최근 슈퍼컴퓨팅 개발 동향 : Peta, Green, Multi-core

Dr. Hongsuk Yi
KISTI Supercomputing Center

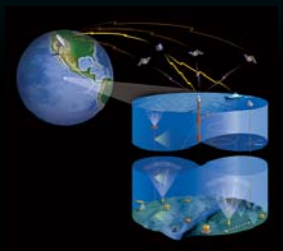
목차

- Supercomputer and Supercomputing
  - Overview of Recent Supercomputers
- Petaflop Supercomputing Project
  - Green
  - Multi-core and GPGPU
- Supercomputing Development in KISTI

슈퍼컴퓨터란 무엇인가?

- 당대에 최고로 빠른 컴퓨터로
  - 가장 어렵고, 큰 규모의 문제에 도전하는 시스템
  - Top500 목록에 포함됨 www.top500.org

CyberInfrastructure

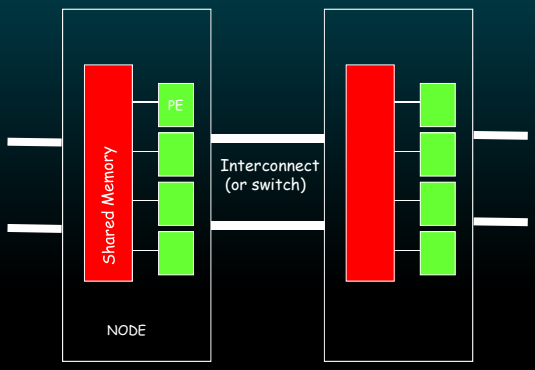


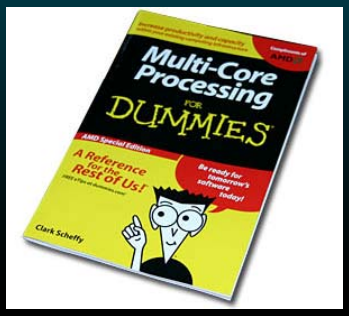
High Performance Computing

최근 2009년 슈퍼컴퓨터....

- 일반 노트북 PC와 비교하면
  - 1000 배 이상 빠른 성능
  - 1000 배 많은 메모리
  - 10,000 ~ 200,000 배 많은 CPU 수
  - 소비전력은 1MegaWatt (전구 1만개 정도)

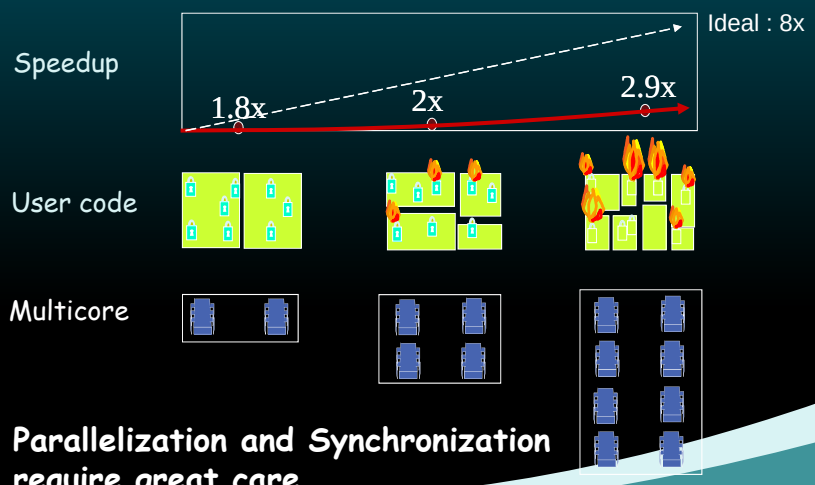
Multi-cores Are Here





5

Real-World Scaling Process



Parallelization and Synchronization require great care…

The Patriot Missile Failure

Patriot가 이라크미사일 요격 실패
- 미국 28명 사망 (1991.2.25)
- Patriot에서 24비트 처리로 시간지연
- 0.34초 지연 발사

0.0000000000000000000000011001100...

- 1996년 6월 4일 프랑스 Ariane5 로켓 폭발
- 10년간 총 7조원이 소요
- 속도계산이 64비트 부동소수-> 16비트 처리

Source: http://www.ima.umn.edu/~arnold/disasters/patriot.html

Recent Supercomputers Overview : Top500 List (2008.11)

Supercomputer Ranking

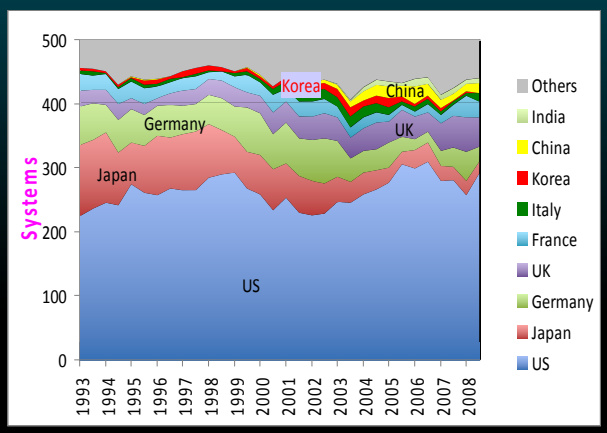
- Started in 1993
- Jack Dongarra, University of Tennessee
- Based on LINPACK benchmark
  - linear algebra (LU factorization)
- Superseded by LAPACK
  - based on BLAS (Basic Lin. Alg. Subprograms)
  - exploits caches
- Measures Floating Point performance
- Fortran code
- see www.top500.org

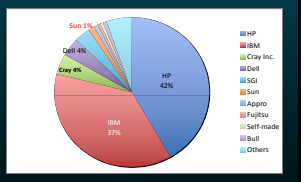
Supercomputing Top500:10

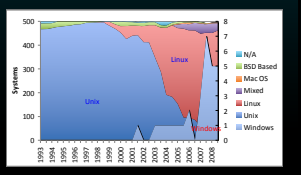
Tera : Tflop/s = 10^{12} flop/s; Peta : Pflop/s = 10^{15} flop/s

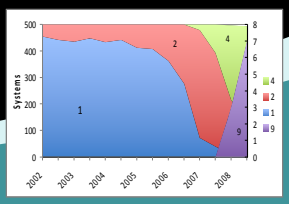
Rank	Site	Manufacturer	Computer	Country	Cores	Rmax [Tflops]	Power [MW]
1	DOE/NNSA/LANL	IBM	Roadrunner - BladeCenter QS22/LS21	USA	129600	1105.0	2.48
2	Oak Ridge National Laboratory	Cray Inc.	Jaguar - Cray XT5 QC 2.3 GHz	USA	150152	1059.0	6.95
3	NASA/Ames Research Center/NAS	SGI	Pleiades - SGI Altix ICE 8200EX	USA	51200	487.0	2.09
4	DOE/NNSA/LLNL	IBM	eServer Blue Gene Solution	USA	212992	478.2	2.32
5	Argonne National Laboratory	IBM	Blue Gene/P Solution	USA	163840	450.3	1.26
6	TACC	Sun	Ranger - SunBlade x6420	USA	62976	433.2	2.0
7	NERSC/LBNL	Cray Inc.	Franklin - Cray XT4	USA	38642	266.3	1.15
8	Oak Ridge National Laboratory	Cray Inc.	Jaguar - Cray XT4	USA	30976	205.0	1.58
9	NNSA/Sandia National Laboratories	Cray Inc.	Red Storm - XT3/4	USA	38208	204.2	2.5
10	Shanghai Supercomputer Center	Dawning	Dawning 5000A, Windows HPC 2008	China	30720	180.6	

Architecture and Country





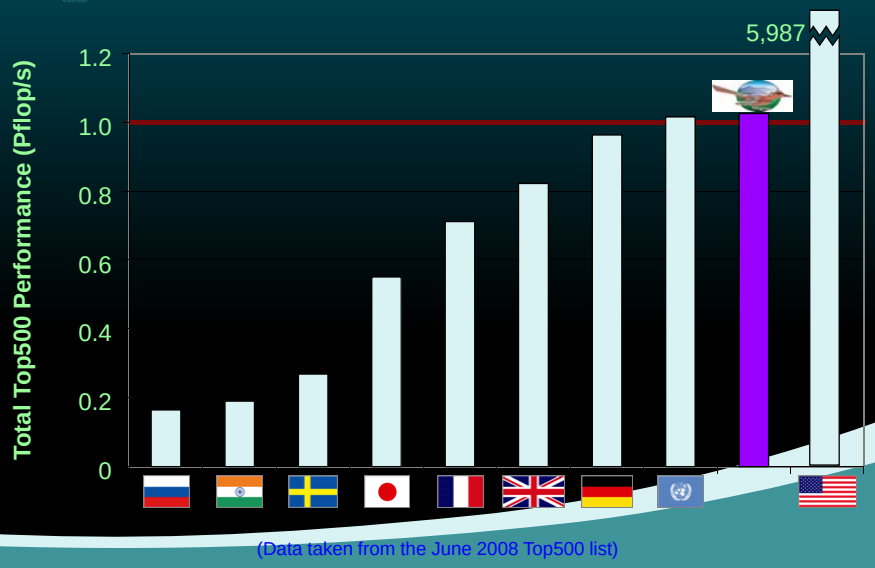




Roadrunner (2009년 1위)

- First supercomputer to achieve
  - 1 Pflop/s on LINPACK
  - 1.38 Pflop/s peak, 1.026 Pflop/s on LINPACK (June 2008)
- First #1 system to use
  - InfiniBand, Linux, Cell+Opteron

Roadrunner Performance

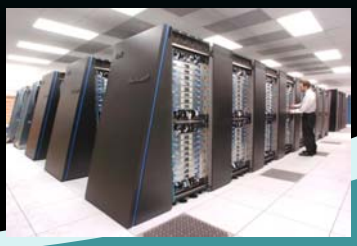


(Data taken from the June 2008 Top500 list)

IBM Blue Gene L/P

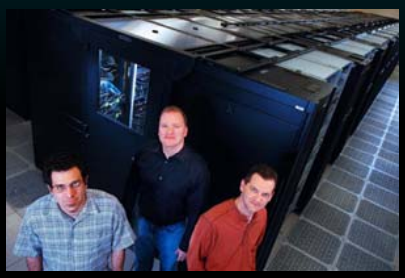
- BlueGene/L@LLNL (#4,2008.11)
  - 478 TFlops
  - Large number of modest processors
  - PowerPC 440 700 Mhz
  - 212,992 cores
  - 3D torus network

- BlueGene/P@ANL (#5,2008.11)
  - 450 TFlops
  - Large number of modest processors
  - PowerPC 450 – 850 Mhz
  - 163,840 cores
  - 3D torus network



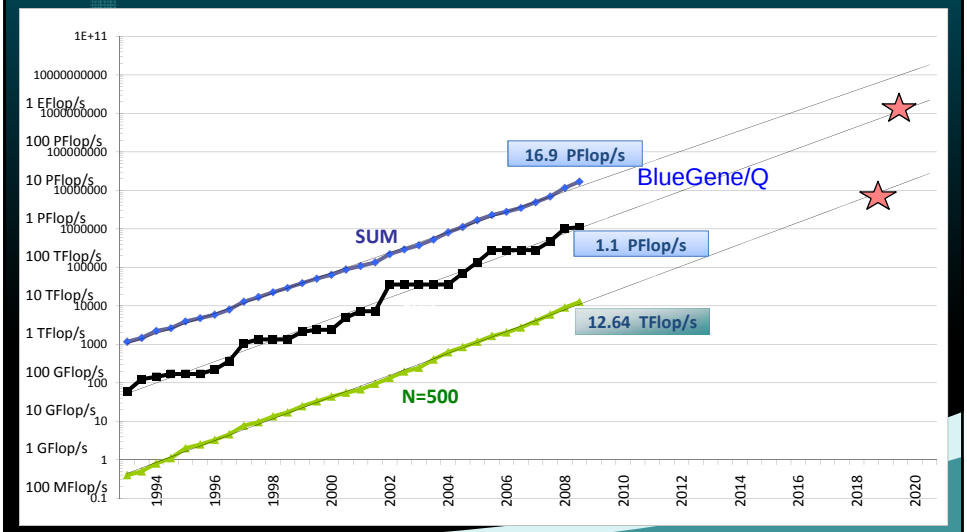
Sun Constellation – Ranger

- TACC Ranger (# 6)
  - 326 TFlops
  - AMD Quad-core Opteron 2 GHz Blades
  - 62,976 cores
  - Full-Clos Infiniband
- KISTI Tachyon(AMD Barcelona 2.0GHz)



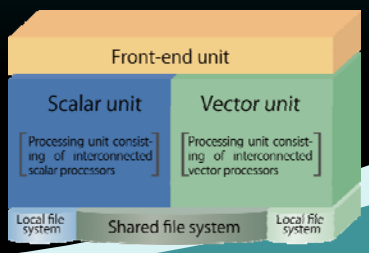
The Billion Dollar Race to a PetaFlop : Japan, US, ..., and KISTI

Petaflop 성능 예측



Japan : NGSC Project

- Next Generation Supercomputer Project
  - 京速計算機 : 京~10^{16} : 10 PF
    - 3-4 PetaFlops in 2010 (official target)
    - 1000 PF in **2017-2020**
  - 비용 : **$1 billion (총 1.3 조원)**
  - 기간 : **2006~2012**
  - **Fujitsu 와 NEC+Hitachi**

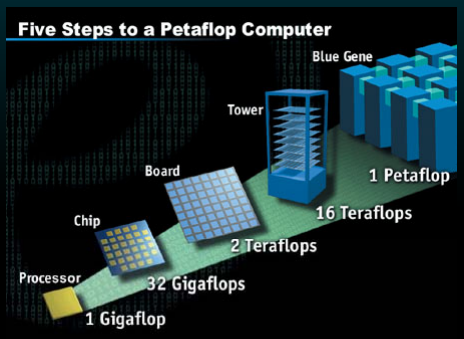


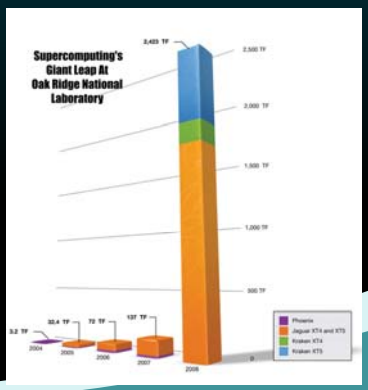
Japan Supercompter Project

- NGSC@2010: 10PF@10MW
  - 100GF/chip(100W)x10^5 (45-32nm)
- Exaflops@2017 : 1000PF@10MW
  - 100GF/chip(1W)x10^7 (11-22nm)
  - 1TF/chip(10W)
  - 100 times improvement required

US Supercomputer Project

- Multi PetaFlops system in 2010-2011





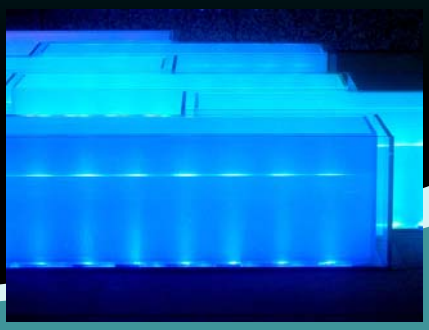
First Sustained PetaFLOP Supercomputer

- NSF's IBM Blue Waters at NCSA
  - On-Line 2011
  - 200,000 Processor Cores, 1 PetaByte Memory
  - 10 PetaBytes of Rotating Storage
  - ¼ of UIUC Electricity

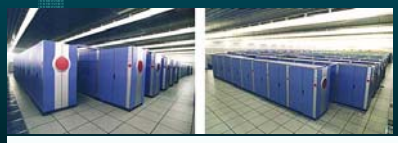
BLUE WATERS
SUSTAINED PETASCALE COMPUTING

IBM Blue Gene/Q project

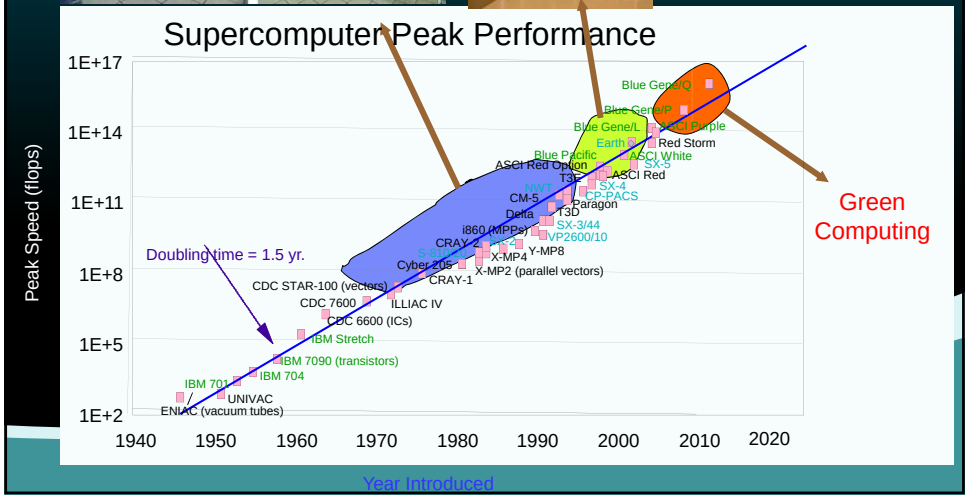
- Sequoia Supercomputer in 2012
  - 20PF@6MW
  - LLNL@13~30MW (additional plants)
  - 핵탄두 개발 등 핵 중심 연구!
  - 160만개 cores와 45nm



Supercomputing Trends



IBM BlueGene/L
소비전력~1.3MW



Looming Power Crisis

- New Constraints
  - Power limits clock rates

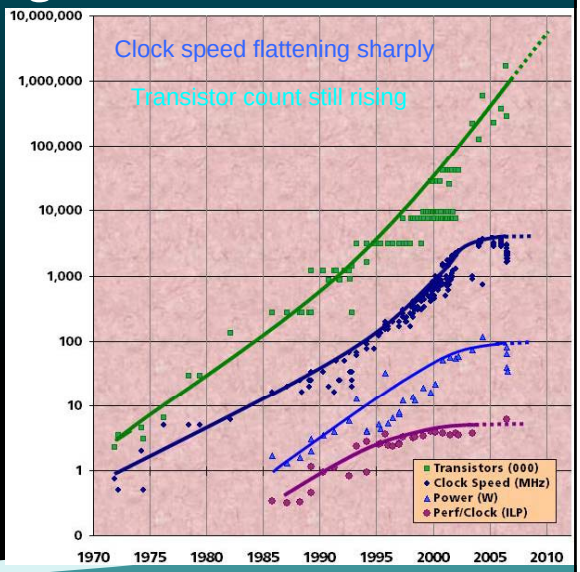
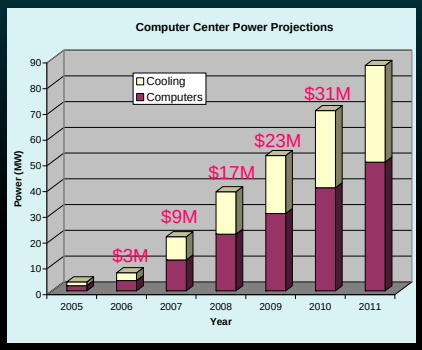


Figure courtesy of Kunle Olukotun, Lance Hammond, Herb Sutter, and Burton Smith

Power is an Industry wide Problem

ORNL Computing power and Cooling



$50 MW/hr (36,000 $/월)

Google's Data center

Cheap power and cooling!

Need 40-50 MW for projected systems
Cost estimates based on $0.05 kW/hr

Green 500 List (As of 30 June, 2008)

- Top 10 Systems are IBM
- 46 out of Top 50 Systems are IBM

Green500 Rank	MFLOPS/W	Site	Computer	Total Power (kW)	TOP500 Rank
1	488.14	IBM Germany	BladeCenter QS22 Cluster	22.76	324
1	488.14	Fraunhofer ITWM	BladeCenter QS22 Cluster	18.97	464
3	437.43	DOE/NNSA/LANL	BladeCenter QS22/LS21 Cluster	2345.50	1
4	371.75	Argonne National Laboratory	Blue Gene/P Solution	31.50	304
4	371.75	Dublin Institute for Advanced Studies/ICHEC	Blue Gene/P Solution	31.50	305
4	371.75	Science and Technology Facilities Council - Daresbury Laboratory	Blue Gene/P Solution	31.50	306
7	371.67	RZG/Max-Planck-Gesellschaft MPI/IPP	Blue Gene/P Solution	94.50	52
7	371.67	Stony Brook/BNL, New York Center for Computational Sciences	Blue Gene/P Solution	63.00	75
7	371.67	ASTRON/University Groningen	Blue Gene/P Solution	94.50	51
7	371.67	IBM - Rochester	Blue Gene/P Solution	126.00	37
7	371.67	DOE/Oak Ridge National Laboratory	Blue Gene/P Solution	63.00	74

Source: www.green500.org

26

KISTI Supercomputing Project

KISTI 슈퍼컴퓨팅센터 Vision

Keep securing/providing world-class supercomputing systems

Resource Center

Help Korea research communities to be equipped with proper knowledge of supercomputing

Technology Center

High Performance Computing & Networking Supercomputing Center
VISION PRODUCT INNOVATION 1967

Testbed Center

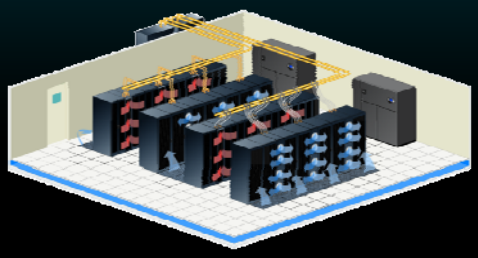
Validate newly emerging concepts, ideas, tools, and systems

Value-adding Center

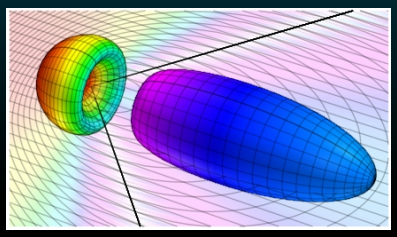
Make the best use of what the center has, to create the new values

KISTI 4th Tachyon Phase I

타키온 (TACHYON)
24 TeraFlops : 1차 2008.4
300 TeraFlops : 2차 2009.9



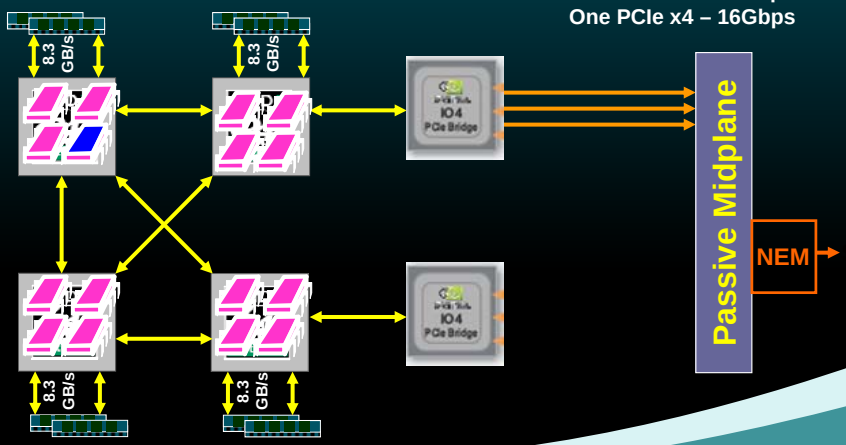
tachyon is any hypothetical particle that travels faster than light.



Multi-cores Are Here

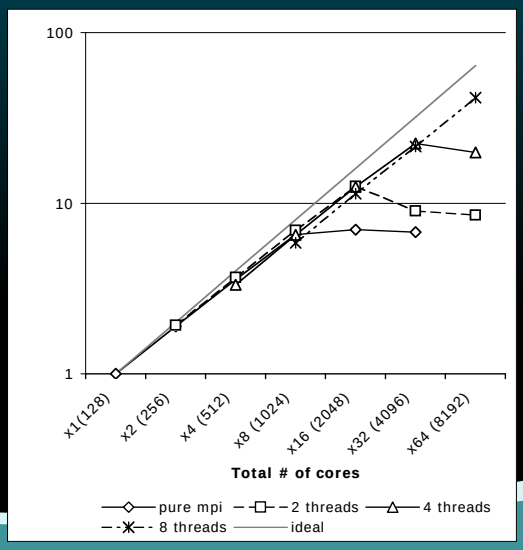
AMD Barcelona (Quad-core : 2.0 GHz)

Two PCIe x8 – 32Gbps
One PCIe x4 – 16Gbps



30

Hybrid Computing : Ranger



Hybrid-Computing
-OpenMP+MPI
- Good Scalability
- NPB BT-MZ3.3 Class E

4호기 2차 시스템 : SUN

- 타키온 2차 시스템
  - 300 Tflops 예정
    - Intel Nehalem 2.93 GHz
    - AMD Shanghai 2.90 GHz

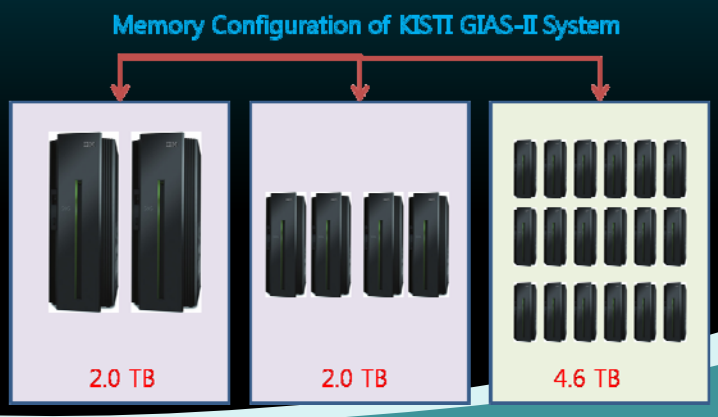
KISTI 4th GAIA : IBM

- IBM POWER6 p595+
  - 5.88 TeraFlops : GAIA 1차 (KISTI,2008)
  - 40 TeraFlops : GAIA 2차 (KISTI,2009)

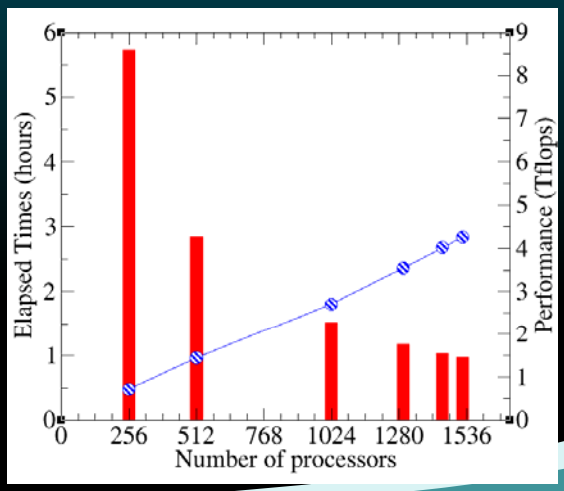
- p5 595+ system
  - 64way SMP System
  - POWER5+ Chip
  - MCM(MultiChip-Module)
  - 2.3GHz Clock Speed
  - 9.2GFLOPS / CPU core
  - 588.8GFLOPS / node

4호기 2차 시스템 : IBM

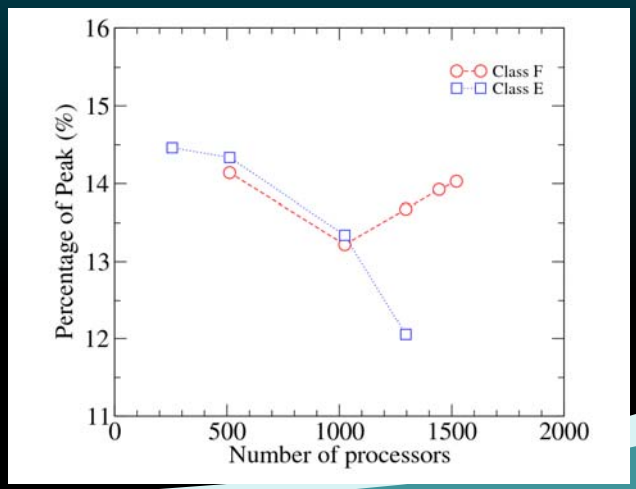
- IBM POWER6 (p595)



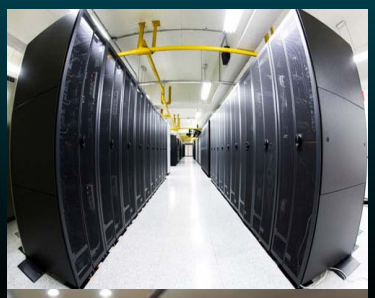
GAIA-II 테스트



NBP-MZ TEST



KISTI Visualization Computer



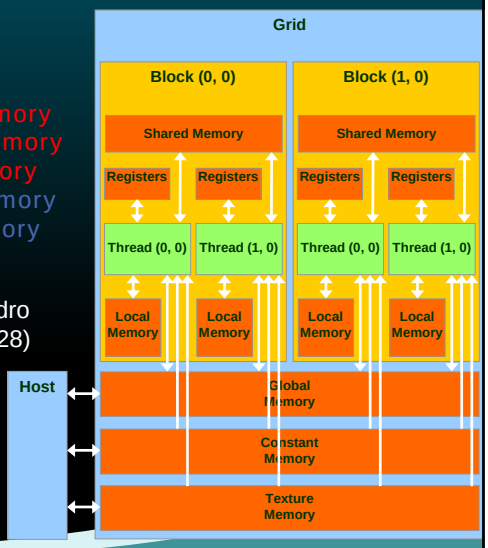
Picasso

R_{peak}	10.4 Tflops
# of computing nodes	90
CPU processor	Intel Xeon x5450@3GHz
# of CPUs	720
Core/Socket/Node	4/2/90
Memory	32GB/node; 4GB/core
GPU	NVIDIA Quadro FX5600
Switch	Infiniband (BW:20Gpbs)

Memory Model

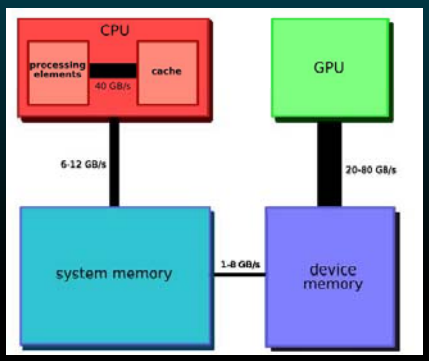
- Each thread can:
  - Read/write per-thread registers
  - Read/write per-thread local memory
  - Read/write per-block shared memory
  - Read/write per-grid global memory
  - Read only per-grid constant memory
  - Read only per-grid texture memory

NVIDIA Quadro FX5600 (#128)



Courtesy NVIDIA

New Hybrid Computing Model



Optimised GPU Kernel

Multiply Square Matrices

	CPU (ms)	GPU (ms)	Optimised GPU (ms)
256	34.9	11.1	0.7
512	286	75.2	5.3
1024	5930	569	41
2048	116169	4465	327
4096	1384086	35498	2640

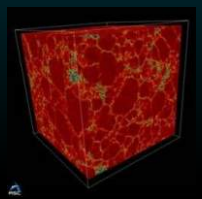
Speed up of around 500x

GPU x40

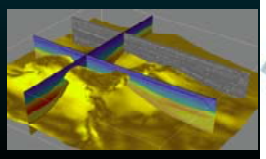
GPU SMP x500

Who needs more computing power?

- Improve understanding – significantly larger scale, more complex and higher resolution models; new science applications
- Multiscale and multiphysics – From atoms to mega-structures; coupled applications
- Shorter time to solution – Answers from months to minutes

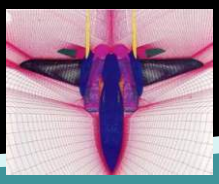


Physics Materials Science Molecular Dynamics

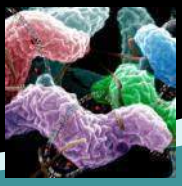


Geophysical Data Processing Upstream Petroleum

Computational Fluid Dynamics

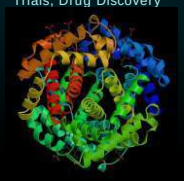


Biological Modeling – Brain Science



Life Sciences: Sequencing

Life Sciences: In-Silico Trials, Drug Discovery



Financial Modeling Streaming Data Analysis

40

감사합니다.

Q&A