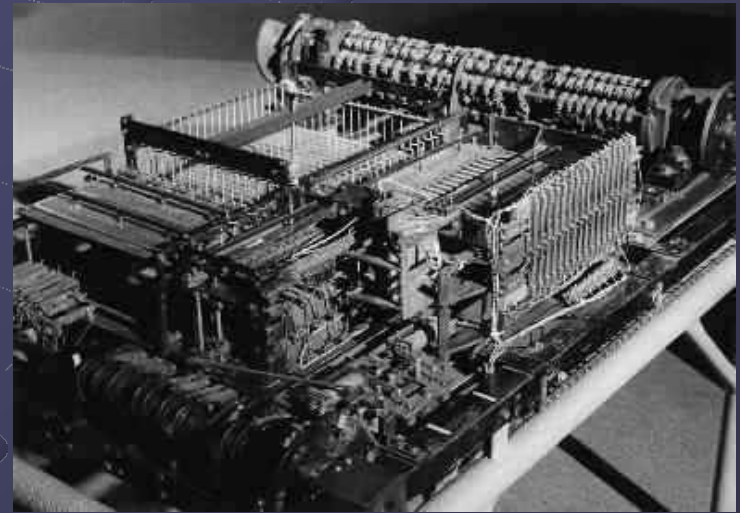




# Computer



**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

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)



# Computer

“The father of supercomputing”



## ● Seymour Cray (1925-1996)



**CRAY1: Vectorregisters (1976)**  
**160 Mflop, 80 MHz, 8 MByte RAM**

**CRAY2: (1984)**  
**1Gflop, 120MHz, 2GByte RAM**

# Computer

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



*Opening Ceremony June 2008*



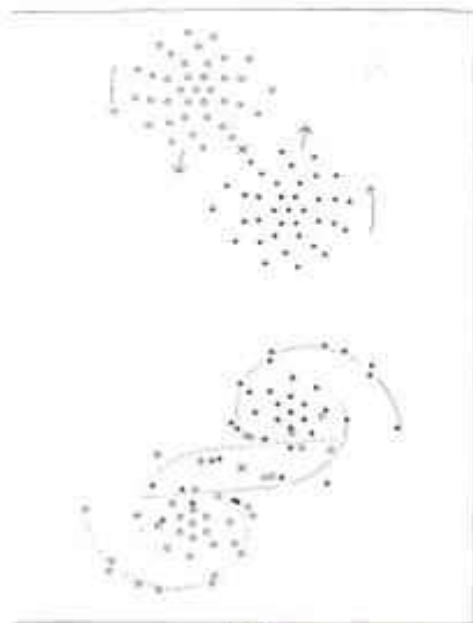


Fig. 4

Holmberg, 1937/1941

Credit: Whitmore (STScI) and NASA



NGC 4038/NGC 4039



FIG. 23.—Symmetric model of NGC 4038/9. Here two identical disks of radius  $0.75R_{\text{min}}$  suffered an  $e \approx 0.5$  encounter with orbit angles  $i_1 = i_2 = 60^\circ$  and  $\omega_1 = \omega_2 = -30^\circ$  that appeared the same to both. The above all-inclusive views of the debris and remnants of these disks have been drawn exactly normal and edge-on to the orbit plane; the latter viewing direction is itself  $30^\circ$  from the line connecting the two pericenters. The viewing time is  $t = 15$ , or slightly past apocenter. The filled and open symbols again disclose the original loyalties of the various test particles.

Toomre & Toomre, 1972, ApJ, 178, 623

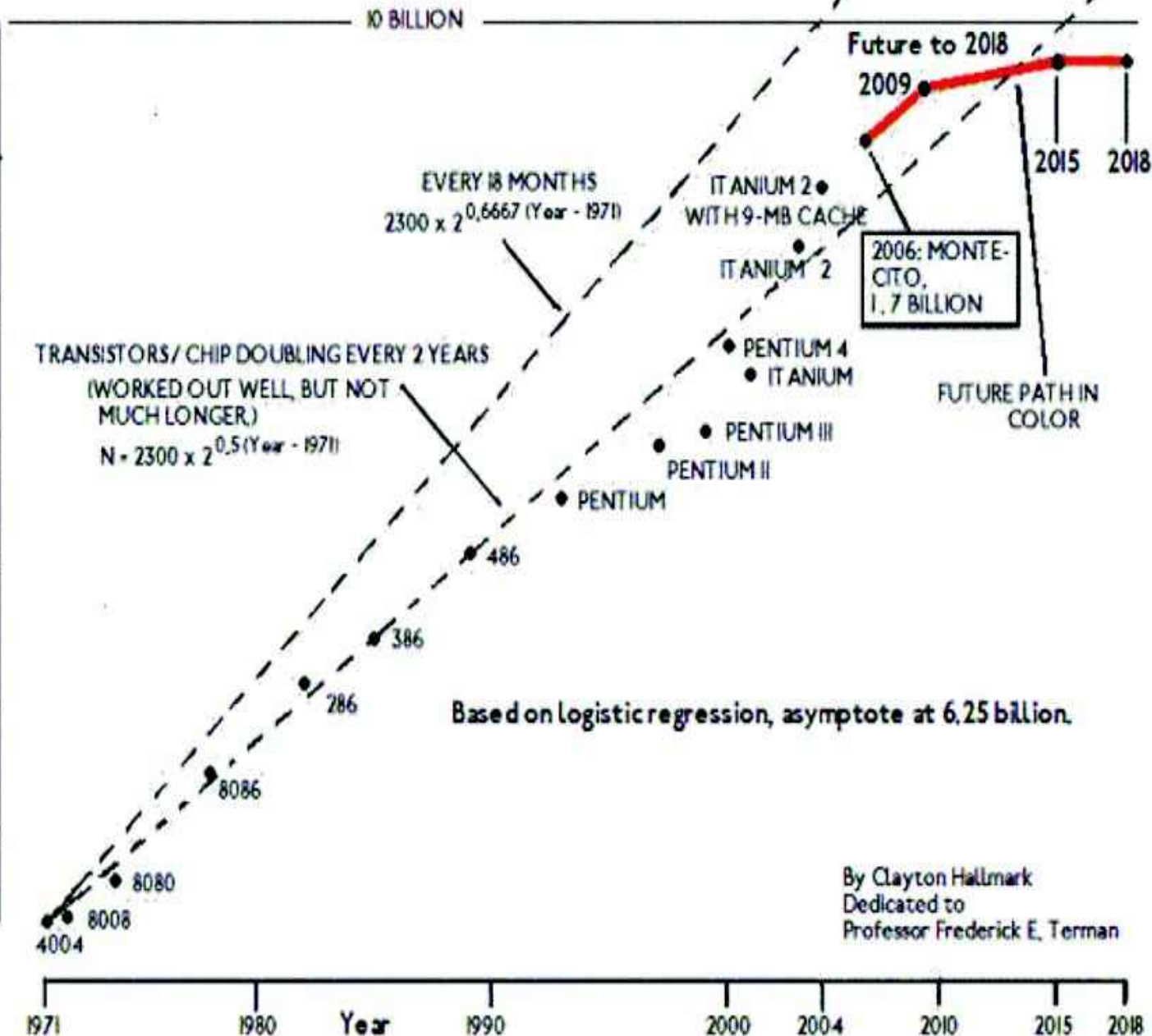




Moore's Law Ending (Red Line):  
 Delayed products, Delayed 45nm / 32 nm, Reduced Capex

Number of transistors on

10 BILLION TRANSISTORS PER CHIP  
 1 BILLION  
 100 MILLION  
 10 MILLION  
 1 MILLION  
 100,000  
 10,000  
 2300



1971 1980 Year 1990 2000 2004 2010 2015 2018

# Computational Science..

...after von Neumann...

Exaflop/s?

Petaflop/s

Teraflop/s

Gigaflop/s

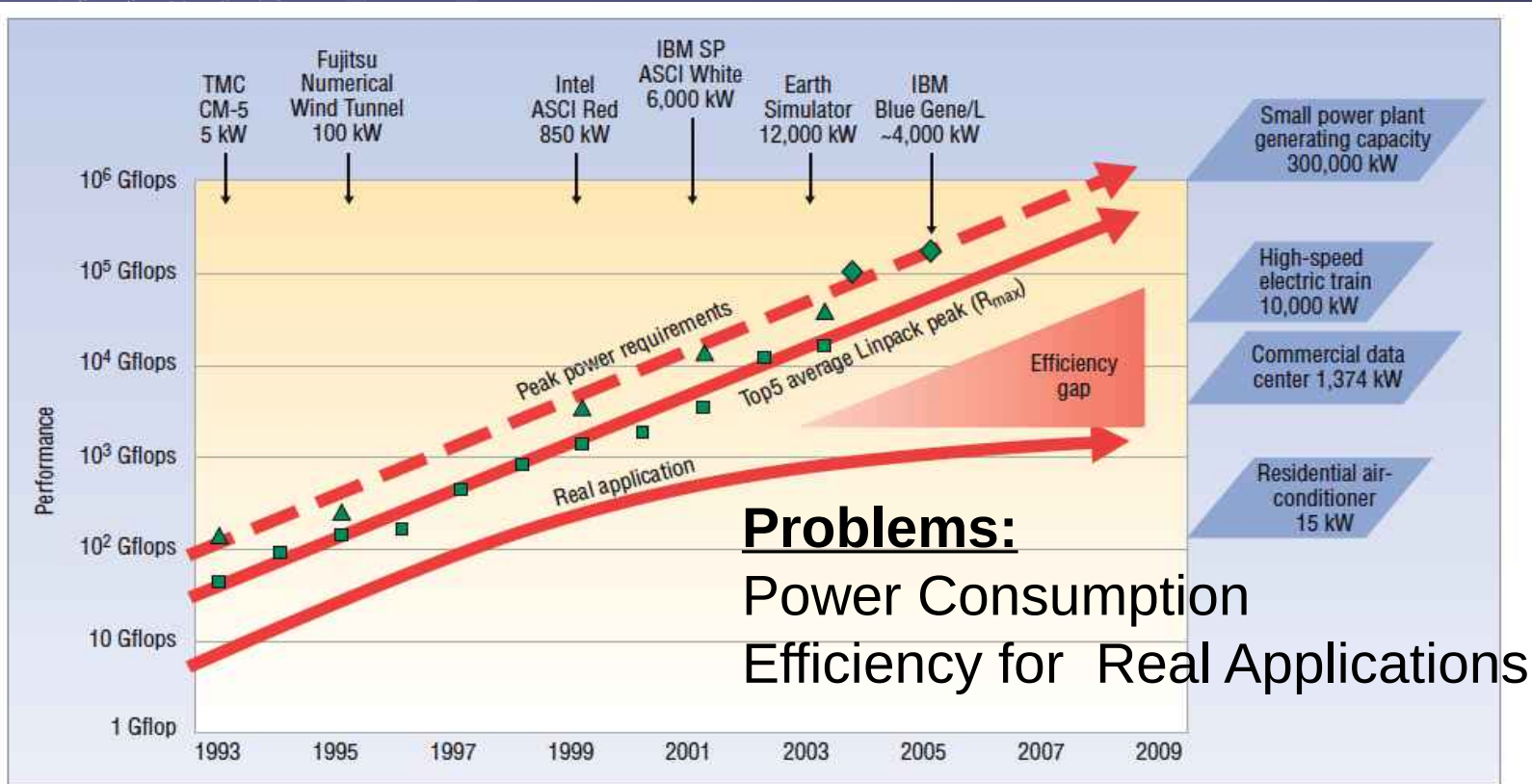


Figure 1. Rising power requirements. Peak power consumption of the top supercomputers has steadily increased over the past 15 years.

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

# 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

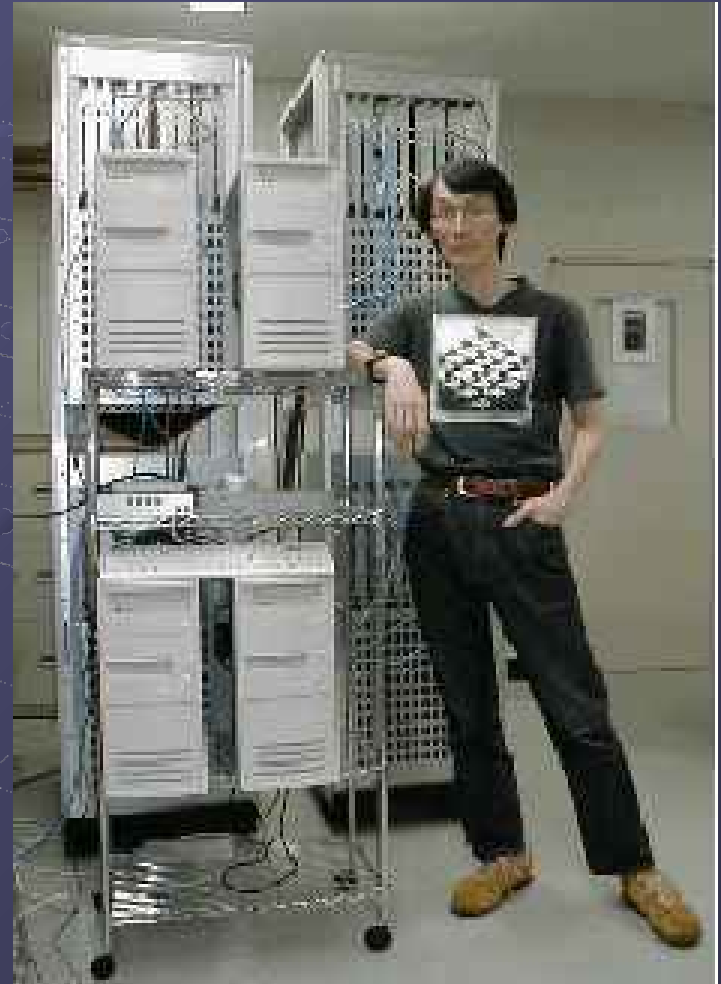




# HARDWARE

## 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](#)

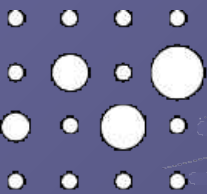
[Sales information](#)

## 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 G2050, Infiniband	36.00

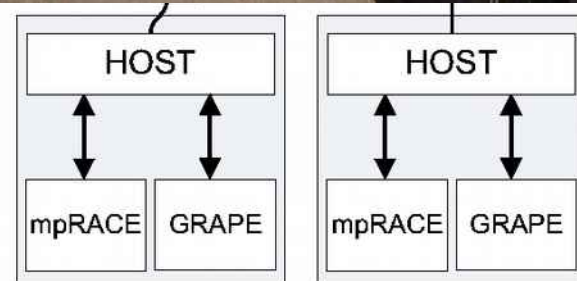


Volkswagen **Stiftung**

## GRACE Cluster

4 Tflops (32 micro-GRAPE6)  
Dual Port Infiniband  
4 MPRACE-1 reconfigurable  
(soon: 32 MPRACE-2)

GRAPE + MPRACE  
= GRACE





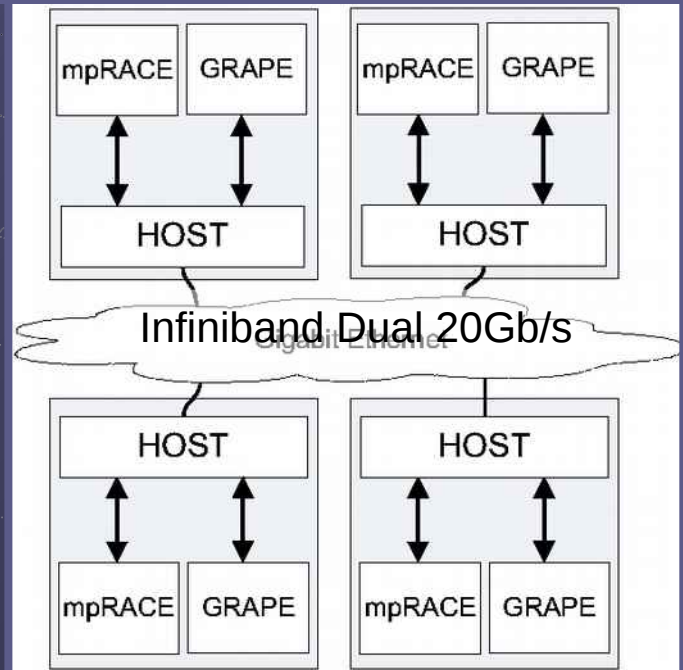
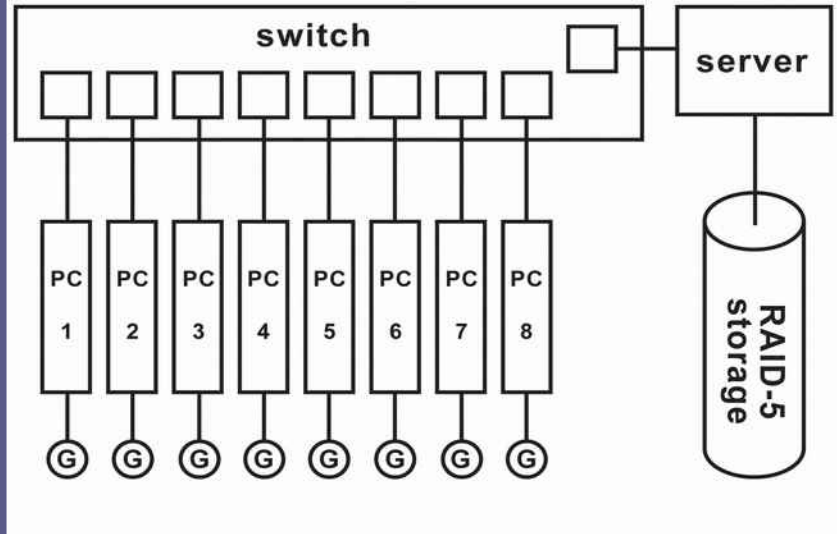


# FPGA...

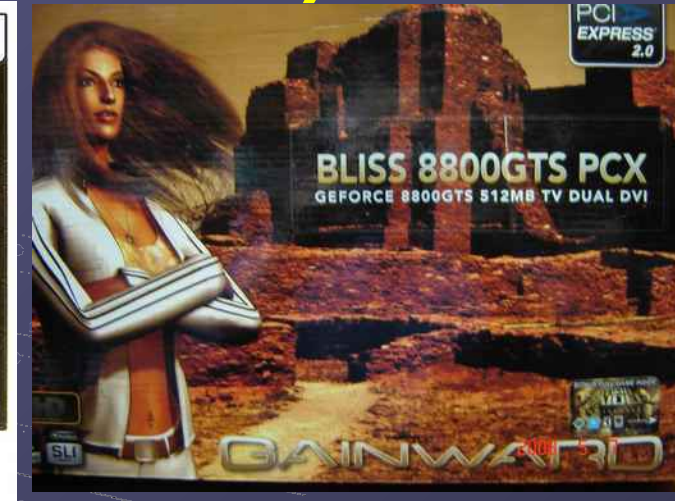
## Heidelberg titan 32 node cluster

**GRACE = GRAPE + MPRACE**

- 32 dual-Xeon 3.2 GHz nodes
- 32 GRAPE6a
- 32 FPGA
- 7 TB RAID
- Dual port Infiniband link (20 Gb/s)
- Speed: ~4 Tflops
- N up to 4M
- Cost: ~380K EUR
- Funding: Volkswagen/Baden-Württemberg



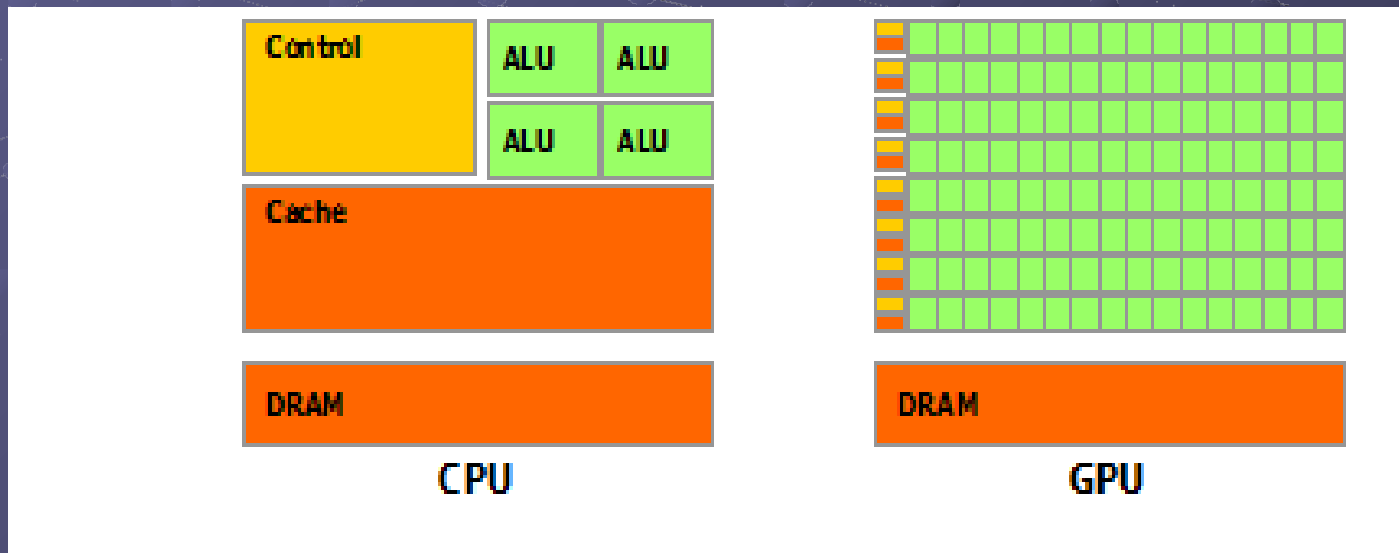
# Graphics Processors (GPU) as General Purpose Supercomputers (GPGPU)



2008...  
GeForce 9800 GTX, 128 Stream Proc., 512 MB  
GeForce 9800 GX2, 256 Stream Proc., 1 GB  
GeForce 9800 GT, 64 Stream Proc., 512 MB  
[...]  
2009: Tesla ~200 Proc., 4GB  
2010: Fermi ~400 Proc., 4GB  
2013: Kepler K20, ~2500 Procs., 6GB



# CPU and GPU; from CUDA NVIDIA Developer Zone at <http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>



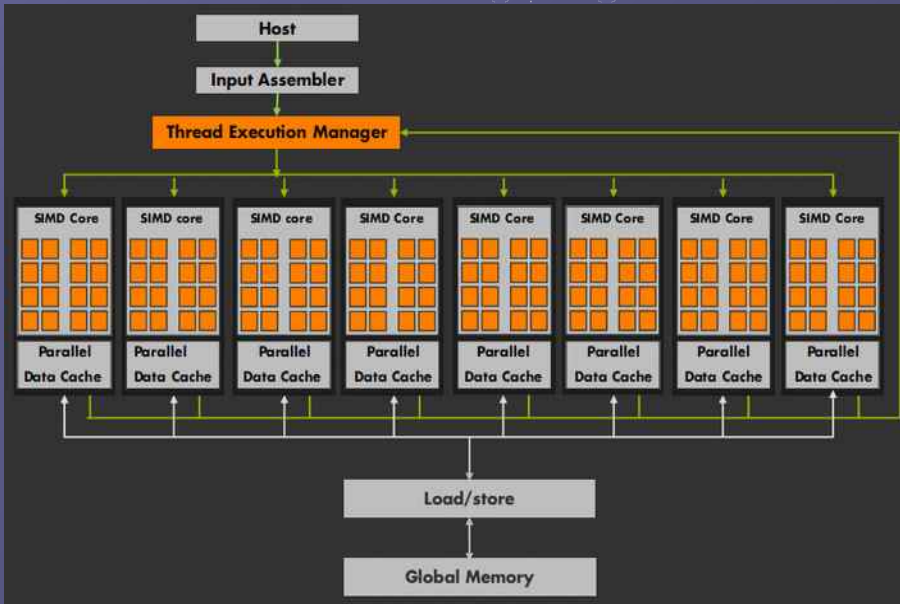
**“The GPU devotes more transistors to computing”  
“favours data parallel operations”**



# CPU vs. GPU speedup timeline

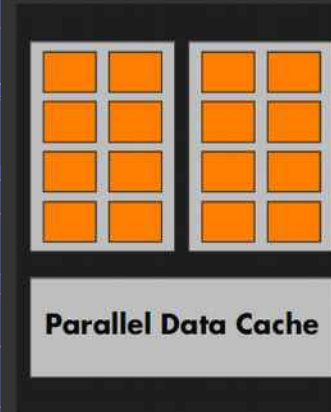


# Hardware around 2006



## Each core

- 8 functional units
- SIMD 16/32 "warp"
- 8-10 stage pipeline
- Thread scheduler
- 128-512 threads/core
- 16 KB shared memory



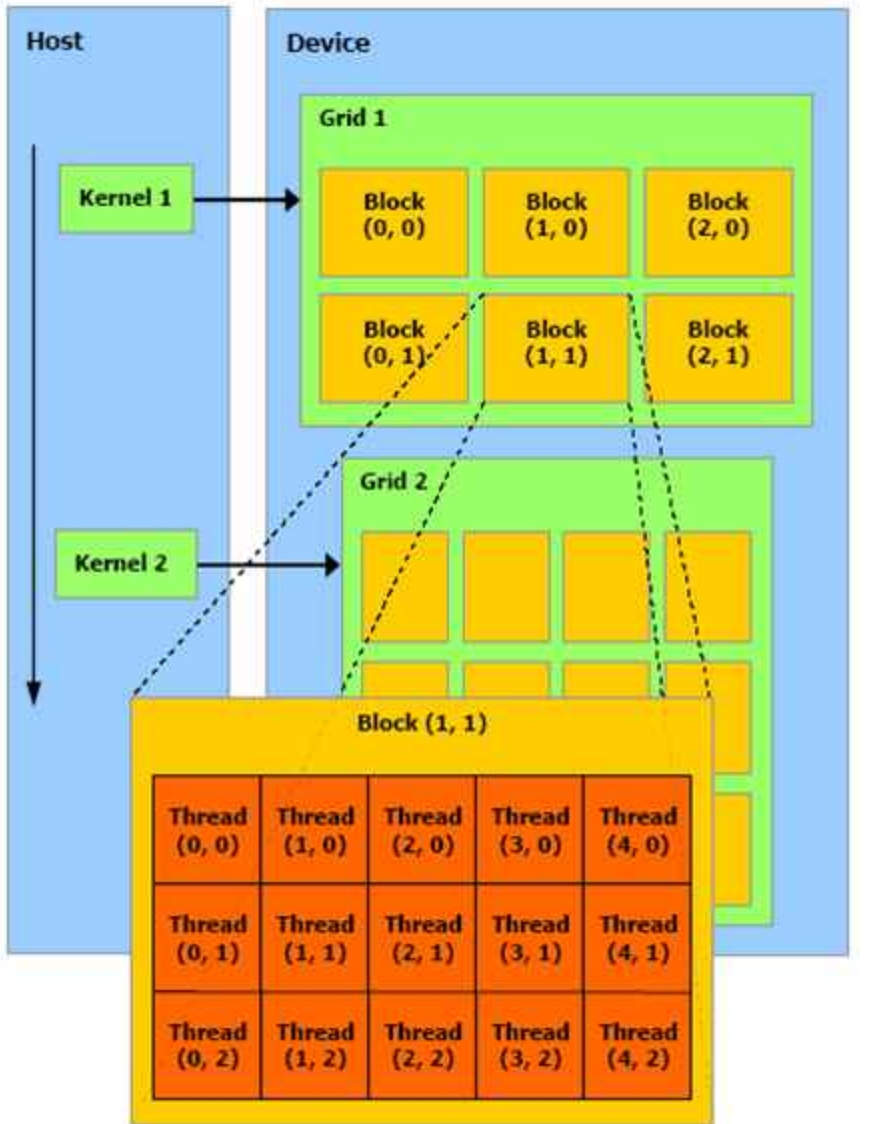
**Total #threads/chip**

$$16 * 512 = 8K$$

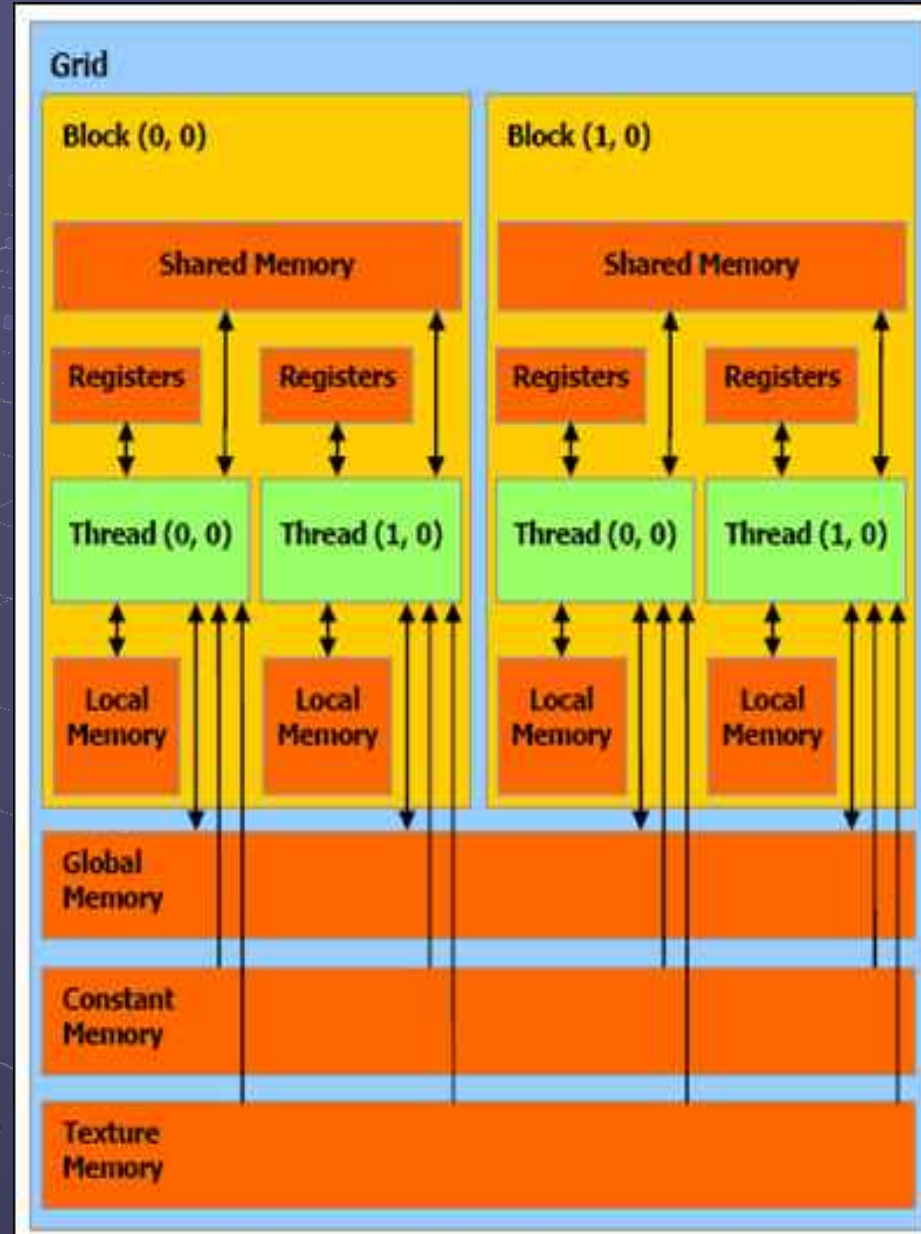
## GeForce 8800 GTX:

**575 MHz \* 128 processors \* 2 flop/inst \* 2 inst/clock = 333 Gflops**

# GPU Structure From: [http://geco.mines.edu/tesla/cuda\\_tutorial\\_mio/](http://geco.mines.edu/tesla/cuda_tutorial_mio/)



The host issues a succession of kernel invocations to the device. Each kernel is executed as a batch of threads organized as a grid of thread blocks

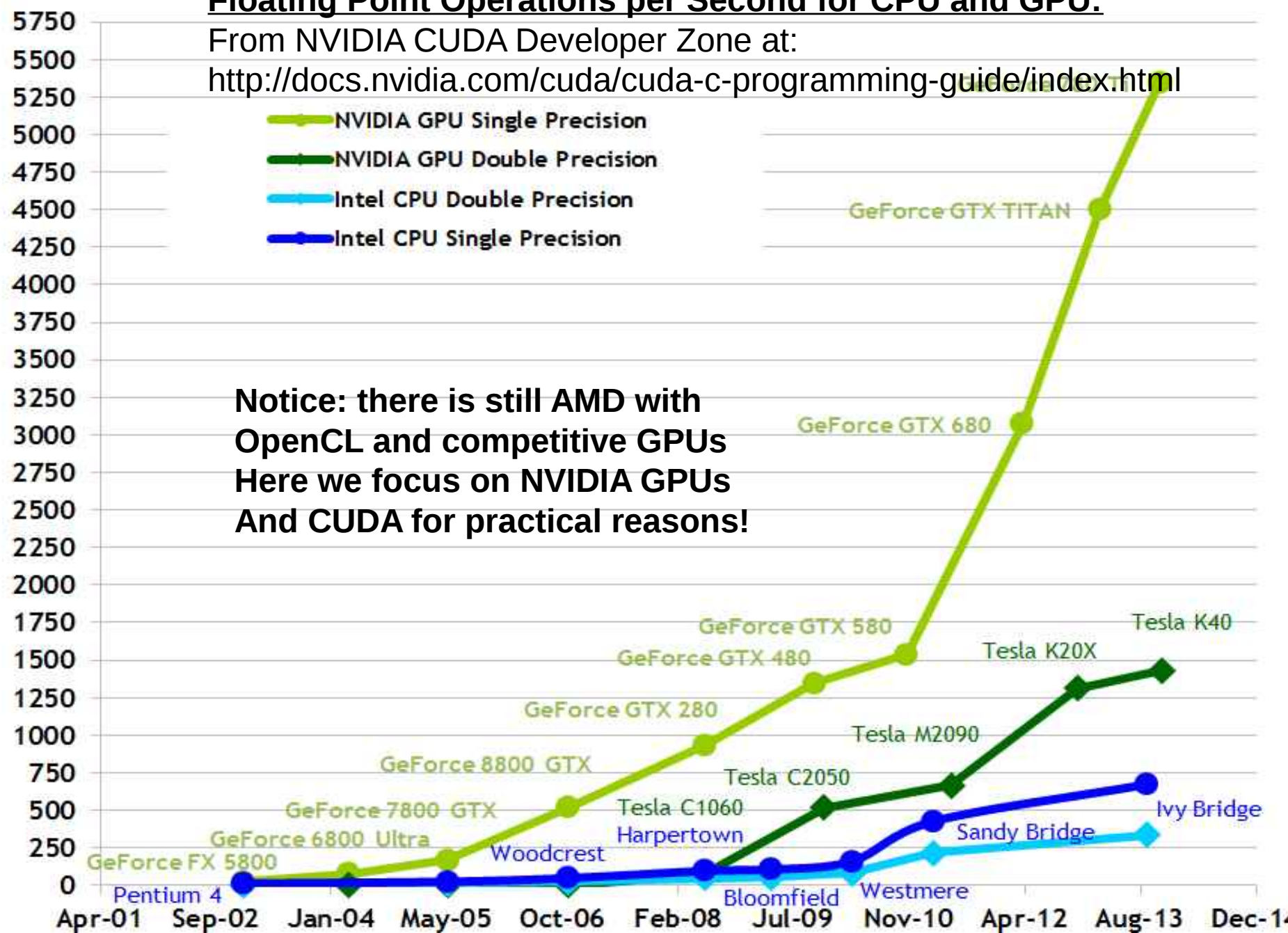




## Floating Point Operations per Second for CPU and GPU:

From NVIDIA CUDA Developer Zone at:

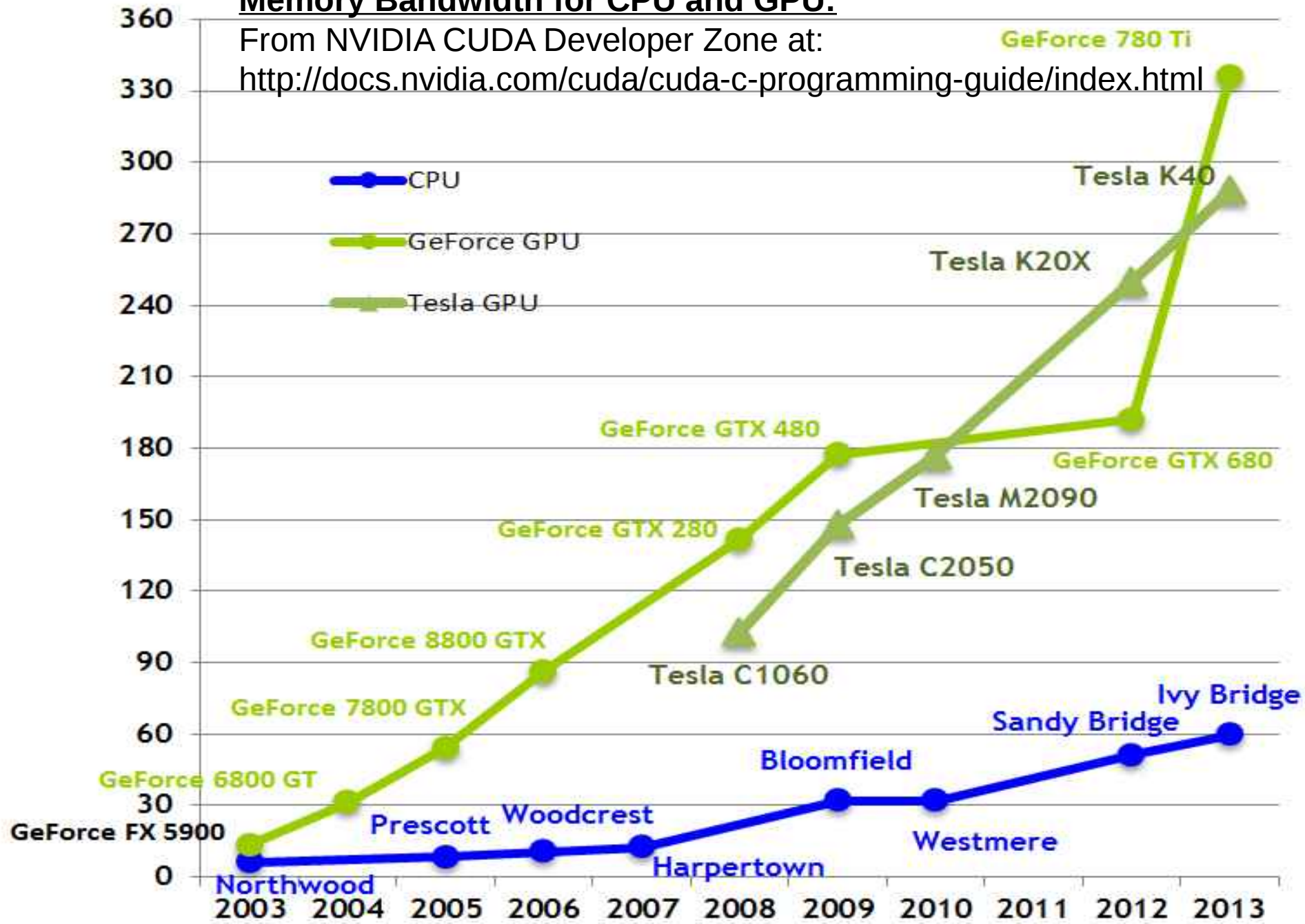
<http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>



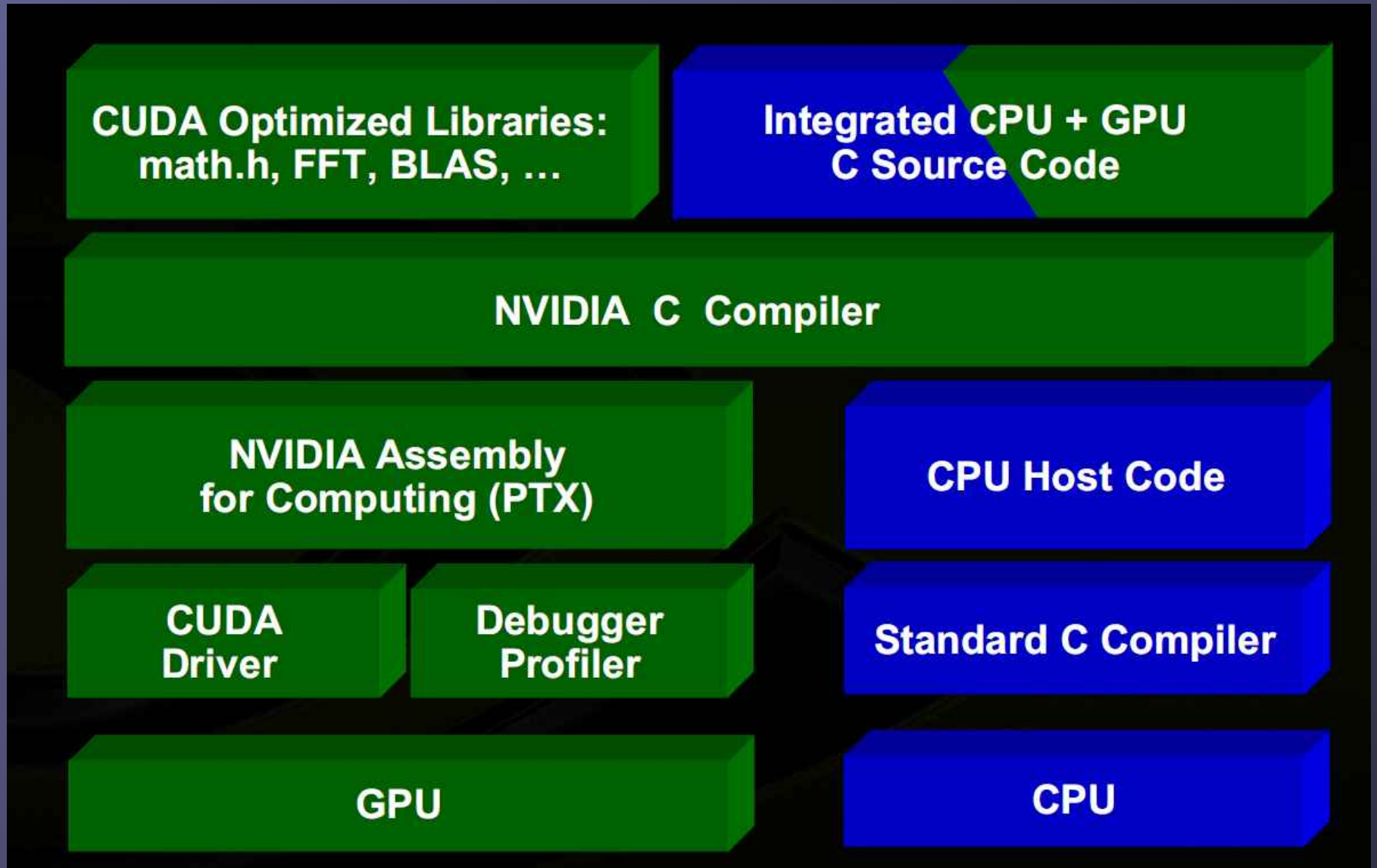
# Theoretical GB/s

## Memory Bandwidth for CPU and GPU:

From NVIDIA CUDA Developer Zone at:  
<http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>



# CUDA





# Simple CUDA example

## CPU C program

```
void addMatrix(float *a, float *b,
              float *c, int N)
{
    int i, j, index;
    for (i = 0; i < N; i++) {
        for (j = 0; j < N; j++) {
            index = i + j * N;
            c[index]=a[index] + b[index];
        }
    }
}

void main()
{
    .....
    addMatrix(a, b, c, N);
}
```

## CUDA C program

```
__global__ void addMatrix(float *a, float *b,
                          float *c, int N)
{
    int i=blockIdx.x*blockDim.x+threadIdx.x;
    int j=blockIdx.y*blockDim.y+threadIdx.y;
    int index = i + j * N;
    if ( i < N && j < N)
        c[index]= a[index] + b[index];
}

void main()
{
    ..... // allocate & transfer data to GPU
    dim3 dimBlk (blocksize, blocksize);
    dim3 dimGrd (N/dimBlk.x, N/dimBlk.y);
    addMatrix<<<dimGrd,dimBlk>>>(a, b, c,N);
}
```

# GPU Computing Applications

## Libraries and Middleware

CUFFT CUBLAS CURAND CUSPARSE	CULA MAGMA	Thrust NPP	VSIPL SVM OpenCurrent	PhysX OptiX	iray	MATLAB Mathematica
---------------------------------------	---------------	---------------	-----------------------------	----------------	------	-----------------------

## Programming Languages

C	C++	Fortran	Java Python Wrappers	DirectCompute	Directives (e.g. OpenACC)
---	-----	---------	----------------------------	---------------	------------------------------



## CUDA-Enabled NVIDIA GPUs

<b>Kepler Architecture</b> (compute capabilities 3.x)	GeForce 600 Series	<u>Quadro Kepler Series</u>	Tesla K20 Tesla K10
<b>Fermi Architecture</b> (compute capabilities 2.x)	GeForce 500 Series GeForce 400 Series	<u>Quadro Fermi Series</u>	Tesla 20 Series
<b>Tesla Architecture</b> (compute capabilities 1.x)	<u>GeForce 200 Series</u> <u>GeForce 9 Series</u> <u>GeForce 8 Series</u>	<u>Quadro FX Series</u> <u>Quadro Plex Series</u> <u>Quadro NVS Series</u>	Tesla 10 Series



# NAOC laohu cluster Beijing, China





# Milky Way GPU cluster.

SFB 881 – The Milky Way System

Collaboration with FZ Jülich, Germany

206 nodes x 24 = 4944 CPU cores (@ 2.8 GHz)

206 x 96 GB ~ 20 TB RAM CPU memory

408 GPUs M2070/M2050 ~ 200k GPU threads

~ 2 TB GPU device memory

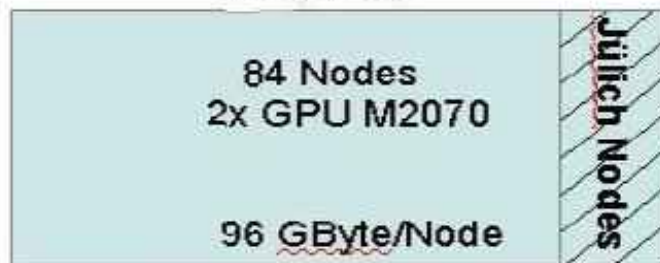
since mid. 2012 jointly operated.

nodes "judge123 - judge206" – MW part.

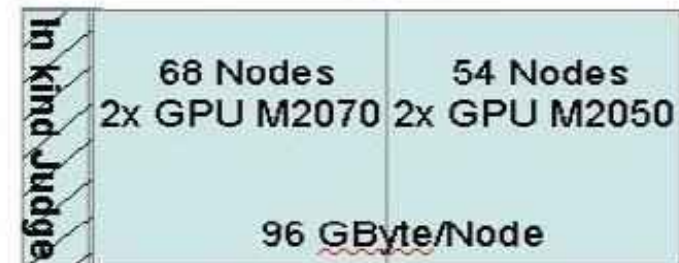


## GPU Cluster

Milky Way

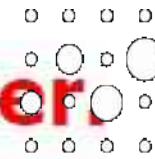


Judge



Heidelberg

# Kepler GPU cluster



VolkswagenStiftung

## 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.





# Top 10 List November 2010





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).

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



## “天河一号” 超级计算机系统 TH-1 supercomputer



Landmark result of the important project "High Efficient Supercomputer and Grid Service Environment" supported by National 863 Program.

- ▶ Built by National University of Defense Technology, with the cooperation of National Supercomputer Center in Tianjin (NSCC-TJ) and Inspur (Beijing) Electronic Information Industry Co., Ltd.

Host system of NSCC-TJ, installed in Tianjin Binhai New Area.

A backbone node of the national grid of China.

# NCSA director: GPU is future of supercomputing

by Brooke Crothers



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.**)

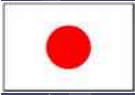


Thom Dunning directs the Institute for Advanced Computing Applications and Technologies and the NCSA.

# Top 10 List 2011 -----

# 2012

1	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom
2	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect
3	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom
4	SuperMUC - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR
5	Tianhe-1A - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050
6	Jaguar - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090
7	Fermi - BlueGene/Q, Power BQC 16C 1.60GHz, Custom
8	JuQUEEN - BlueGene/Q, Power BQC 16C 1.60GHz, Custom
9	Curie thin nodes - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR
10	Nebulae - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050



Rank	Site	System
1	National University of Defense Technology China	<b>Tianhe-2 (MilkyWay-2)</b> - TH-IVB-FEP Cluster Intel Xeon E5-2692 12C 2.200GHz, TH Express Intel Xeon Phi 31S1P NUDT
2	DOE/SC/Oak Ridge National Laboratory United States	<b>Titan</b> - Cray XK7, Opteron 6274 16C 2.200GHz Cray Gemini interconnect, NVIDIA K20x Cray Inc.
3	DOE/NNSA/LLNL United States	<b>Sequoia</b> - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu
5	DOE/SC/Argonne National Laboratory United States	<b>Mira</b> - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM
6	Texas Advanced Computing Center/Univ. of Texas United States	<b>Stampede</b> - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell
7	Forschungszentrum Juelich (FZJ) Germany	<b>JUQUEEN</b> - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM
8	DOE/NNSA/LLNL United States	<b>Vulcan</b> - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM
9	Leibniz Rechenzentrum Germany	<b>SuperMUC</b> - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR

Xeon $\phi$

GPU

Xeon $\phi$



Nr. 1,2 Supercomputer from China: 96/33 Pflop/s Linpack  
Wuxi/Guangzhou/Tianjin National Supercomputing Center  
Taihu 10 mill. cores






Tianhe-2 (MilkyWay-2) - TH-IV  
E5-2692 12C 2.200GHz, TH Ex  
31S1P



32000 Intel Xeon 12 core,  
48000 Intel Phi Accelerators 57 Core

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

# Top 10 List June 2016

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	 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
<b><u>Chinese Processor</u></b>						
2	 National Super Computer Center in Guangzhou China	<b>Tianhe-2 (MilkyWay-2)</b> - 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
<b><u>Xeonφ</u></b>						
3	<b>USA</b> DOE/SC/Oak Ridge National Laboratory United States	<b>Titan</b> - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x, Cray Inc.	560,640	17,590.0	27,112.5	8,209
<b><u>GPU</u></b>						
4	<b>USA</b> DOE/NNSA/LLNL United States	<b>Sequoia</b> - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
5	 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	 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
6	<b>USA</b> DOE/SC/Argonne National Laboratory United States	<b>Mira</b> - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945
7	<b>USA</b> DOE/NNSA/LANL/SNL United States	<b>Trinity</b> - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	301,056	8,100.9	11,078.9	
8	<b>Swiss</b> Swiss National Supercomputing Centre (CSCS) Switzerland	<b>Piz Daint</b> - Cray XC30, Xeon E5-2670 8C 2.600GHz, Aries Interconnect , NVIDIA K20x Cray Inc.	115,984	6,271.0	7,788.9	2,325
9	 HLRS - Hochleistungsrechenzentrum Stuttgart Germany	<b>Hazel Hen</b> - Cray XC40, Xeon E5-2680v3 12C 2.5GHz, Aries interconnect Cray Inc.	185,088	5,640.2	7,403.5	
10	<b>Saudi-A.</b> King Abdullah University of Science and Technology Saudi Arabia	<b>Shaheen II</b> - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	196,608	5,537.0	7,235.2	2,834

<b>Green500 Rank</b>	<b>MFLOPS/W</b>	<b>Site</b>	<b>System</b>	<b>Total Power(kW)</b>
1	7031.4	Institute of Physical and Chemical Research (RIKEN)	ExaScaler-1.4 80Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SC	50.3
2	5331.5	GSIC Center, Tokyo Institute of Technology	LX 1U-4GPU/104Re-1G Cluster, Intel Xeon E5-2620v2 6C 2.1GHz, Infiniband FDR, NVIDIA Tesla K80	51.1
3	5272.1	GSI Helmholtz Center	ASUS ESC4000 FDR/G2S, Intel Xeon E5-2690v2 10C 3GHz, Infiniband FDR, AMD FirePro S9150	57.2
4	4778.5	Institute of Modern Physics (IMP), Chinese Academy of Sciences	Sugon Cluster W780I, Xeon E5-2640v3 8C 2.6GHz, Infiniband QDR, NVIDIA Tesla K80	65
5	4112.1	Stanford Research Computing Center	Cray CS-Storm, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR, Nvidia K80	190
6	3856.9	IT Company	Inspur TS10000 HPC Server, Xeon E5-2620v3 6C 2.4GHz, 10G Ethernet, NVIDIA Tesla K40	58
7	3775.5	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110
8	3775.5	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110
9	3775.5	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110
10	3775.5	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110



# Intel MIC Hardware

INSPUR, NAOC - 2013.XI.26



icpc ... "-mmic" ...  $61 \times 4 = 244 \times 1.1$  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	
5110P	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	
	PCIe Card, Passively Cooled	> 1 TF			5.0	240	6	28.5	300	



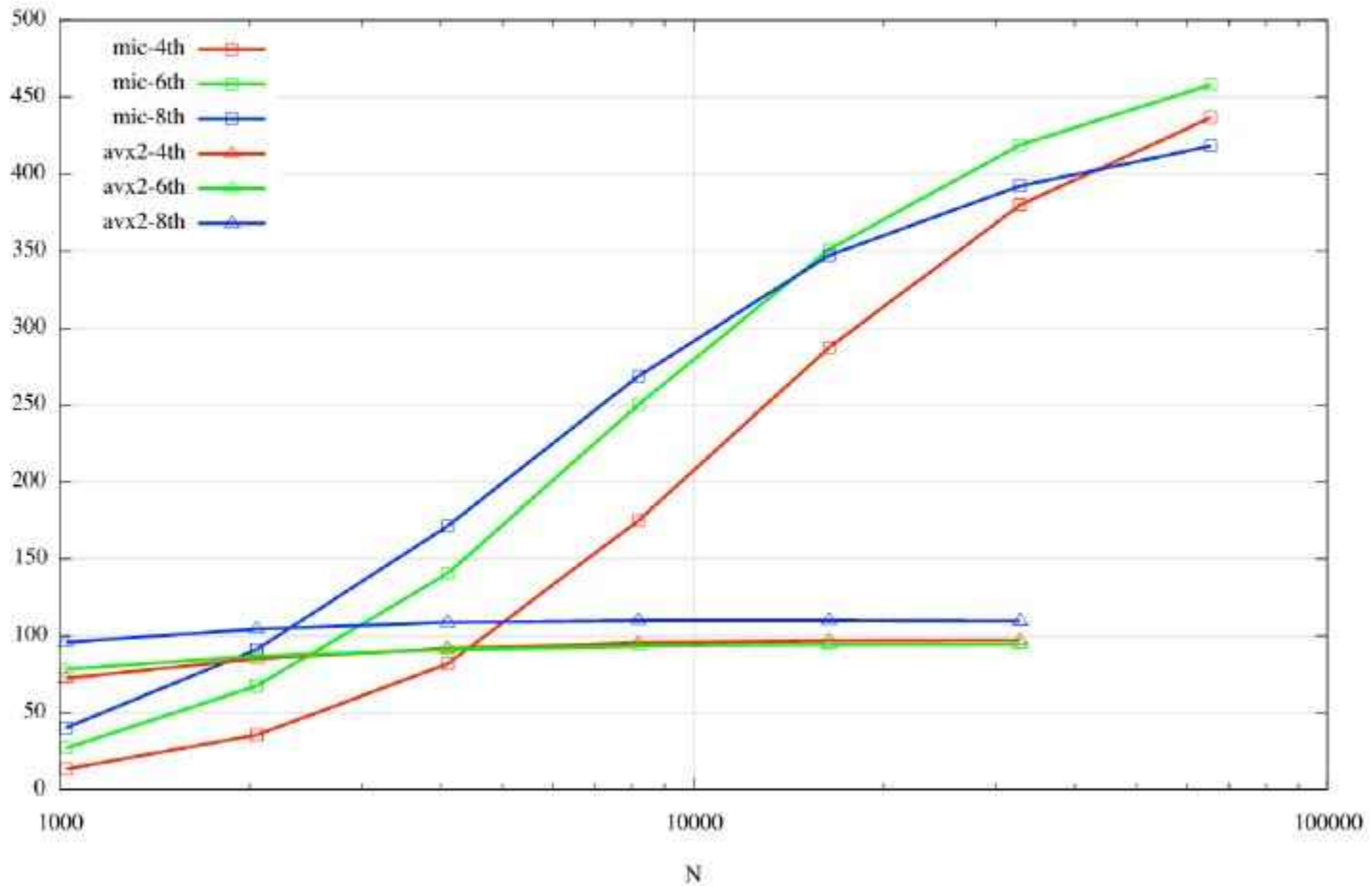
PCIe Card, Actively Cooled



PCIe Card, Passively Cooled

# $\phi$ GPU Hermite results

GFLOPS

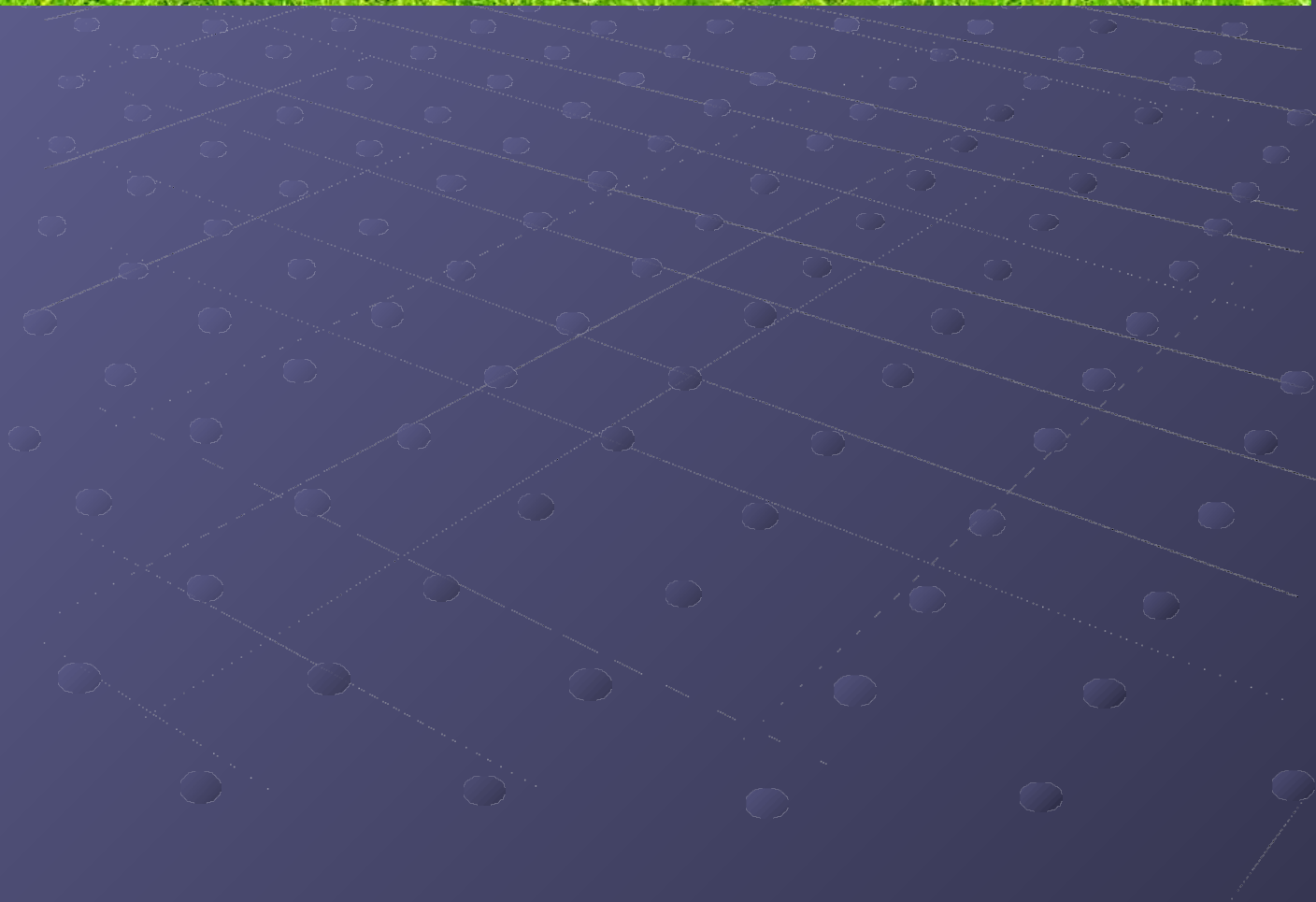




[www.green500.org](http://www.green500.org)

# Ranking the World's Most ENERGY-EFFICIENT SUPERCOMPUTERS

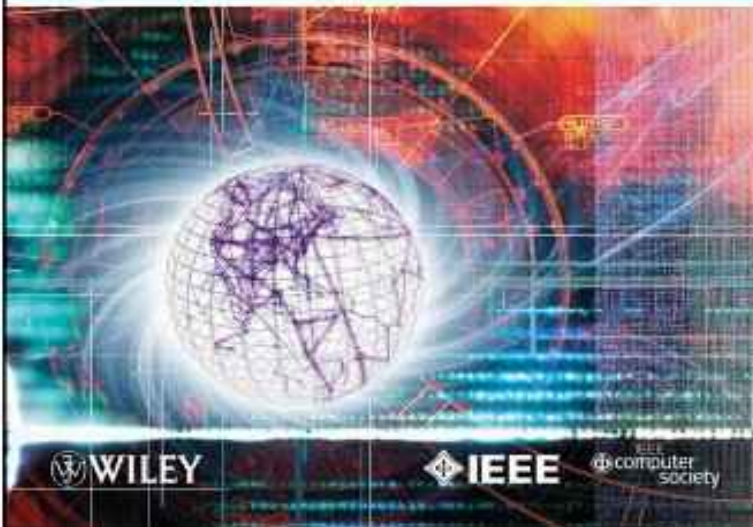
[www.green500.org](http://www.green500.org)



Wiley Series on Parallel and Distributed Computing • Albert Y. Zomaya, Series Editor

# LARGE-SCALE COMPUTING TECHNIQUES FOR COMPLEX SYSTEM SIMULATIONS

WERNER DUBITZKY • KRZYSZTOF KUROWSKI • BERNHARD SCHOTT



Edited Volume  
To appear November 22

With Paper

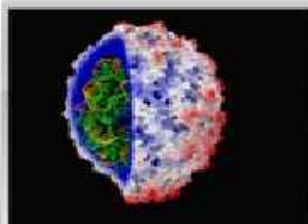
Spurzem, et al.,  
Accelerated Many-Core  
GPU Computing for  
Physics and Astrophysics  
on Three Continents

# Speedups using GPU vs. CPU



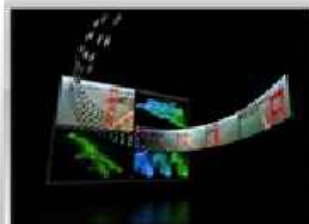
**146X**

Interactive visualization of volumetric white matter connectivity<sup>1</sup>



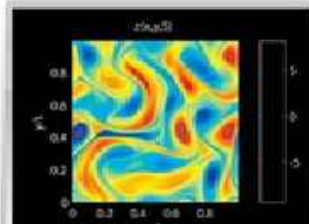
**36X**

Ionic placement for molecular dynamics simulation on GPU<sup>2</sup>



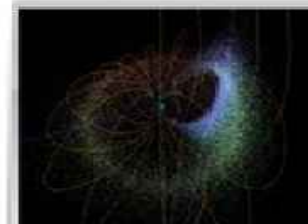
**18X**

Transcoding HD video stream to H.264 for portable video<sup>3</sup>



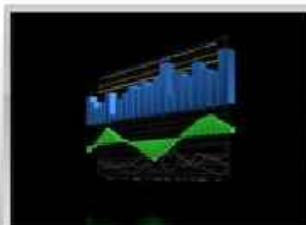
**17X**

Simulation in Matlab using mex file CUDA function<sup>4</sup>



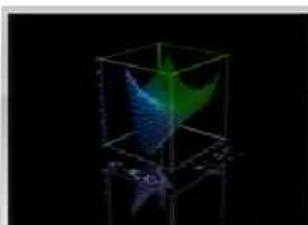
**100X**

Astrophysics N-body simulation<sup>5</sup>



**149X**

Financial simulation of LIBOR model with swaptions<sup>6</sup>



**47X**

GLAME@lab: M-script API for linear Algebra operations on GPU<sup>7</sup>



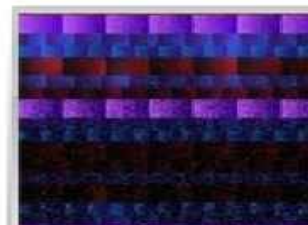
**20X**

Ultrasound medical imaging for cancer diagnostics<sup>8</sup>



**24X**

Highly optimized object oriented molecular dynamics<sup>9</sup>



**30X**

Cmatch exact string matching - find similar proteins & gene sequences<sup>10</sup>





# Towards Peta-Scale Green Computation

— applications of the GPU supercomputers in CAS

<http://www.nvidia.com/gtc2010-content>



## GPU TECHNOLOGY CONFERENCE

GTC 2010 | Sept 20-23, 2010

San Jose Convention Center, San Jose, California

Watch the Keynote Recordings

Algorithms & Numerical Techniques

Astronomy & Astrophysics

Audio Processing

Cloud Computing

Computational Fluid Dynamics

Computer Graphics

Computer Vision

Databases & Data Mining

Digital Content Creation

Embedded & Automotive

Energy Exploration

Film

Finance

General Interest

GPU Accelerated Internet

High Performance Computing

Imaging

Life Sciences

Machine Learning & Artificial

Intelligence

Medical Imaging & Visualization

Mobile & Tablet & Phone

Molecular Dynamics

Neuroscience

Physics Simulation

Programming Languages &

Techniques

Quantum Chemistry

Ray Tracing

Signal Processing

Stereoscopic 3D

Tools & Libraries

Video Processing

Wei Ge  
Xiaowei Wang

Inst. of Proc. Eng.



Yunquan Zhang

Inst. of Software



Rainer Spurzem

Nat. Astro. Obs.  
Chn.



Long Wang

SC Center

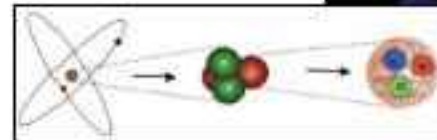
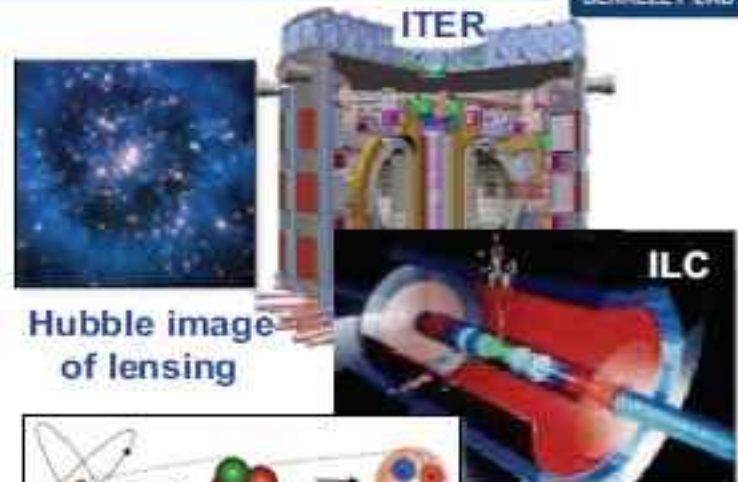


# Exascale simulation will enable fundamental advances in basic science

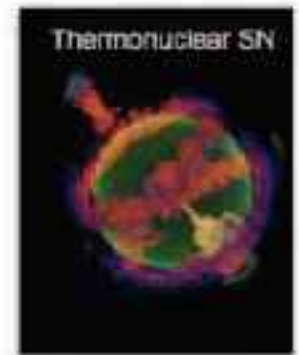


- High Energy & Nuclear Physics
  - Dark-energy and dark matter
  - Fundamentals of fission fusion reactions
- Facility and experimental design
  - Effective design of accelerators
  - Probes of dark energy and dark matter
  - ITER shot planning and device control
- Materials / Chemistry
  - Predictive multi-scale materials modeling: observation to control
  - Effective, commercial technologies in renewable energy, catalysts, batteries and combustion
- Life Sciences
  - Better biofuels
  - Sequence to structure to function

These breakthrough scientific discoveries and facilities require exascale applications and resources



Structure of nucleons





# Advanced Computation in Energy Science at LBNL



Probe natural systems under constraints that are difficult or impossible to impose in the field or laboratory

Reveal the manner in which large-scale phenomena arise from smaller-scale properties

Discover new materials for green technology applications through first-principles calculations

Global Scale Reactive Transport Modeling of CH<sub>4</sub> hydrates (M. Reagan)



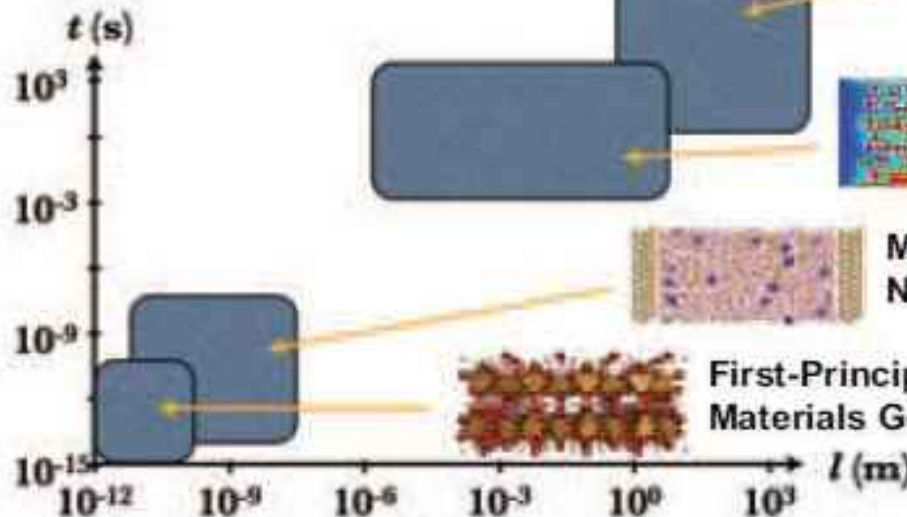
Pore Scale Reactive Transport Modeling of CO<sub>2</sub> sequestration (D. Trebotich)



Molecular Dynamics Simulations of Natural Nanofluids (I. Bourg)

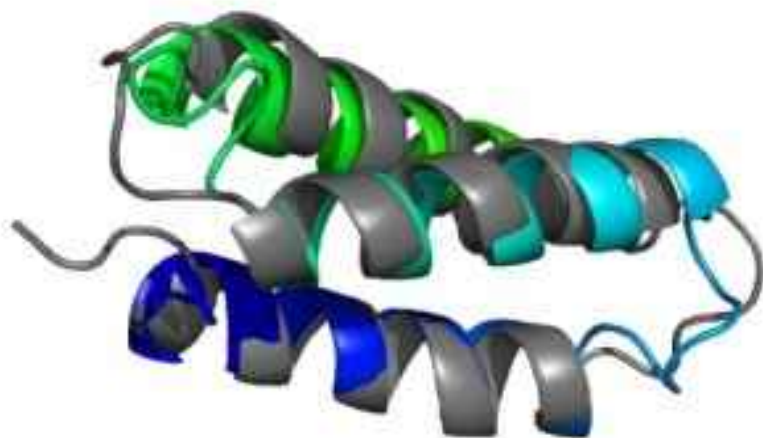


First-Principles Calculations of Materials Genome (K. Persson)





JSC's research and development concentrates on mathematical modelling and numerical, especially parallel algorithms for quantum chemistry, molecular dynamics and Monte-Carlo simulations. The focus in the computer sciences is on cluster computing, performance analysis of parallel programs, visualization, computational steering and grid computing.



## Modelling and Simulation

The simulation of complex systems in natural science or engineering depends on the development of adequate mathematical models. Thus the development of realistic and yet efficient models is a core activity at JSC. Examples of simulations are:

- Computational Plasma Physics
- Protein Folding
- Quantum Information Processing
- Civil Security and Traffic

## Algorithms and Methods

Efficient simulations need powerful algorithms and methods. JSC focusses on the development of the following methods:

- Fast Coulomb Solvers
- Parallel-In-Time Integration
- Fast Multipole Method
- Parallel I/O



Computer Physics - Astrophysics

**Molecular Dynamics**

# Fermi-based GPU supercomputer IPE (2010.04.24)

Rpeak SP : 2Pflops  
Rpeak DP : 1Pflops  
Linpack: 207.3T (Top500 19th)  
Mflops/Watt: 431 (Green500 8th)  
Total RAM : 17.2TB  
Total VRAM : 6.6TB  
Total HD : 360TB  
Inst. Comm. : H3C GE  
Data Comm. : Mellanox QDR IB  
Occupied  
area : 150 sq.m.  
Weight : 12.6 tons  
Max Power : 600kW(computing)  
200kW(cooling)  
System : CentOS 5.4, PBS  
Monitor : Ganglia, GPU monitor  
Languages : C, C++, CUDA 3.1 , OpenCL



中国科学院过程工程研究所

Institute Of Process Engineering, Chinese Academy Of Sciences



**IPE CAS 372 node 6xC2050 cluster**

**2232 GPU = 2.2 Pflops SP / 1.1 Pflops DP**

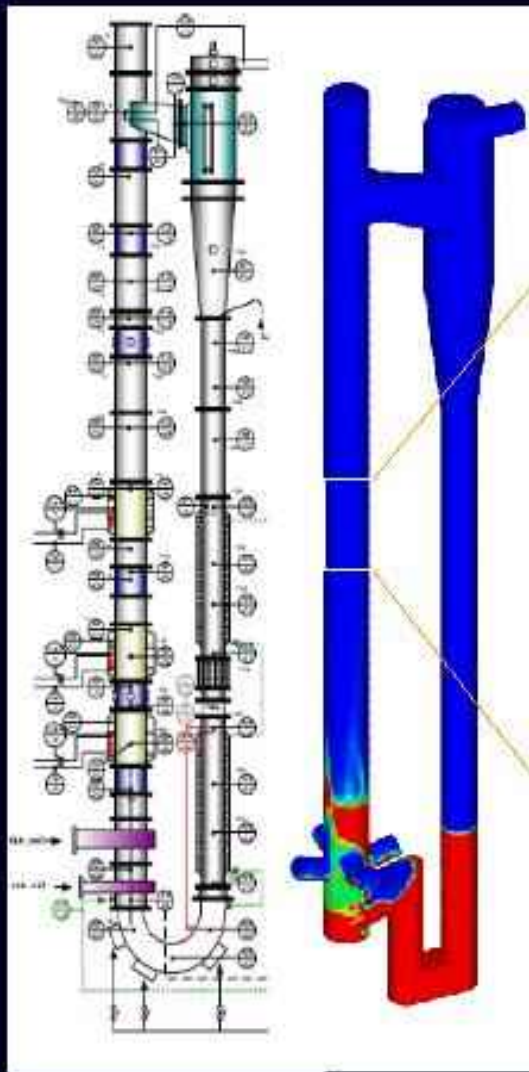




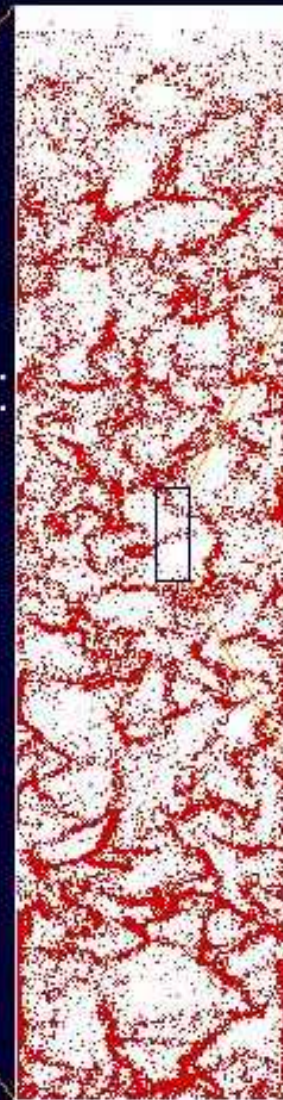
DNS of gas-solid flow : **>20x speedup** (1C1060/1E5430 core)

120K Particles + 400M pseudo-particles

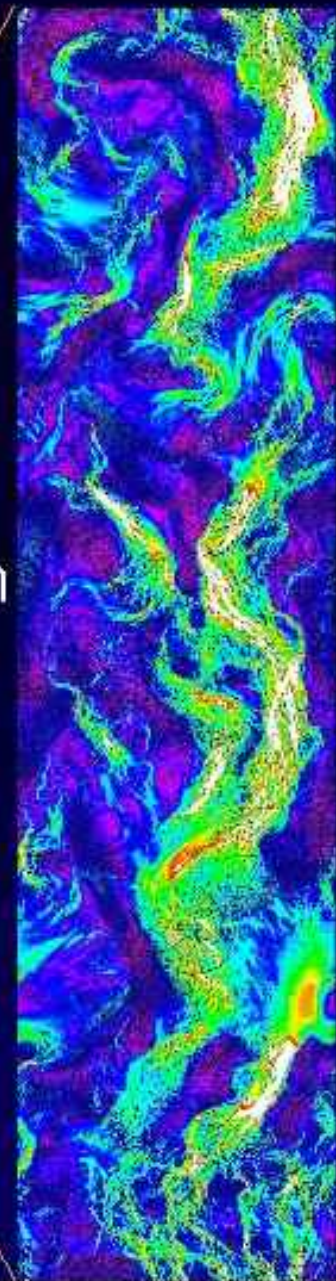
Reactor:  
0.4\*20m  
3D



Section:  
0.4\*1m  
2D



Cell:  
2\*10cm  
2D

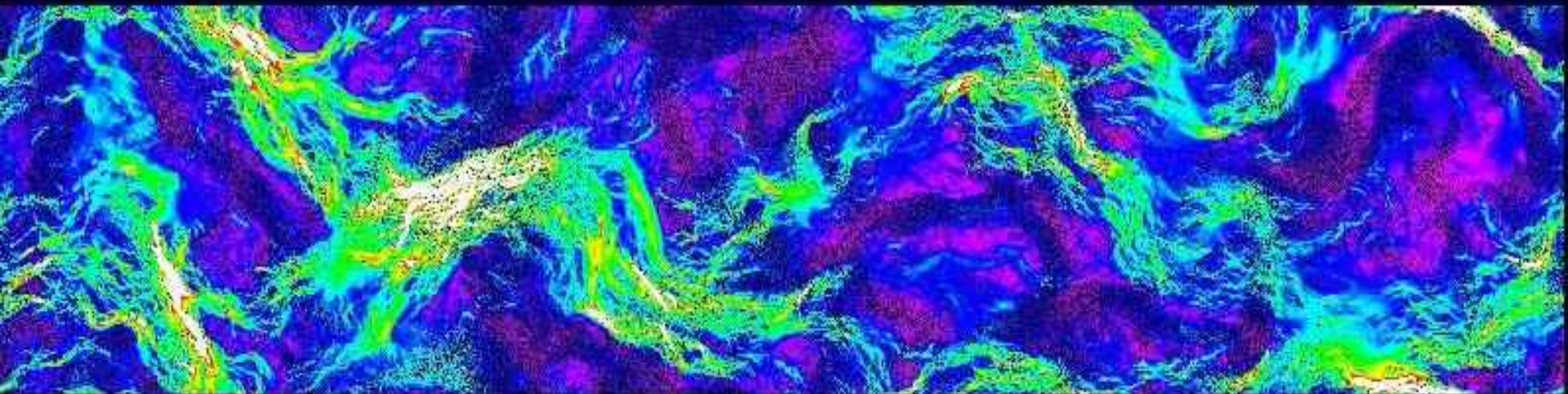




## Animation Challenge:

9600x2400 → 1200x300 pixels

1000 → 17 frames





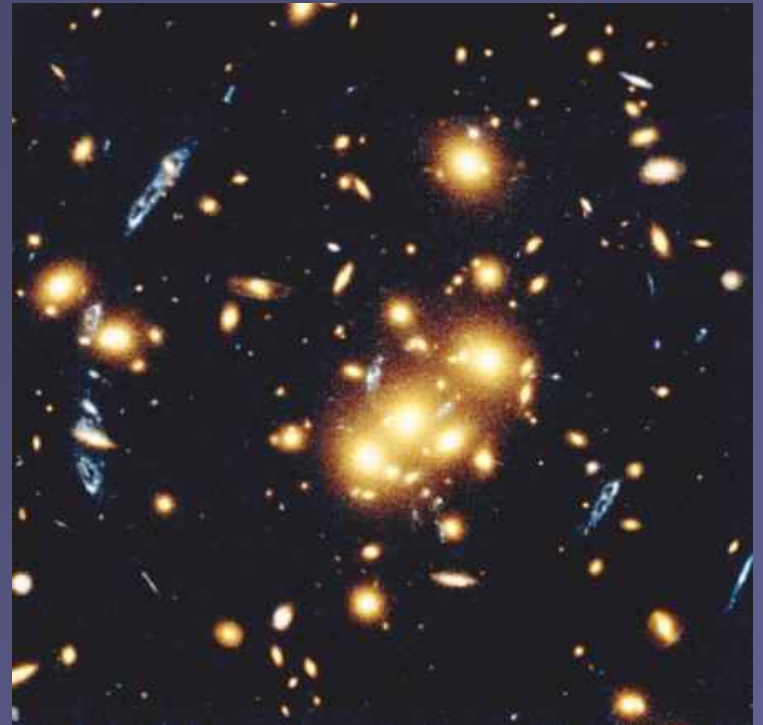
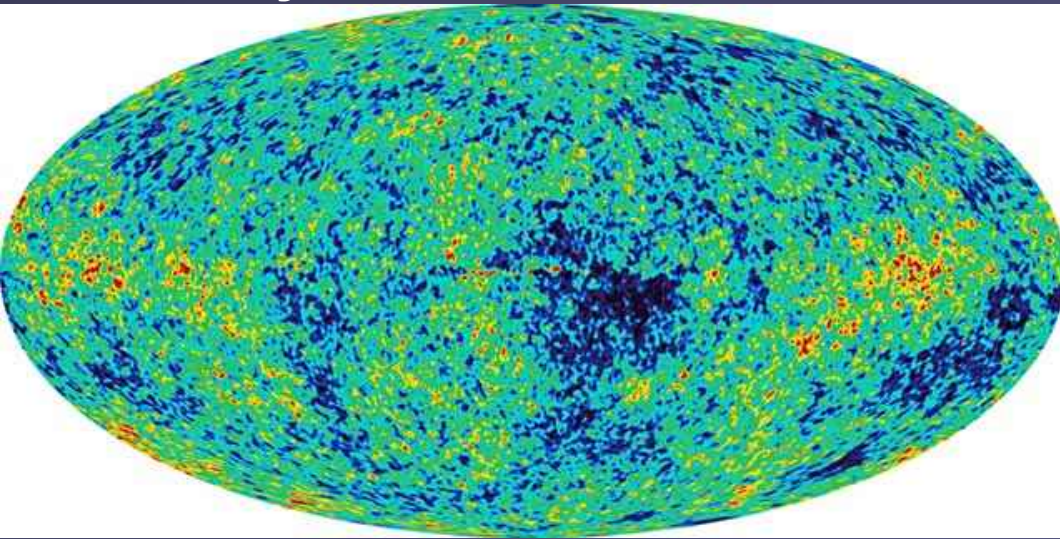
Computer Physics - Astrophysics

Cosmology

# Computer Physics - Astrophysics

## ● Structure Formation in the Universe

In the year 100.000....



● Wilkinson Microwave Anisotropy Probe (WMAP)  
(Cosmic Microwave Background)

...and ``today``



A visualization of the Millennium Simulation, showing a vast field of particles. The particles are color-coded by mass, with larger particles appearing in shades of red and orange, and smaller particles in shades of blue and purple. The overall structure is a complex, interconnected web of filaments and nodes, representing the large-scale structure of the universe. A horizontal double-headed arrow at the top indicates a scale of 1 Gpc/h.

1 Gpc/h

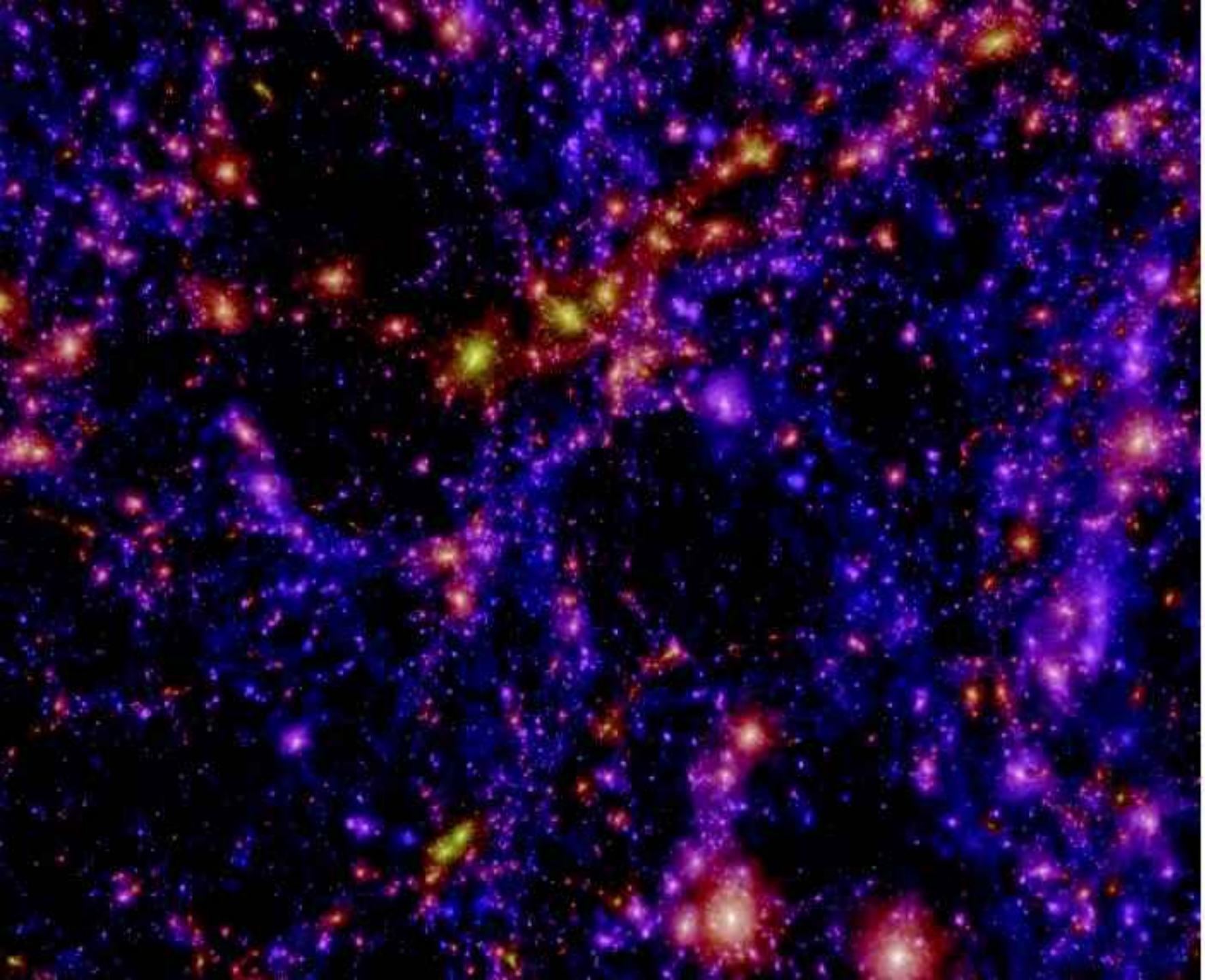
**Millennium Simulation**

10,077,696,000 particles

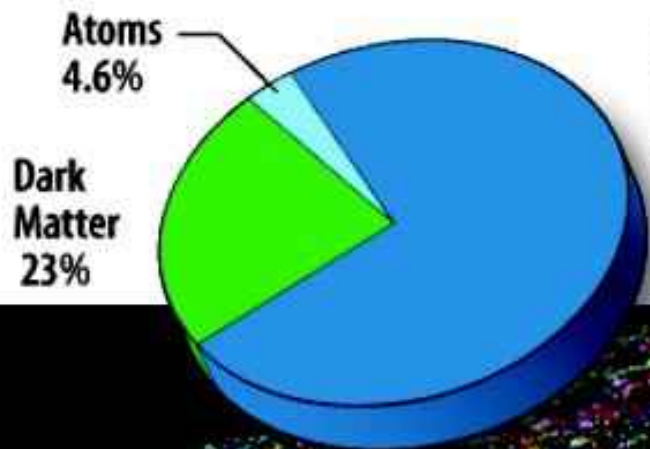
( $z = 0$ )



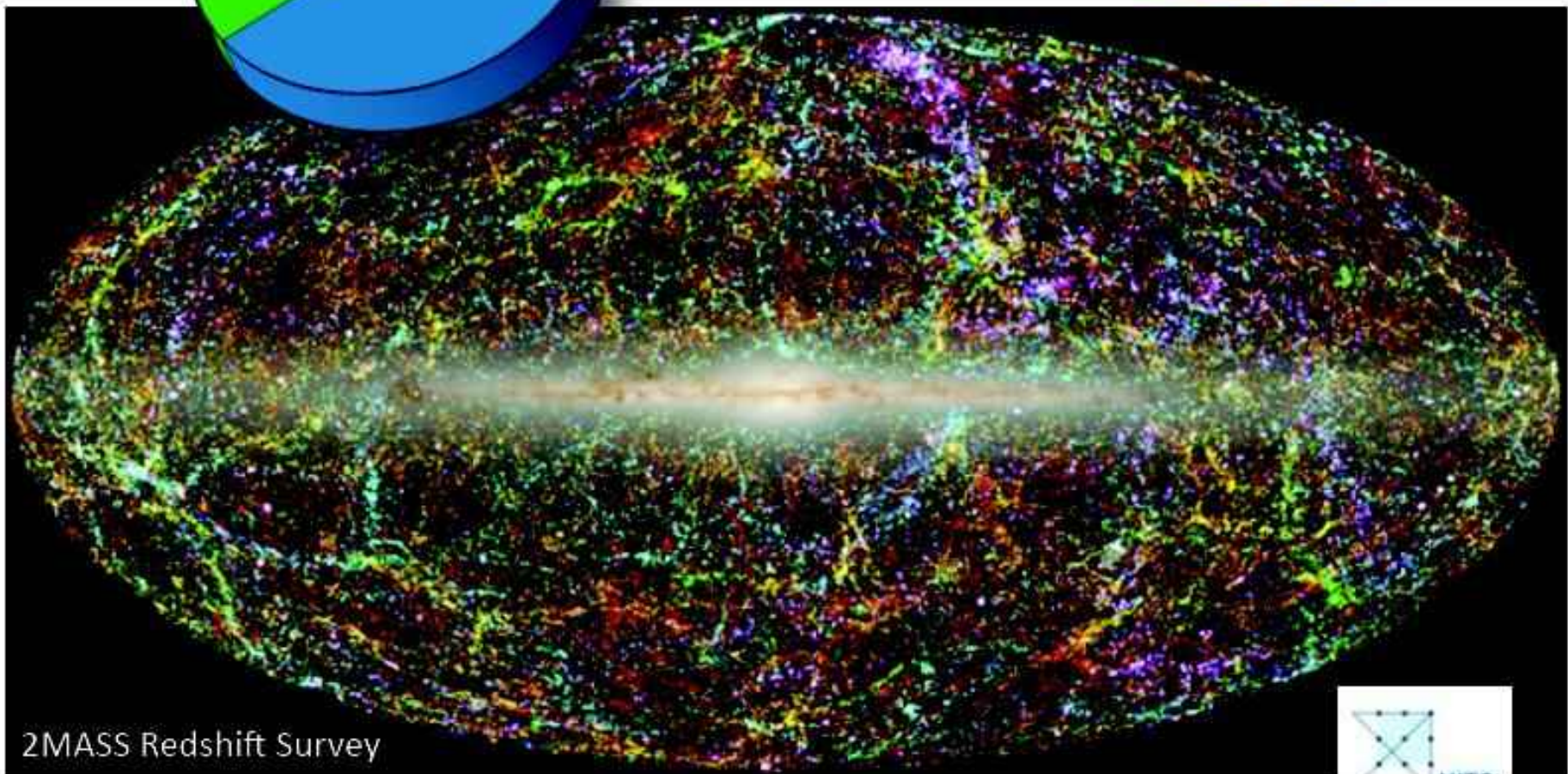
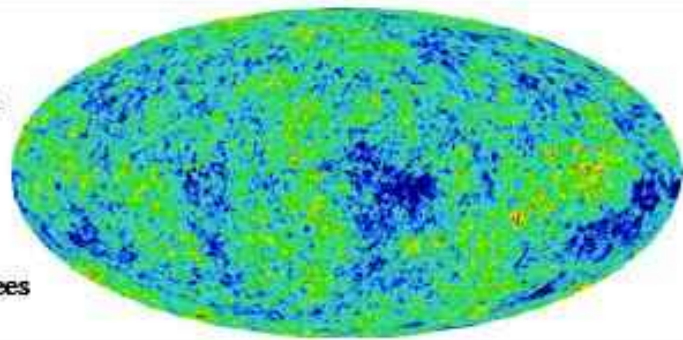
Millennium Simulation (Springel et al.)







WMAP  
2.725 Kelvin  
0.0002 degrees



2MASS Redshift Survey

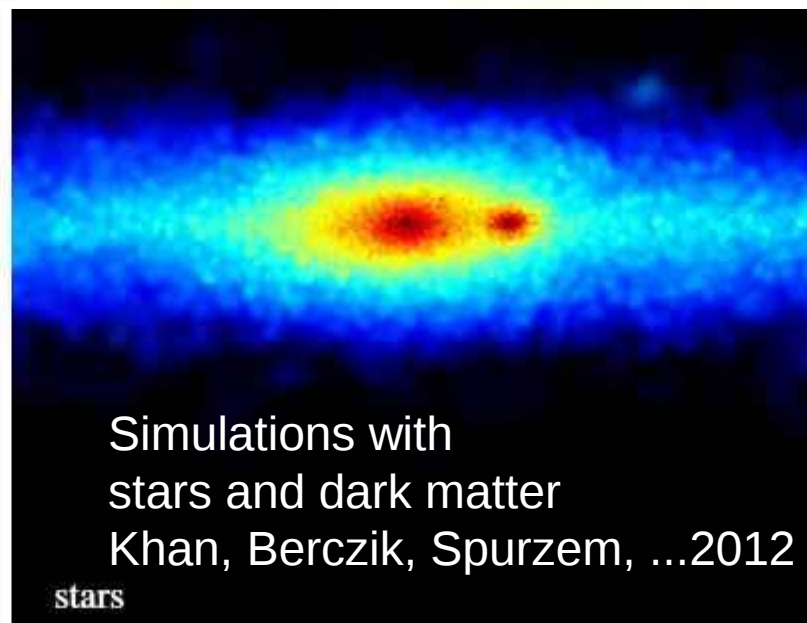
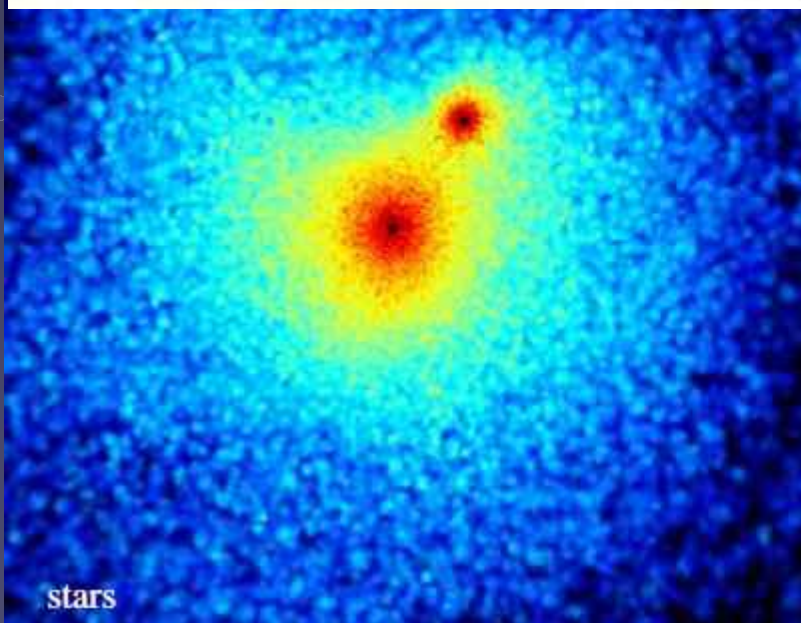
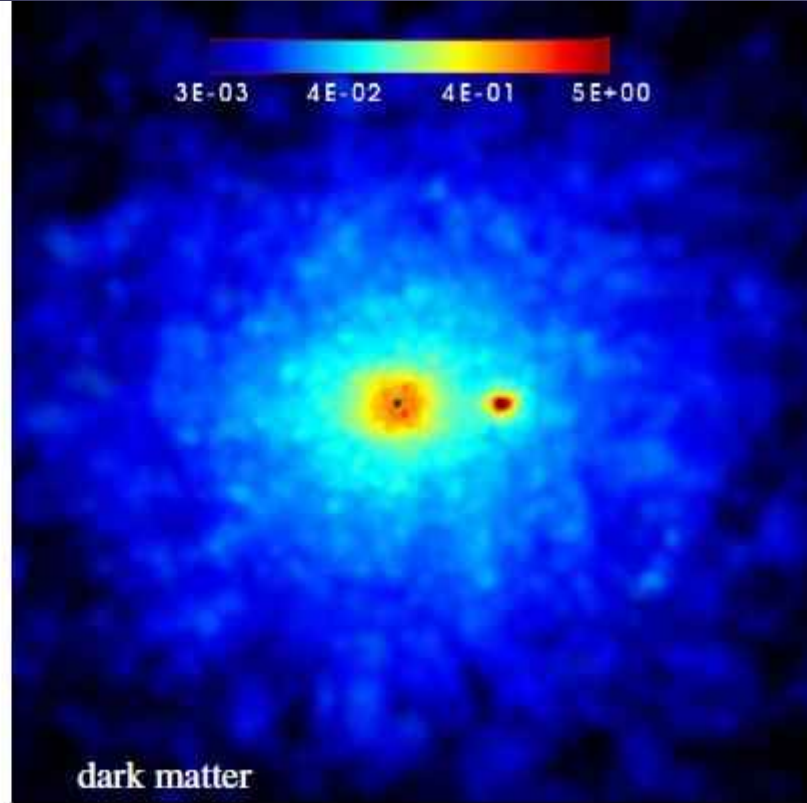
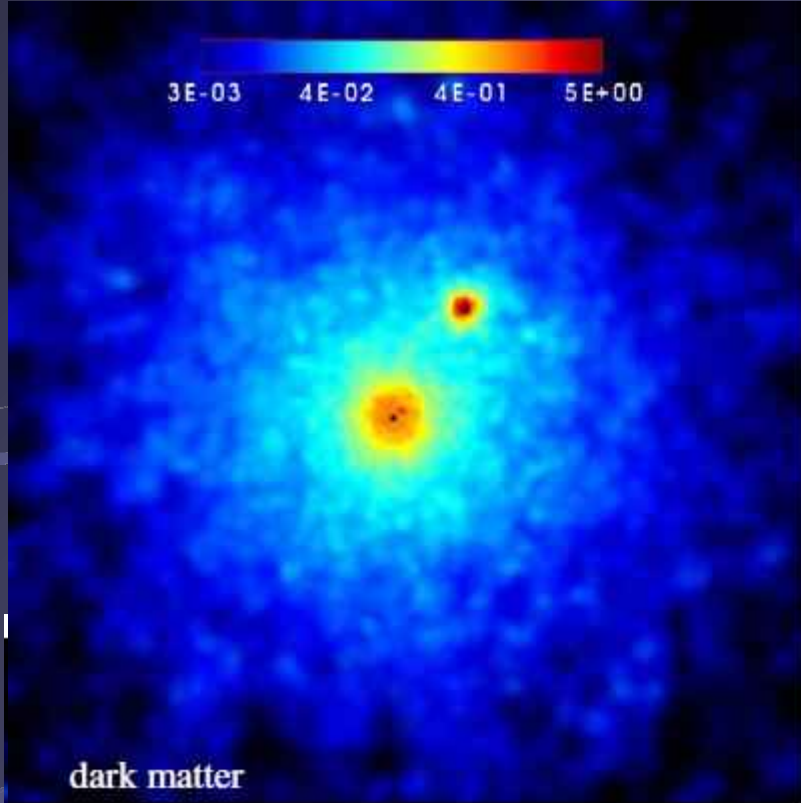
(Image: T.H. Jarrett (IPAC/SSC))



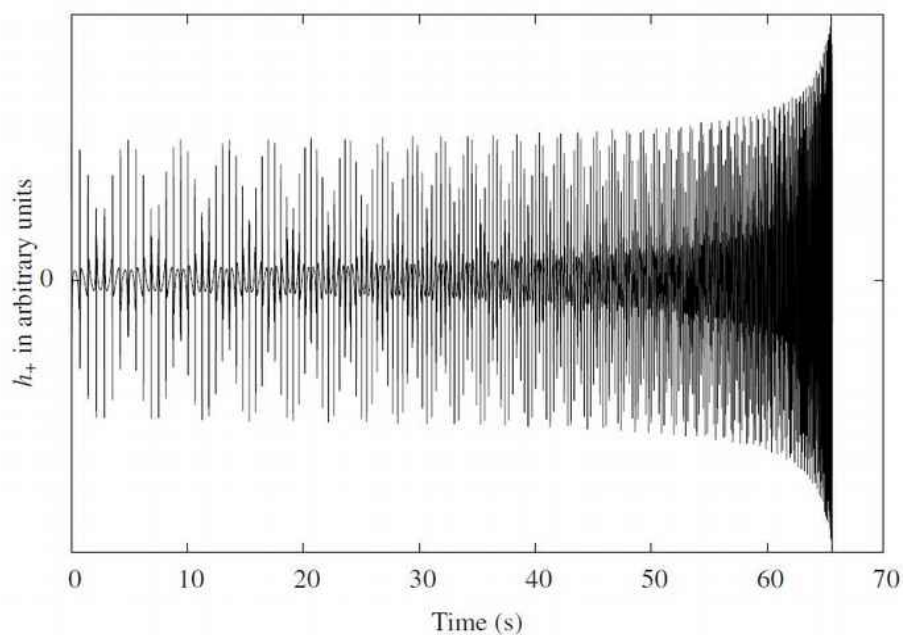
Computer Physics - Astrophysics

Black Holes in  
Galaxies and  
Star Clusters





Simulations with stars and dark matter  
Khan, Berczik, Spurzem, ...2012



# Post-Newtonian Dynamics Gravitational Wave Templates

Figure 3.11: Waveform for two equal mass objects on an orbit with  $e = 0.5$ .

Handle spin-orbit and  
spin-spin coupling  
(P. Brem, R. Spurzem,  
Univ. Heidelberg)

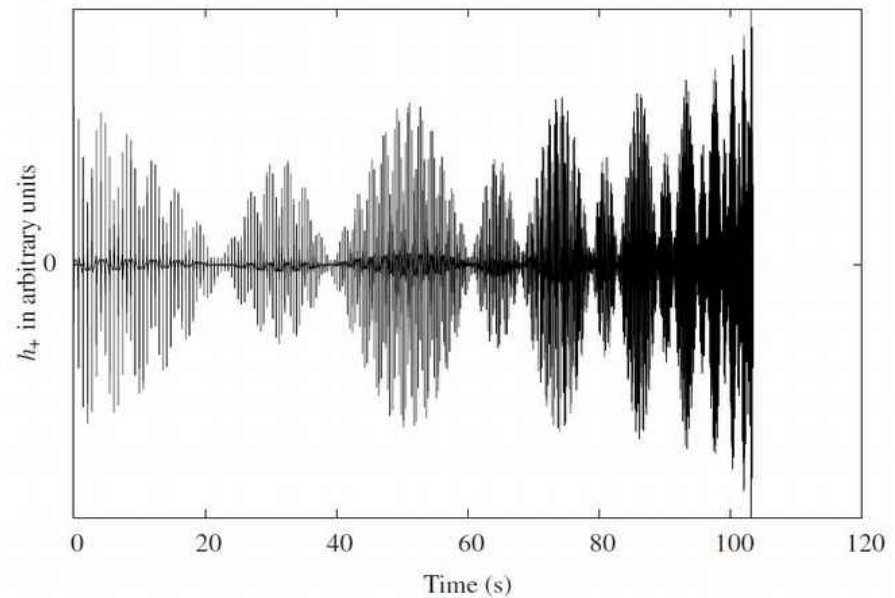


Figure 3.12: Waveform for two objects with a mass ratio of  $q = 1/10$  on an orbit with  $e = 0.5$  and spins  $a_{1,x} = 1.0$ ,  $a_{2,y} = 1.0$ .

EUROPEAN GRAVITATIONAL OBSERVATORY

EGO



Consortium of

Example: VIRGO Detector in Cascina near Pisa, Italy







VIRGO – Pisa 3km  
LIGO – Livingston, LA  
Hanford, WA  
1km  
GEO600 – Hannover  
600m  
AIGO – Australien  
(planned, 5 km)

<http://www.ligo-la.caltech.edu/>  
<http://www.ego-gw.it>  
<http://www.geo600.uni-hannover.de>

Outreach to 50 Millionen  
light years (Neutron Stars)