GPU Club presentation on Friday 15 July (2pm)
in the John Casken Theatre, Martin Harris Centre for Music and Drama.

Emerging Technologies for High Performance Computing

Jack Dongarra

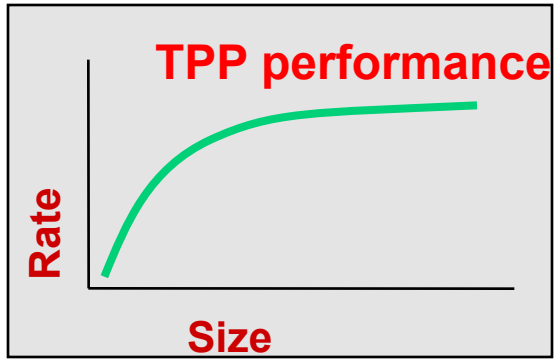
University of Tennessee
Oak Ridge National Laboratory
University of Manchester

Top500 List of Supercomputers

H. Meuer, H. Simon, E. Strohmaier, & JD

- Listing of the 500 most powerful Computers in the World
- Yardstick: Rmax from LINPACK MPP

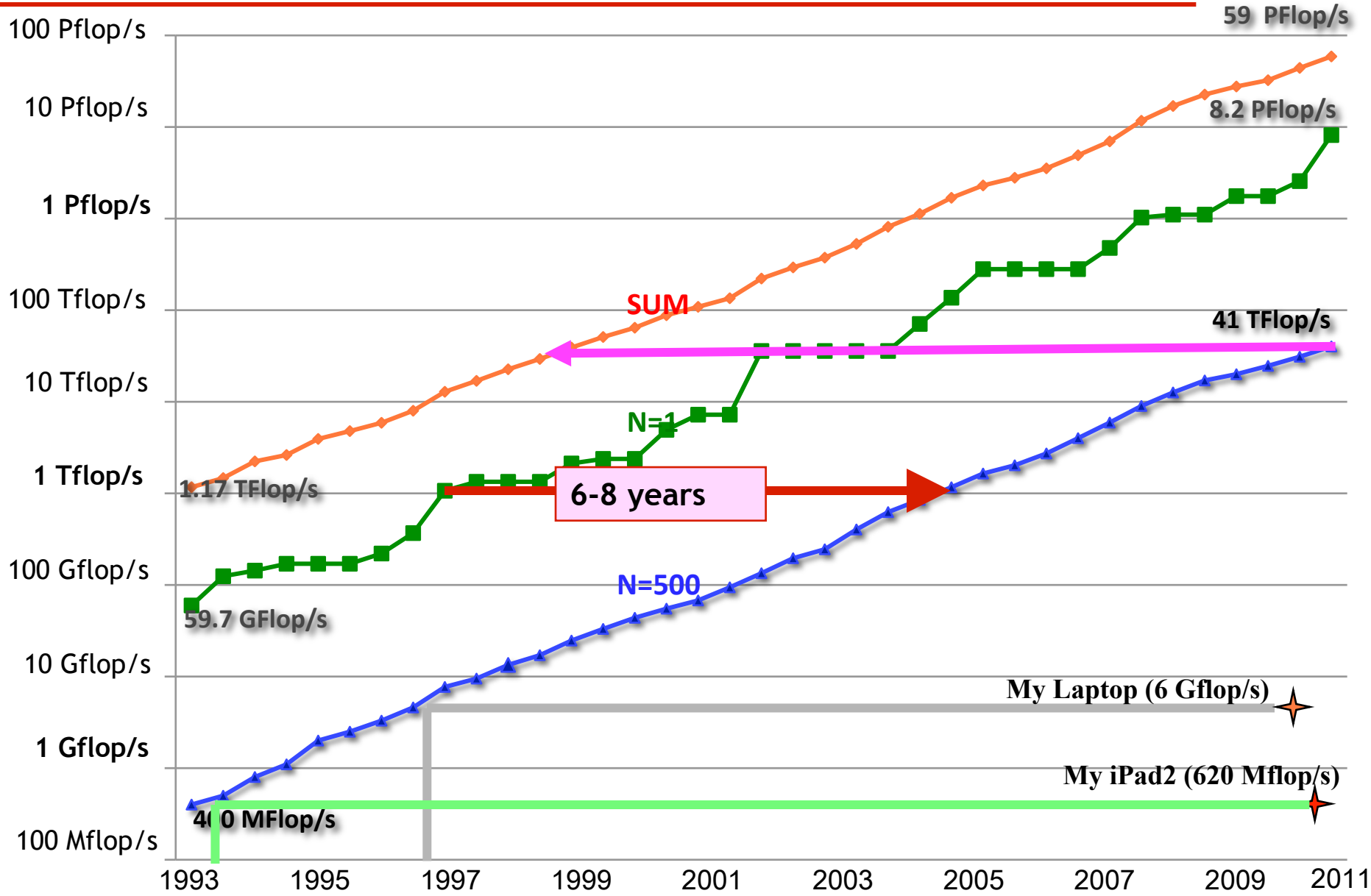
$Ax=b$, *dense problem*



- Updated twice a year
  SC‘xy in the States in November
  Meeting in Germany in June

- All data available from **www.top500.org**

Performance Development



Emerging Computer Architectures

- Are needed by applications

- Applications are given (as function of time)
- Architectures are given (as function of time)

- Algorithms and software must be adapted or created to bridge to computer architectures for the sake of the complex applications

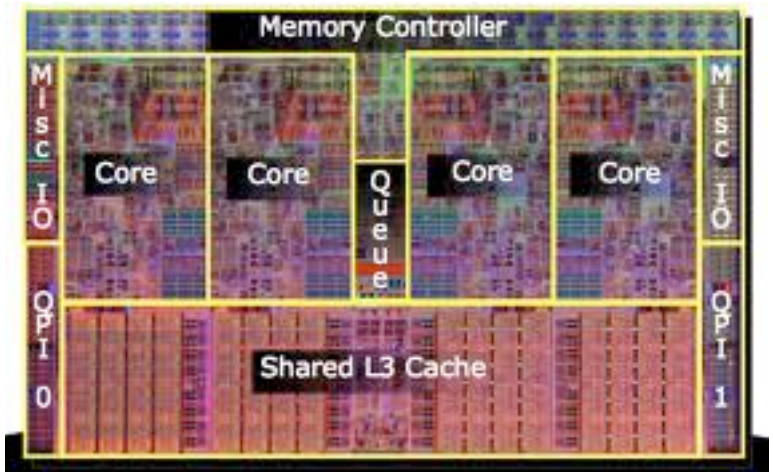
Three Design Points Today

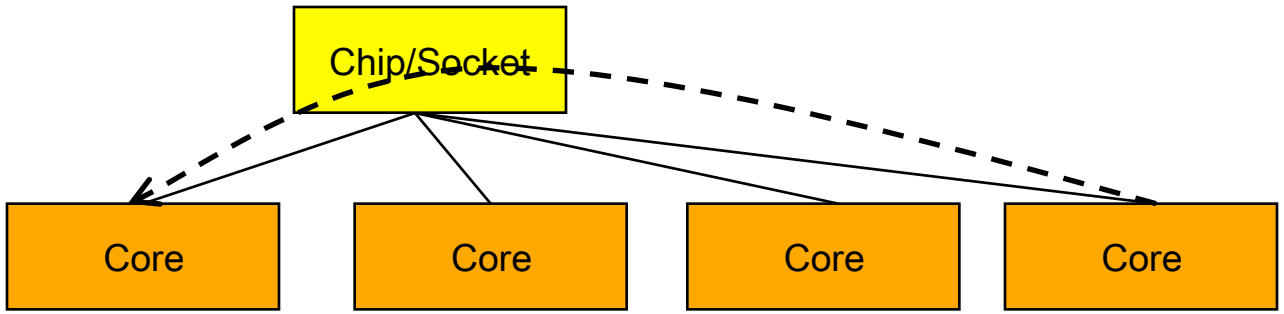
- Gigascale Laptop: Uninode-Multicore
  (Your iPhone and iPad are Mflop/s devices)

- Terascale Deskside: Multinode-Multicore

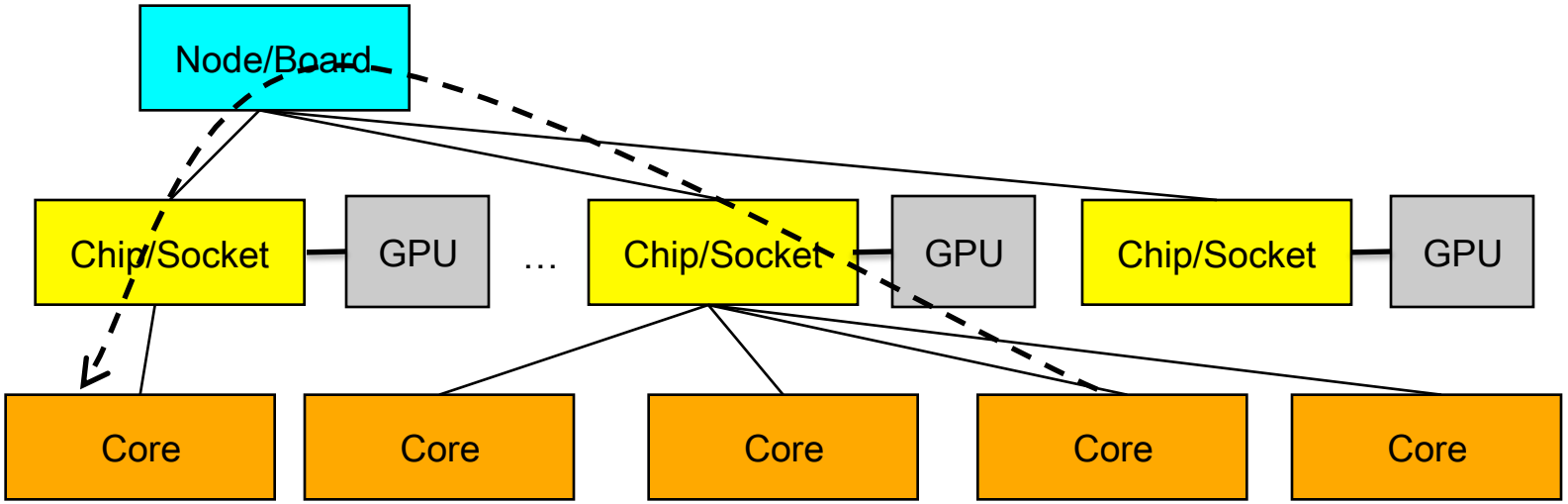
- Petacale Center: Multinode-Multicore

Example of typical parallel machine



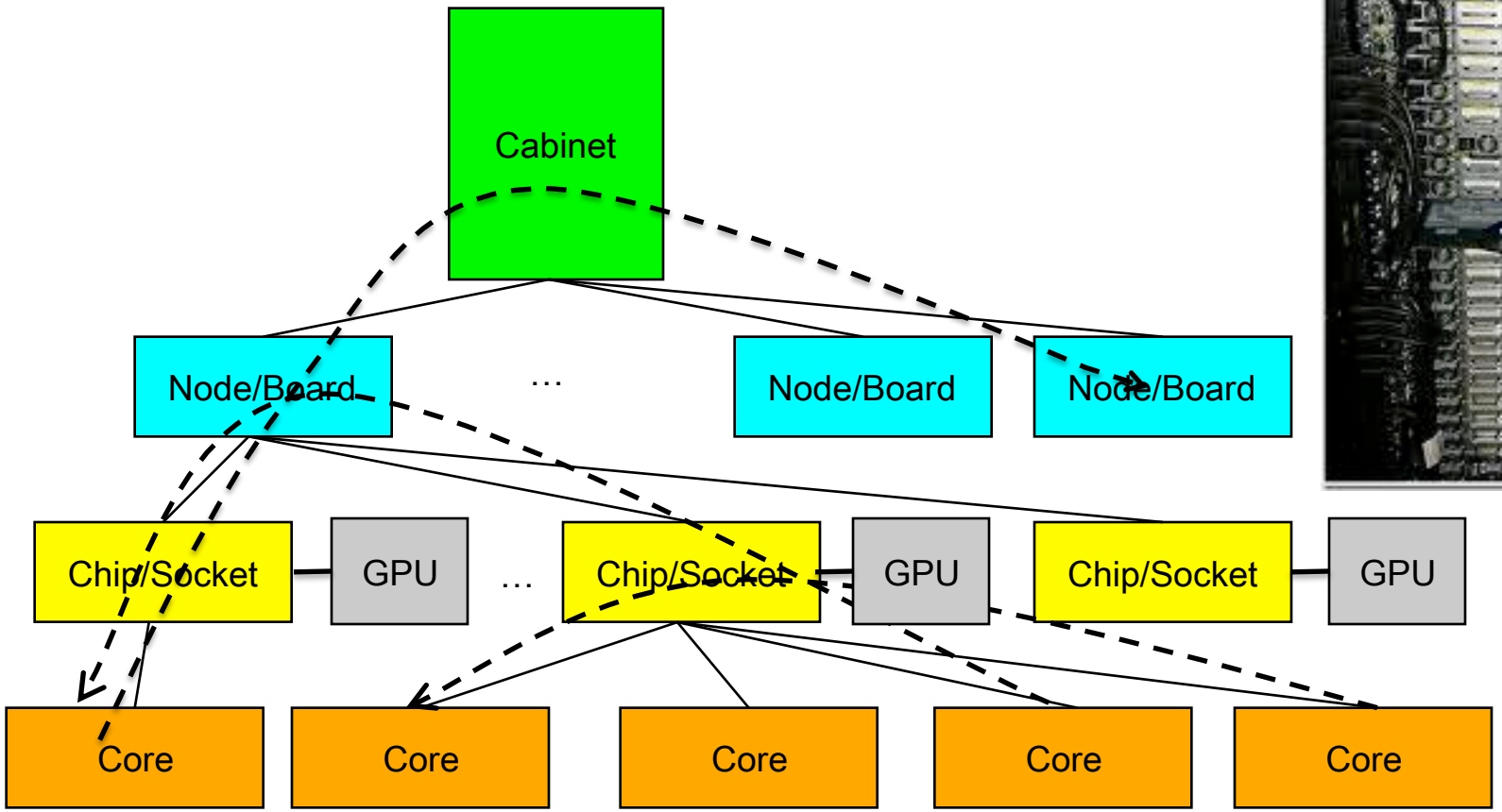


Example of typical parallel machine

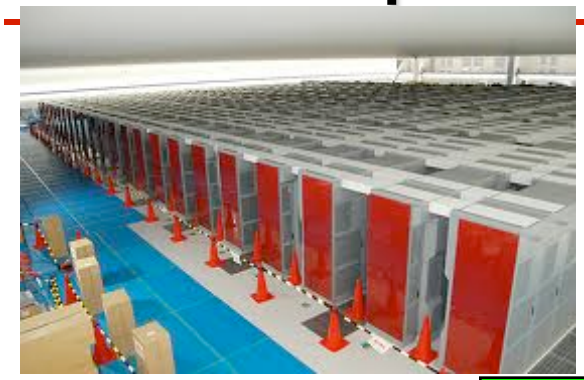


Example of typical parallel machine

Shared memory programming between processes on a board and
a combination of shared memory and distributed memory programming
between nodes and cabinets



Example of typical parallel machine



Combination of shared memory and distributed memory programming

June 2011: The TOP10

Rank	Site	Computer	Country	Cores	Rmax [Pflops]	% of Peak
1	RIKEN Advanced Inst for Comp Sci	K Computer Fujitsu SPARC64 VIIIfx + custom	Japan	548,352	8.16	93
2	Nat. SuperComputer Center in Tianjin	Tianhe-1A, NUDT Intel + Nvidia GPU + custom	China	186,368	2.57	55
3	DOE / OS Oak Ridge Nat Lab	Jaguar, Cray AMD + custom	USA	224,162	1.76	75
4	Nat. Supercomputer Center in Shenzhen	Nebulea, Dawning Intel + Nvidia GPU + IB	China	120,640	1.27	43
5	GSIC Center, Tokyo Institute of Technology	Tusbame 2.0, HP Intel + Nvidia GPU + IB	Japan	73,278	1.19	52
6	DOE / NNSA LANL & SNL	Cielo, Cray AMD + custom	USA	142,272	1.11	81
7	NASA Ames Research Center/NAS	Plelades SGI Altix ICE 8200EX/8400EX + IB	USA	111,104	1.09	83
8	DOE / OS Lawrence Berkeley Nat Lab	Hopper, Cray AMD + custom	USA	153,408	1.054	82
9	Commissariat a l'Energie Atomique (CEA)	Tera-10, Bull Intel + IB	France	138,368	1.050	84
10	DOE / NNSA Los Alamos Nat Lab	Roadrunner, IBM AMD + Cell GPU + IB	USA	122,400	1.04	76

June 2011: The TOP10

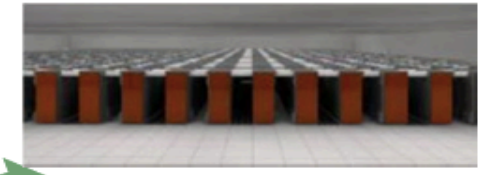
Rank	Site	Computer	Country	Cores	Rmax [Pflops]	% of Peak	Power [MW]	GFlops/ Watt
1	RIKEN Advanced Inst for Comp Sci	K Computer Fujitsu SPARC64 VIIIfx + custom	Japan	548,352	8.16	93	9.9	824
2	Nat. SuperComputer Center in Tianjin	Tianhe-1A, NUDT Intel + Nvidia GPU + custom	China	186,368	2.57	55	4.04	636
3	DOE / OS Oak Ridge Nat Lab	Jaguar, Cray AMD + custom	USA	224,162	1.76	75	7.0	251
4	Nat. Supercomputer Center in Shenzhen	Nebulae, Dawning Intel + Nvidia GPU + IB	China	120,640	1.27	43	2.58	493
5	GSIC Center, Tokyo Institute of Technology	Tusbame 2.0, HP Intel + Nvidia GPU + IB	Japan	73,278	1.19	52	1.40	850
6	DOE / NNSA LANL & SNL	Cielo, Cray AMD + custom	USA	142,272	1.11	81	3.98	279
7	NASA Ames Research Center/NAS	Plelades SGI Altix ICE 8200EX/8400EX + IB	USA	111,104	1.09	83	4.10	265
8	DOE / OS Lawrence Berkeley Nat Lab	Hopper, Cray AMD + custom	USA	153,408	1.054	82	2.91	362
9	Commissariat a l'Energie Atomique (CEA)	Tera-10, Bull Intel + IB	France	138,368	1.050	84	4.59	229
10	DOE / NNSA Los Alamos Nat Lab	Roadrunner, IBM AMD + Cell GPU + IB	USA	122,400	1.04	76	2.35	446
500	Energy Comp	IBM Cluster, Intel + GigE	China	7,104	.041	53		

Quiz: How Many of the Top500 systems use GPUs?

K computer Specifications

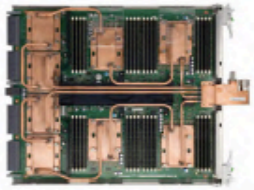
CPU (SPARC64 VIIIfx)	Cores/Node	8 cores (@2GHz)
	Performance	128GFlops
	Architecture	SPARC V9 + HPC extension
	Cache	L1(I/D) Cache : 32KB/32KB L2 Cache : 6MB
	Power	58W (typ. 30 C)
	Mem. bandwidth	64GB/s.
Node	Configuration	1 CPU / Node
	Memory capacity	16GB (2GB/core)
System board(SB)	No. of nodes	4 nodes /SB
Rack	No. of SB	24 SBs/rack
System	Nodes/system	> 80,000

Inter-connect	Topology	6D Mesh/Torus
	Performance	5GB/s. for each link
	No. of link	10 links/ node
	Additional feature	H/W barrier, reduction
	Architecture	Routing chip structure (no outside switch box)
Cooling	CPU, ICC*	Direct water cooling
	Other parts	Air cooling



System
LINPACK 10 PFlops
over 1PB mem.
800 racks
80,000 CPUs
640,000 cores

Rack
12.3 TFlops
15TB memory



System Board
512 GFlops
64 GB memory

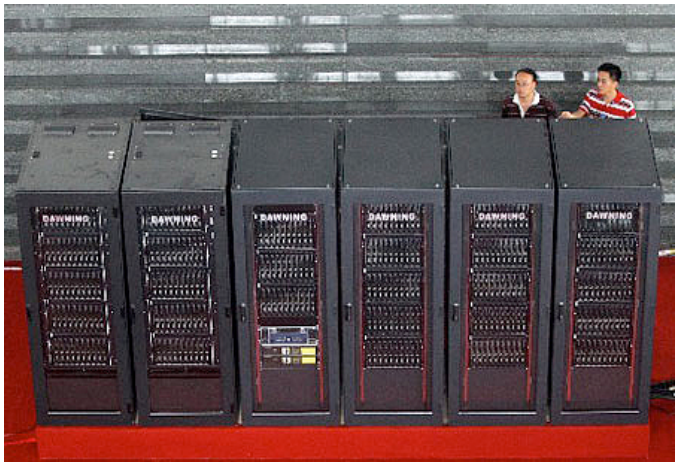
Node
128 GFlops
16GB Memory
64GB/s Memory band width

CPU
128GFlops
SPARC64™ VIIIfx
8 Cores@2.0GHz

* ICC : Interconnect Chip

China's Very Aggressive Deployment of HPC

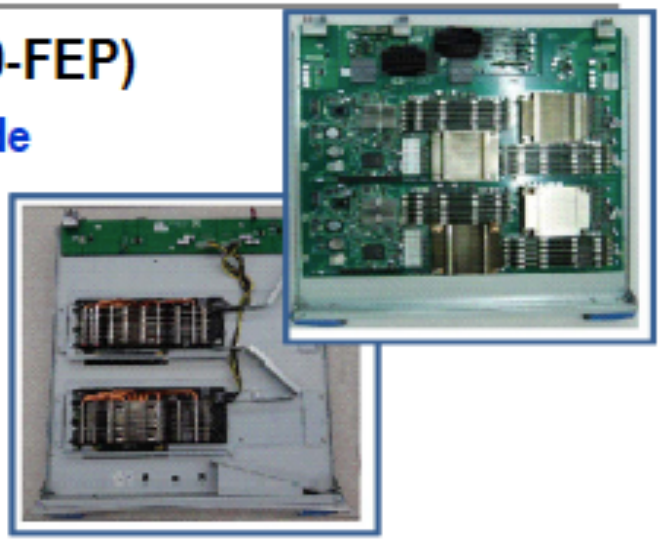
- Has 3 Pflops systems
  - NUDT, Tianhe-1A, located in Tianjin
    Dual-Intel 6 core + Nvidia Fermi w /custom interconnect
    - Budget 600M RMB
      - MOST 200M RMB, Tianjin Government 400M RMB
  - CIT, Dawning 6000, Nebulea, located in Shenzhen
    Dual-Intel 6 core + Nvidia Fermi w /QDR Ifiniband
    - Budget 600M RMB
      - MOST 200M RMB, Shenzhen Government 400M RMB
  - Mole-8.5 Cluster/320x2 Intel QC Xeon E5520 2.26 Ghz + 320x6 Nvidia Tesla C2050/QDR Infiniband
- Fourth one planned for Shandong



Tianhe-1A

Main configuration of TH-1A system

- 7,168 compute nodes (YH-X5670-FEP)
  - 2 six-core CPU and 1 GPU per node
  - CPU: Xeon X5670 (Westmere)
    - Processor speed - 2.93GHz
  - GPU: nVIDIA M2050
    - Connected with CPU by PCI-E
  - 32GB memory per node
  - 2U height



7168(nodes) × 2(CPU) ×2.93(GHz)× 6(Cores) ×4 =1.008PFlops

7168(nodes) × 1(GPU) ×1.15(GHz)*448(CUDA Cores) =3.692PFlops

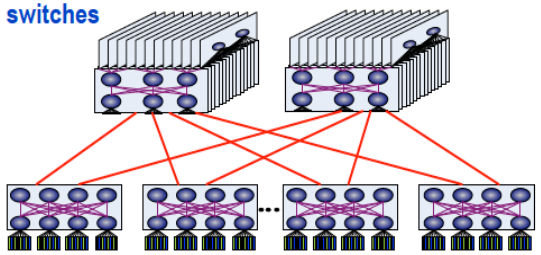
+

Total: 4,701,061 GFlops

Tianhe-1A

- The interconnect on the Tianhe-1A is a proprietary fat-tree.
- The router and network interface chips where designed by NUDT.
- It has a bi-directional bandwidth of 160 Gb/s, double that of QDR infiniband, a latency for a node hop of 1.57 microseconds, and an aggregated bandwidth of 61 Tb /sec.
- On the MPI level, the bandwidth and latency is 6.3GBps(one direction)/9.3 GBps(bi- direction) and 2.32us, respectively.

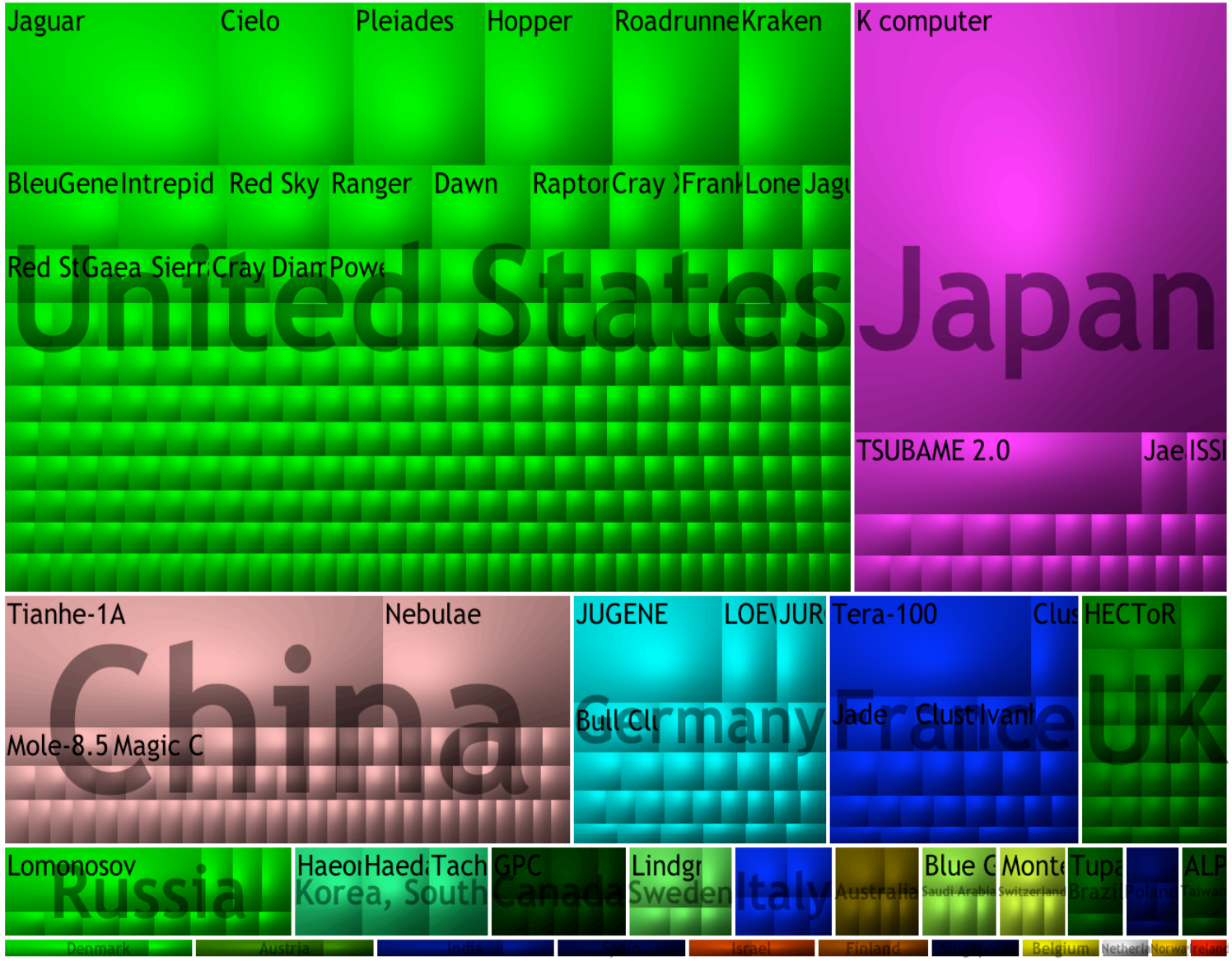
- First stage: 16 nodes connected by 16-port switching board
- Second stage: all parts connected to eleven 384-port switches



- High radix router ASIC：NRC
  - Feature size ： 90nm
  - Die size：17.16mm x 17.16mm
  - Package ： FC-PBGA
  - 2577 pins
  - Throughput of single NRC: 2.56Tbps

- Network interface ASIC：NIC
  - Same Feature size and package
  - Die size ： 10.76mm x 10.76mm
  - 675 pins, PCI-E G2 16X

Countries Share



Absolute Counts

US:	251
China:	64
Germany:	31
UK:	28
Japan:	26
France:	25

28 Supercomputers in the UK

Rank	Site	Computer	Cores	Rmax Tflop/s
24	University of Edinburgh	Cray XE6 12-core 2.1 GHz	44376	279
65	Atomic Weapons Establishment	Bullx B500 Cluster, Xeon X56xx 2.8Ghz, QDR Infiniband	12936	124
69	ECMWF	Power 575, p6 4.7 GHz, Infiniband	8320	115
70	ECMWF	Power 575, p6 4.7 GHz, Infiniband	8320	115
93	University of Edinburgh	Cray XT4, 2.3 GHz	12288	95
154	University of Southampton	iDataPlex, Xeon QC 2.26 GHz, Ifband, Windows HPC2008 R2	8000	66
160	IT Service Provider	Cluster Platform 4000 BL685c G7, Opteron 12C 2.2 Ghz, GigE	14556	65
186	IT Service Provider	Cluster Platform 3000 BL460c G7, Xeon X5670 2.93 Ghz, GigE	9768	59
190	Computacenter (UK) LTD	Cluster Platform 3000 BL460c G1, Xeon L5420 2.5 GHz, GigE	11280	58
191	Classified	xSeries x3650 Cluster Xeon QC GT 2.66 GHz, Infiniband	6368	58
211	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
212	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
213	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
228	IT Service Provider	Cluster Platform 4000 BL685c G7, Opteron 12C 2.1 Ghz, GigE	12552	54
233	Financial Institution	iDataPlex, Xeon X56xx 6C 2.66 GHz, GigE	9480	53
234	Financial Institution	iDataPlex, Xeon X56xx 6C 2.66 GHz, GigE	9480	53
278	UK Meteorological Office	Power 575, p6 4.7 GHz, Infiniband	3520	51
279	UK Meteorological Office	Power 575, p6 4.7 GHz, Infiniband	3520	51
339	Computacenter (UK) LTD	Cluster Platform 3000 BL460c, Xeon 54xx 3.0GHz, GigEthernet	7560	47
351	Asda Stores	BladeCenter HS22 Cluster, WM Xeon 6-core 2.93Ghz, GigE	8352	47
365	Financial Services	xSeries x3650M2 Cluster, Xeon QC E55xx 2.53 Ghz, GigE	8096	46
404	Financial Institution	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	7872	44
405	Financial Institution	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	7872	44
415	Bank	xSeries x3650M3, Xeon X56xx 2.93 GHz, GigE	7728	43
416	Bank	xSeries x3650M3, Xeon X56xx 2.93 GHz, GigE	7728	43
482	IT Service Provider	Cluster Platform 3000 BL460c G6, Xeon L5520 2.26 GHz, GigE	8568	40
484	IT Service Provider	Cluster Platform 3000 BL460c G6, Xeon X5670 2.93 GHz, 10G	4392	40

28 Supercomputers in the UK

Rank	Site	Computer	Cores	Rmax Tflop/s
24	University of Edinburgh	Cray XE6 12-core 2.1 GHz	44376	279
65	Atomic Weapons Establishment	Bullx B500 Cluster, Xeon X56xx 2.8Ghz, QDR Infiniband	12936	124
69	ECMWF	Power 575, p6 4.7 GHz, Infiniband	8320	115
70	ECMWF	Power 575, p6 4.7 GHz, Infiniband	8320	115
93	University of Edinburgh	Cray XT4, 2.3 GHz	12288	95
154	University of Southampton	iDataPlex, Xeon QC 2.26 GHz, Ifband, Windows HPC2008 R2	8000	66
160	IT Service Provider	Cluster Platform 4000 BL685c G7, Opteron 12C 2.2 Ghz, GigE	14556	65
186	IT Service Provider	Cluster Platform 3000 BL460c G7, Xeon X5670 2.93 Ghz, GigE	9768	59
190	Computacenter (UK) LTD	Cluster Platform 3000 BL460c G1, Xeon L5420 2.5 GHz, GigE	11280	58
191	Classified	xSeries x3650 Cluster Xeon QC GT 2.66 GHz, Infiniband	6368	58
211	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
212	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
213	Classified	BladeCenter HS22 Cluster, WM Xeon 6-core 2.66Ghz, Ifband	5880	55
228	IT Service Provider	Cluster Platform 4000 BL685c G7, Opteron 12C 2.1 Ghz, GigE	12552	54
233	Financial Institution	iDataPlex, Xeon X56xx 6C 2.66 GHz, GigE	9480	53
234	Financial Institution	iDataPlex, Xeon X56xx 6C 2.66 GHz, GigE	9480	53
278	UK Meteorological Office	Power 575, p6 4.7 GHz, Infiniband	3520	51
279	UK Meteorological Office	Power 575, p6 4.7 GHz, Infiniband	3520	51
339	Computacenter (UK) LTD	Cluster Platform 3000 BL460c, Xeon 54xx 3.0GHz, GigEthernet	7560	47
351	Asda Stores	BladeCenter HS22 Cluster, WM Xeon 6-core 2.93Ghz, GigE	8352	47
365	Financial Services	xSeries x3650M2 Cluster, Xeon QC E55xx 2.53 Ghz, GigE	8096	46
404	Financial Institution	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	7872	44
405	Financial Institution	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	7872	44
415	Bank	xSeries x3650M3, Xeon X56xx 2.93 GHz, GigE	7728	43
416	Bank	xSeries x3650M3, Xeon X56xx 2.93 GHz, GigE	7728	43
482	IT Service Provider	Cluster Platform 3000 BL460c G6, Xeon L5520 2.26 GHz, GigE	8568	40
484	IT Service Provider	Cluster Platform 3000 BL460c G6, Xeon X5670 2.93 GHz, 10G	4392	40

Commodity plus Accelerator

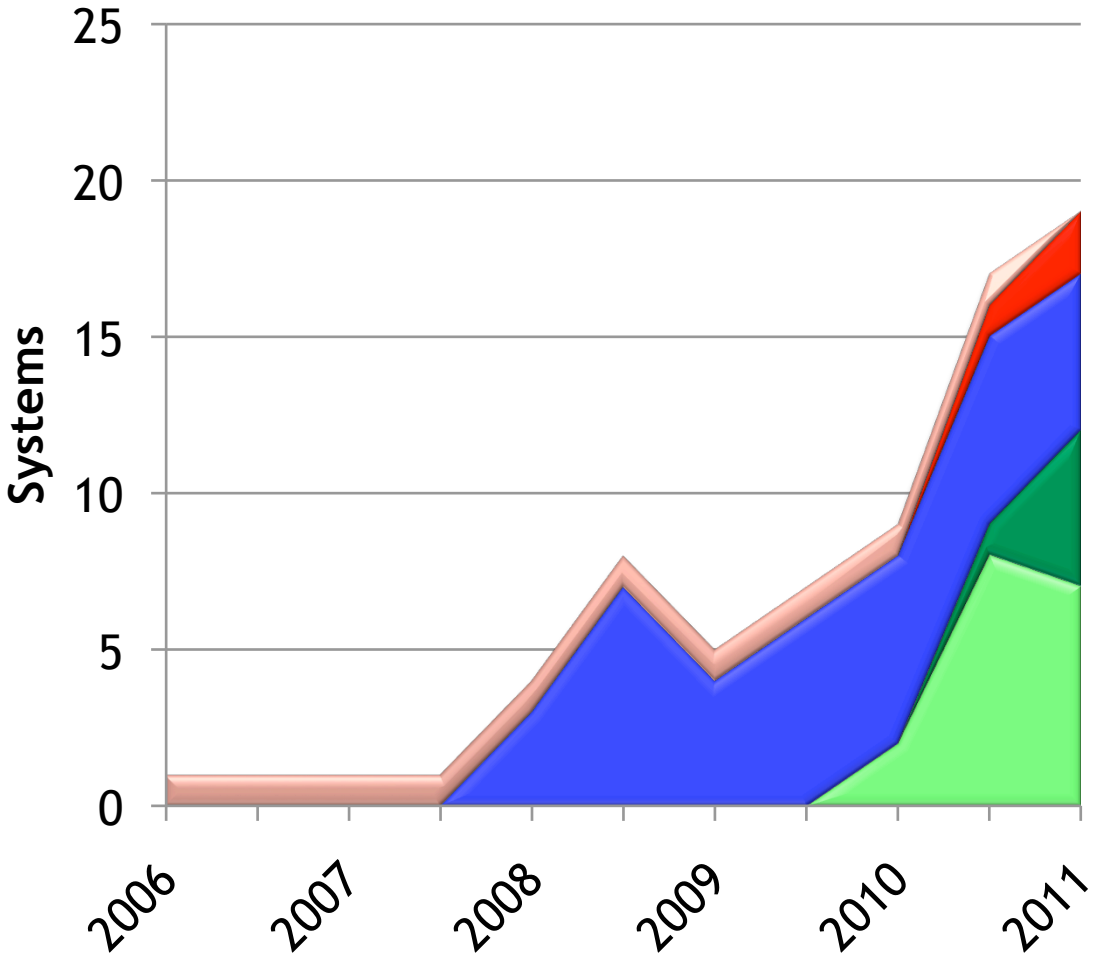
Commodity	Accelerator (GPU)
Intel Xeon	Nvidia C2050 "Fermi"
3 GHz	1.15 GHz
8*4 ops/cycle	448 ops/cycle
96 Gflop/s (DP)	515 Gflop/s (DP)

Quiz: How Many of the Top500 systems use GPUs?

Answer:
Today only 19 systems on the TOP500 use GPUs

Interconnect
PCI-X 16 lane
64 Gb/s
1 GW/s

Accelerators



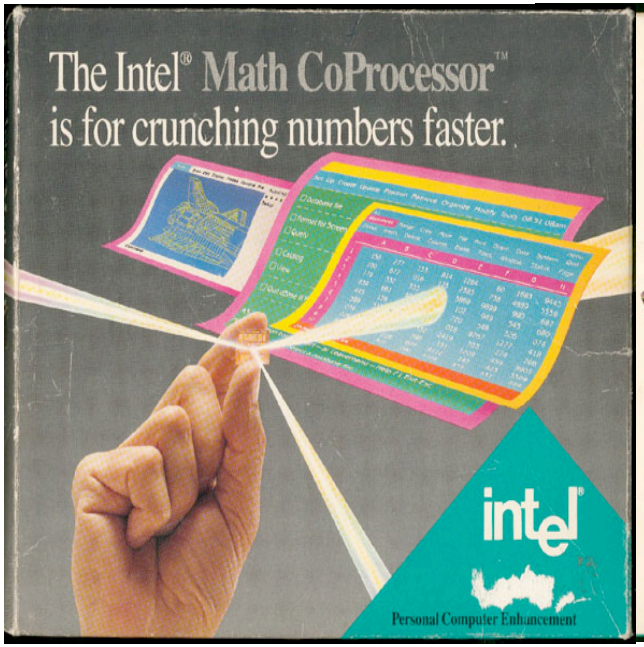
- Clearspeed CSX60022
- ATI GPU
- IBM PowerXCell 8i
- NVIDIA 2070
- NVIDIA 2050

6 US
4 Germany
3 China
3 Japan
1 Australia
1 Italy
1 Russia

Rank	Site	Manufacturer	Computer	Country	Cores	RMax	RPeak	%	Accelerator	Interconnect Family
2	National Supercomputing Center in Tianjin	NUDT	NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C	China	186368	2566000	4701000	0.55	NVIDIA 2050	Custom
4	National Supercomputing Centre in Shenzhen (NSCS)	Dawning	Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU	China	120640	1271000	2984300	0.43	NVIDIA 2050	Infiniband
5	GSIC Center, Tokyo Institute of Technology	NEC/HP	HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows	Japan	73278	1192000	2287630	0.52	NVIDIA 2050	Infiniband
10	DOE/NNSA/LANL	IBM	BladeCenter QS22/LS21 PowerXCell 8i 3.2 Ghz / Opteron 1.8 GHz, Voltaire Infiniband	United States	122400	1042000	1375780	0.76	IBM PowerXCell 8i	Infiniband
13	Moscow State University - Research Computing Center	T-Platforms	T-Platforms T-Blade2/1.1, Xeon X5570/X5670 2.93 GHz, Nvidia 2070 GPU, Infiniband QDR	Russia	33072	674105	1373060	0.49	NVIDIA 2070	Infiniband
22	Universitaet Frankfurt	Clustervision/ Supermicro	Supermicro Cluster, QC Opteron 2.1 GHz, ATI Radeon GPU, Infiniband	Germany	16368	299300	508499	0.59	ATI GPU	Infiniband
33	Institute of Process Engineering, Chinese Academy of Sci	IPE, Nvidia, Tyan	Mole-8.5 Cluster Xeon L5520 2.26 Ghz, nVidia Tesla, Infiniband	China	33120	207300	1138440	0.18	NVIDIA 2050	Infiniband
54	CINECA / SCS - SuperComputing Solution	IBM	iDataPlex DX360M3, Xeon 2.4, nVidia GPU, Infiniband	Italy	3072	142700	293274	0.49	NVIDIA 2070	Infiniband
60	DOE/NNSA/LANL	IBM	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Infiniband	United States	14400	126500	161856	0.78	IBM PowerXCell 8i	Infiniband
85	Lawrence Livermore National Laboratory	Appro International	Appro GreenBlade Cluster Xeon X5660 2.8Ghz, nVIDIA M2050, Infiniband	United States	8240	100500	239866	0.42	NVIDIA 2050	Infiniband
126	National Institute for Environmental Studies	NSSOL / SGI Japan	Asterism ID318, Intel Xeon E5530, NVIDIA C2050, Infiniband	Japan	5760	75350	177120	0.43	NVIDIA 2050	Infiniband
148	University of California, Los Angeles	Hewlett-Packard	HP ProLiant SL390s G7 Xeon X5650, Nvidia M2070, Infiniband QDR	United States	2482	68100	160577	0.42	NVIDIA 2070	Infiniband
169	Georgia Institute of Technology	Hewlett-Packard	HP ProLiant SL390s G7 Xeon 6C X5660 2.8Ghz, nVidia Fermi, Infiniband QDR	United States	6048	63920	188092	0.34	NVIDIA 2070	Infiniband
273	CSIRO	Xenon Systems	Supermicro Xeon Cluster, E5462 2.8 Ghz, Nvidia Tesla s2050 GPU, Infiniband	Australia	4608	52550	143300	0.37	NVIDIA 2050	Infiniband
388	Hewlett-Packard	Hewlett-Packard	HP ProLiant SL390s G7 Xeon X5650, Nvidia M2070, Infiniband QDR	United States	1352	45316.2	86979.4	0.52	NVIDIA 2070	Infiniband
406	Forschungszentrum Juelich (FZJ)	IBM	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	Germany	4608	44500	55705.6	0.80	IBM PowerXCell 8i	Custom
407	Universitaet Regensburg	IBM	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	Germany	4608	44500	55705.6	0.80	IBM PowerXCell 8i	Custom
408	Universitaet Wuppertal	IBM	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	Germany	4608	44500	55705.6	0.80	IBM PowerXCell 8i	Custom
429	Nagasaki University	Self-made	DEGIMA Cluster, Intel i5, ATI Radeon GPU, Infiniband QDR	Japan	7920	42830	111150	0.39	ATI GPU	Infiniband

We Have Seen This Before

- Floating Point Systems FPS-164/MAX Supercomputer (1976)
- Intel Math Co-processor (1980)
- Weitek Math Co-processor (1981)

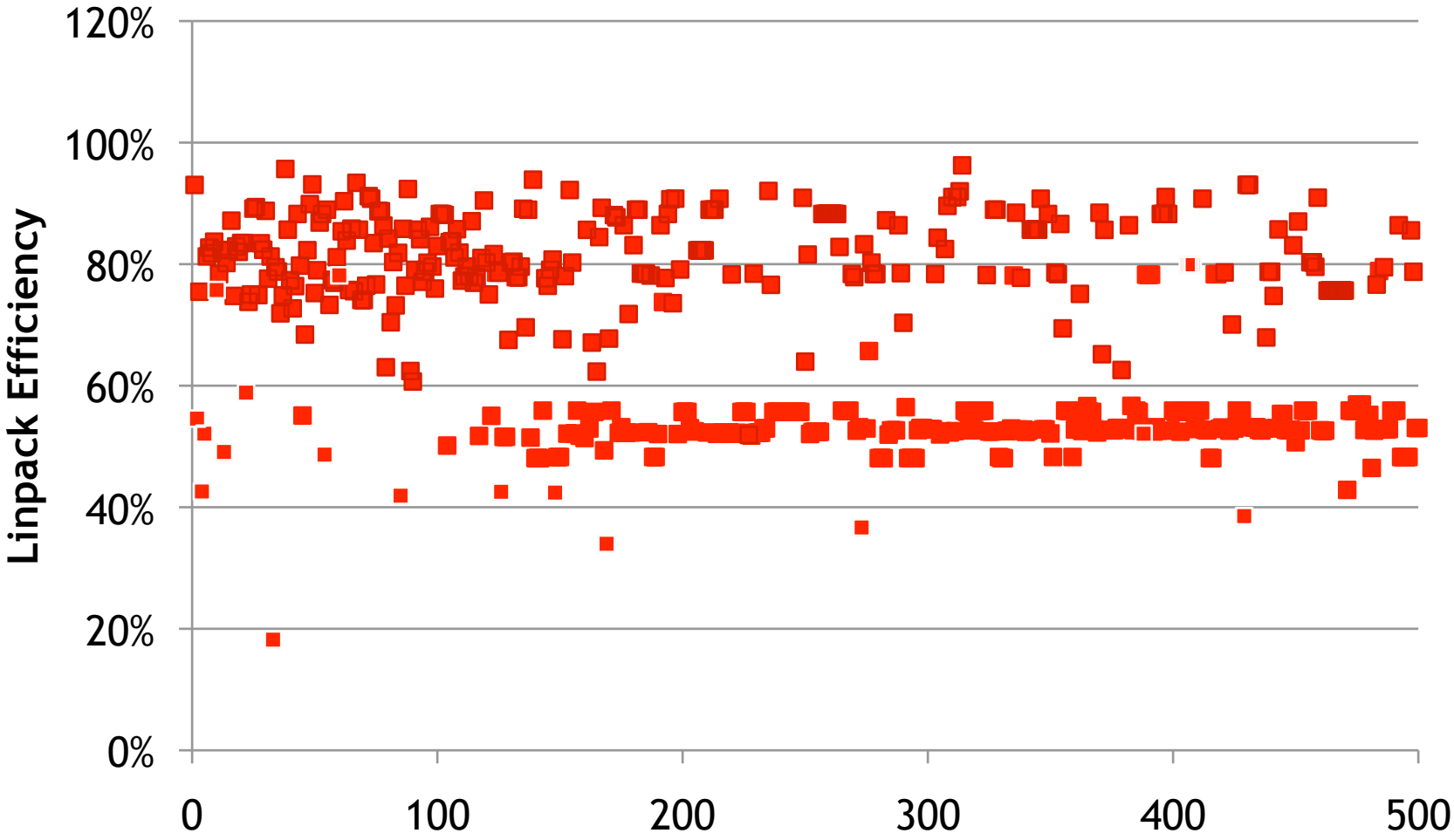


Balance Between Data Movement and Floating point

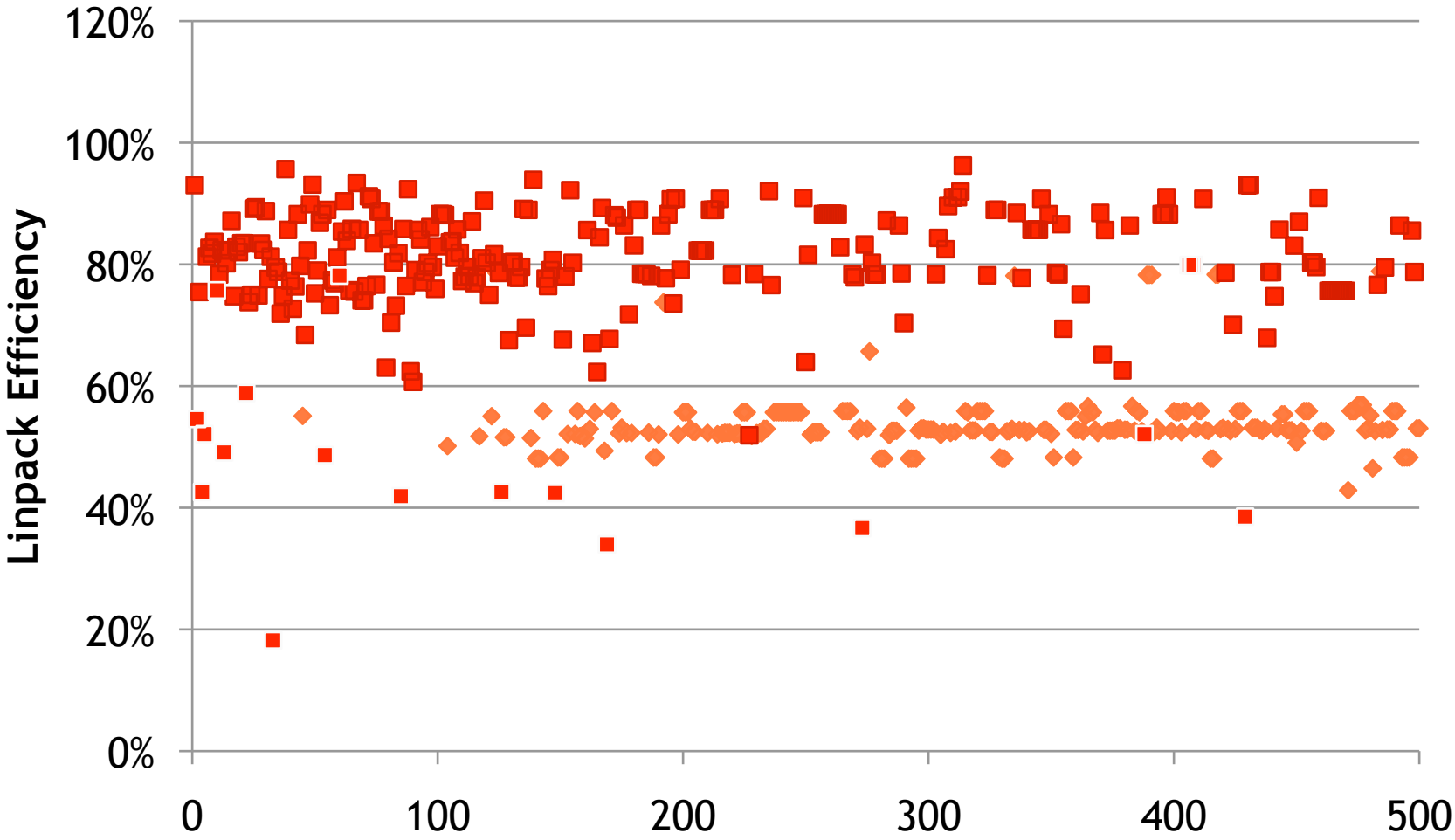
- FPS-164 and VAX (1976)
  - 11 Mflop/s; transfer rate 44 MB/s
  - Ratio of flops to bytes of data movement:
    1 flop per 4 bytes transferred
- Nvidia Fermi and PCI-X to host
  - 500 Gflop/s; transfer rate 8 GB/s
  - Ratio of flops to bytes of data movement:
    62 flops per 1 byte transferred

- Flop/s are cheap, so are provisioned in excess

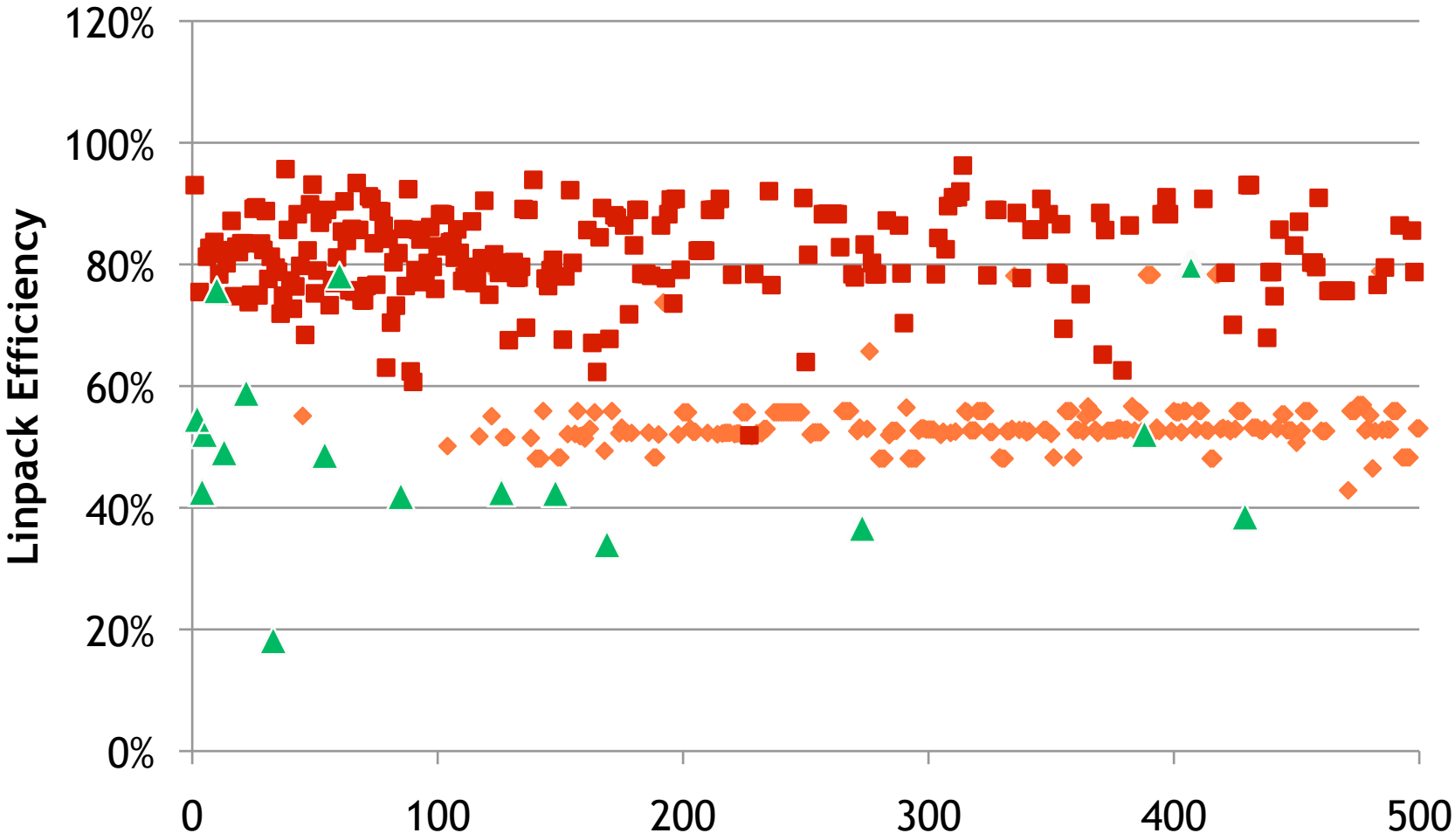
Linpack Efficiency



Linpack Efficiency

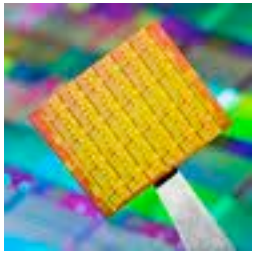


Linpack Efficiency



Future Computer Systems

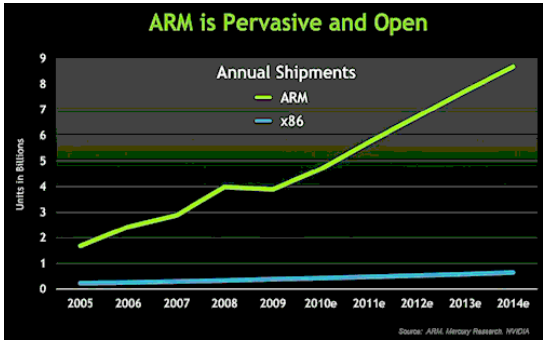
- Most likely be a hybrid design
  - Think standard multicore chips and accelerator (GPUs)
- Today accelerators are attached
- Next generation more integrated
- Intel's MIC architecture "Knights Ferry" and "Knights Corner" to come.
  - 48 x86 cores



- AMD's Fusion in 2012 - 2013
  - Multicore with embedded graphics ATI

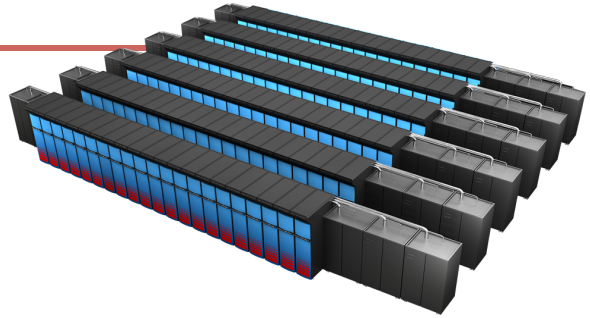
AMD
The future is fusion

- Nvidia's Project Denver plans to develop an integrated chip using ARM architecture in 2013.



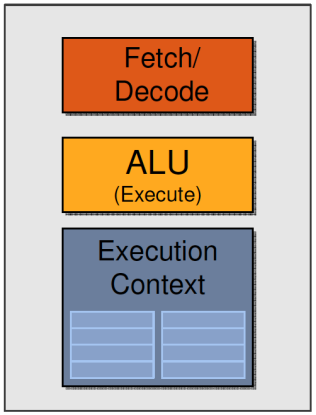
10+ Pflop/s Systems Planned in the States

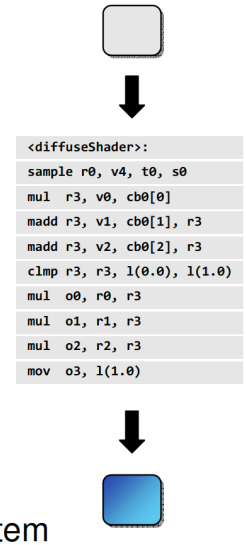
- DOE Funded, Titan at ORNL, Based on Cray design with accelerators, 20 Pflop/s, 2012
- DOE Funded, Sequoia at Lawrence Livermore Nat. Lab, Based on IBM's BG/Q, 20 Pflop/s, 2012
- DOE Funded, BG/Q at Argonne National Lab, Based on IBM's BG/Q, 10 Pflop/s, 2012
- NSF Funded, Blue Waters at University of Illinois UC, Based on IBM's Power 7 Proc, 10 Pflop/s, 2012



How to Count Cores?

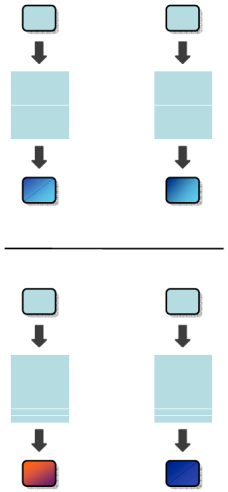
- **CPU Conventional Core**

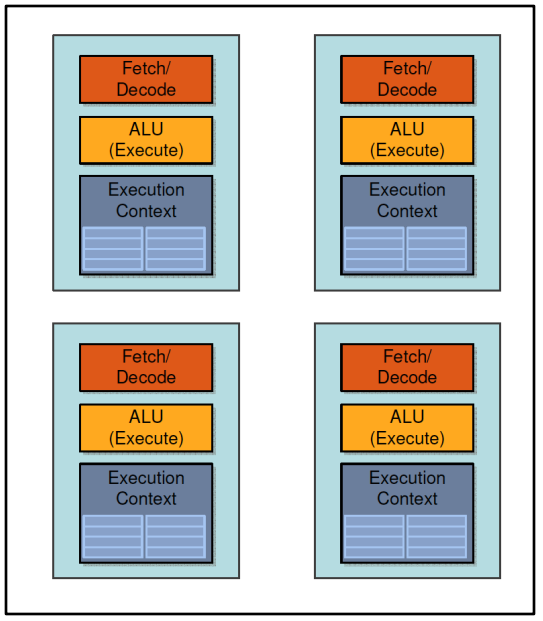




One instruction stream per work-item

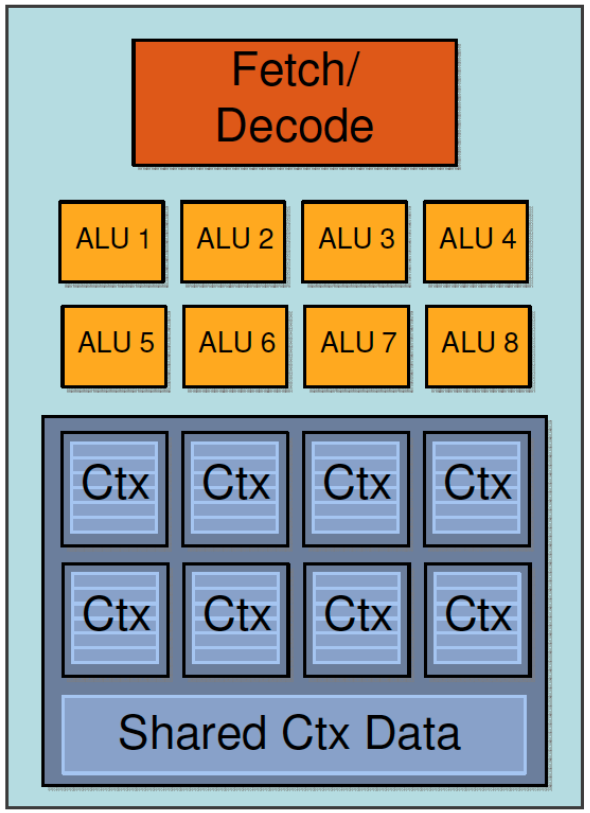
Quad



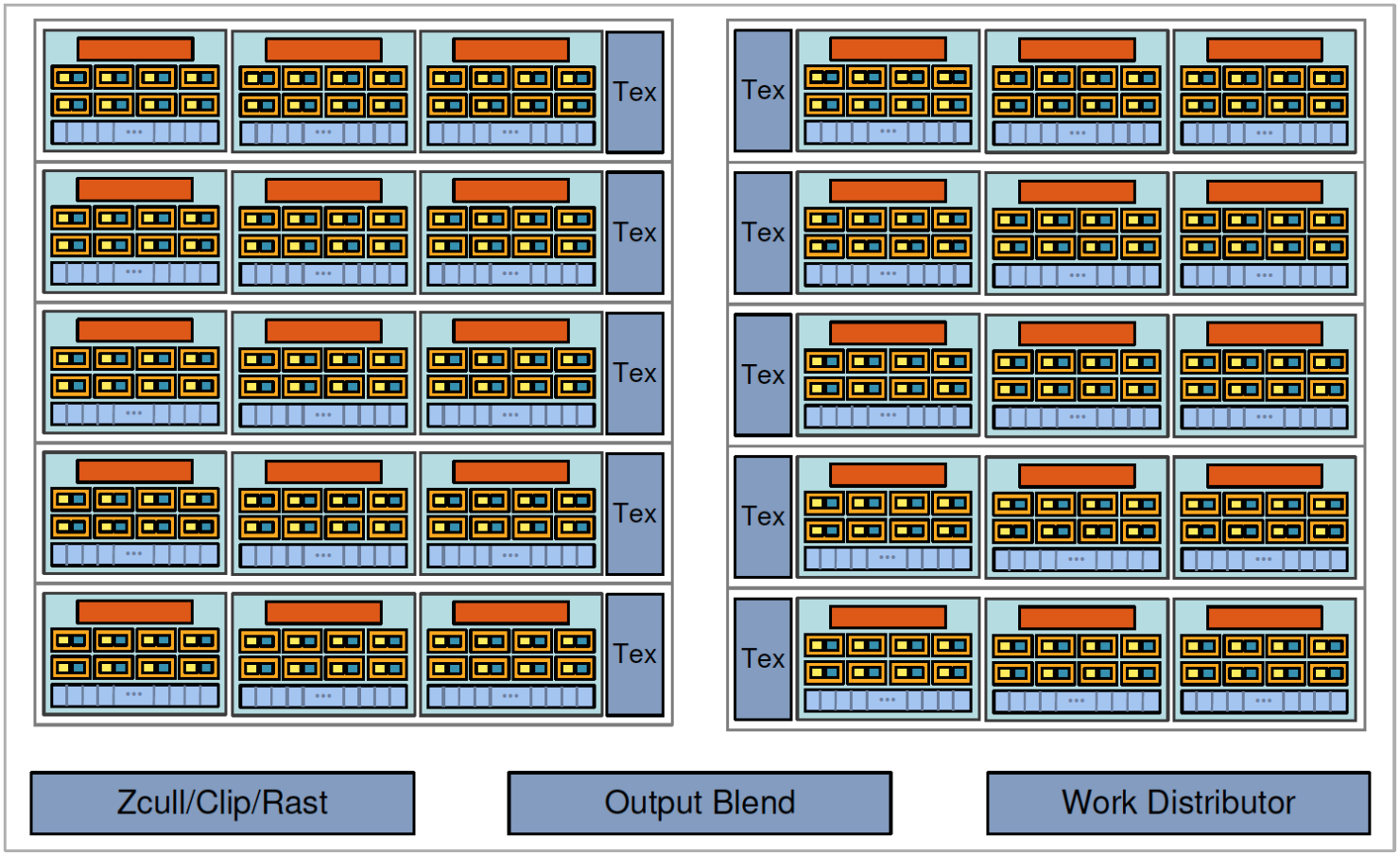


In GPUs - Add ALUs

- SIMD Processing
- Amortize cost /complexity of managing an instruction stream across many ALUs.
- NVIDIA refers to these ALUs as “CUDA Cores” (also streaming processors)

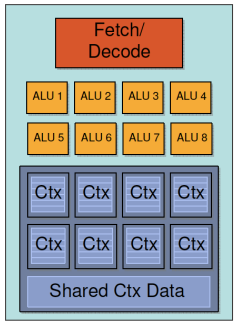


NVIDIA GT280 “old Telsa”



- Equivalent to 30 processing cores, each with 8 “CUDA cores”
- 240 streaming processors (CUDA Cores) (ALUs)

NVIDIA GeForce GTX 280 (Tesla)

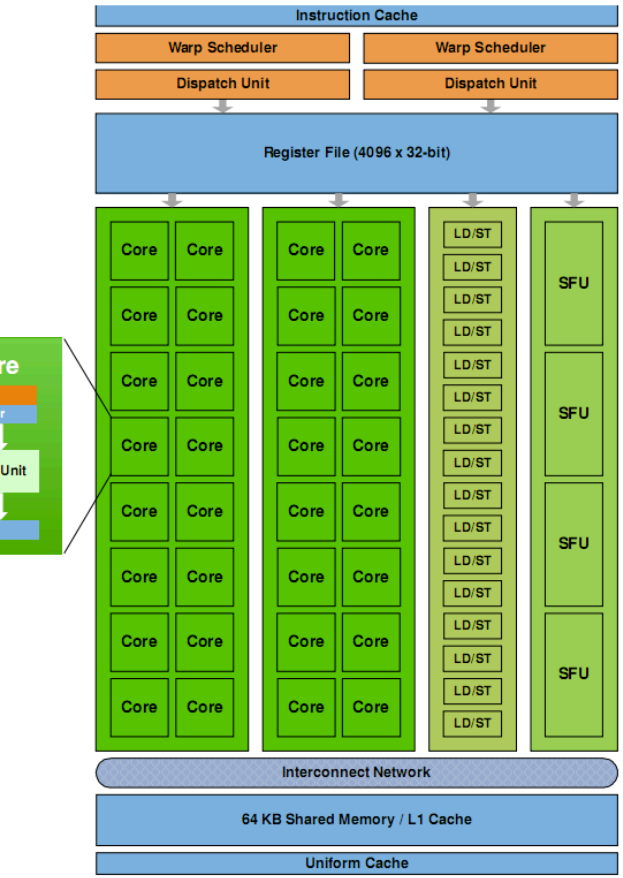


Processing Core

- NVIDIA-Speak
  - 240 CUDA cores (ALUs)
- Generic speak
  - 30 processing cores
    - 8 CUDA Cores (SIMD functional units) per core
  - 1 mul-add (2 flops) + 1 mul per functional unit (3 flops/cycle)
  - Best case theoretically: 240 mul-adds + 240 muls per cycle
    - 1.3 GHz clock
    - 30 * 8 * (2 + 1) * 1.33 = 933 Gflop/s peak
  - Best case reality: 240 mul-adds per clock
    - Just able to do the mul-add so 2/3 or 624 Gflop/s
  - All this is single precision
    - Double precision is 78 Gflop/s peak (Factor of 8 from SP; exploit mixed prec)
  - 141 GB/s bus, 1 GB memory
  - 4 GB/s via PCIe (we see: T = 11 us + Bytes/3.3 GB/s)
  - In SP SGEMM performance 375 Gflop/s

NVIDIA Fermi

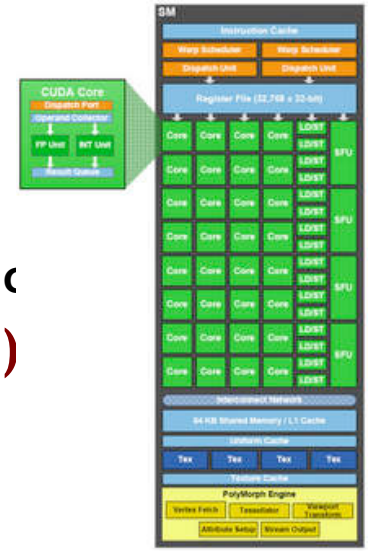
- Fermi GTX 480 has 448 CUDA cores (ALUs)
- 32 CUDA Cores (ALUs) in each of the 14 processing Cores

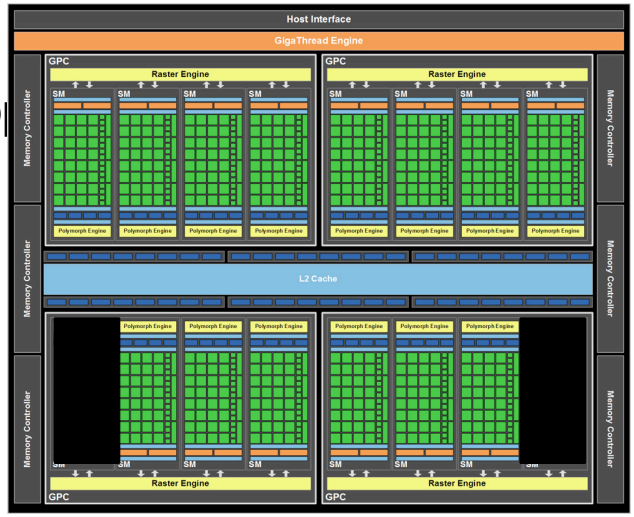


NVIDIA Tesla C2070 (Fermi), GF100 Chip

- **NVIDIA-Speak**
  - 448 CUDA cores (ALUs)
- **Generic speak**
  - 14 processing cores
    - 32 CUDA Cores (SIMD functional units) per c
  - 1 mul-add (2 flops) per ALU (2 flops/cycle)
  - Best case theoretically: 448 mul-adds
    - 1.15 GHz clock
    - 14 * 32 * 2 * 1.15 = 1.03 Tflop/s peak
  - All this is single precision
    - Double precision is half this rate, 515 Gflo
  - In SP SGEMM performance 635Gflop/s
  - In DP DGEMM performance 305 Gflop/s
  - Interface PCI-x16

Processing Core





Potential System Architecture

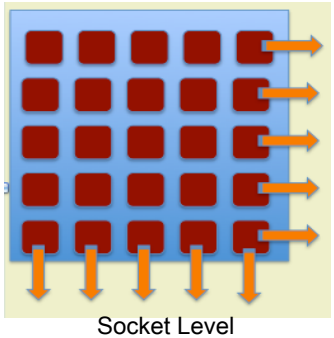
Systems	2011 K Computer
System peak	**8.7 Pflop/s**
Power	**10 MW**
System memory	1.6 PB
Node performance	128 GF
Node memory BW	64 GB/s
Node concurrency	8
Total Node Interconnect BW	20 GB/s
System size (nodes)	68,544
Total concurrency	548.352
MTTI	days

Potential System Architecture with a cap of $200M and 20MW

Systems	2011 K Computer	2019	Difference Today & 2019
System peak	**8.7 Pflop/s**	**1 Eflop/s**	**O(100)**
Power	**10 MW**	**~20 MW**	
System memory	1.6 PB	32 - 64 PB	O(10)
Node performance	128 GF	1,2 or 15TF	O(10) – O(100)
Node memory BW	64 GB/s	2 - 4TB/s	O(100)
Node concurrency	8	O(1k) or 10k	O(100) – O(1000)
Total Node Interconnect BW	20 GB/s	200-400GB/s	O(10)
System size (nodes)	68,544	O(100,000) or O(1M)	O(10) – O(100)
Total concurrency	548.352	O(billion)	O(1,000)
MTTI	days	O(1 day)	- O(10)

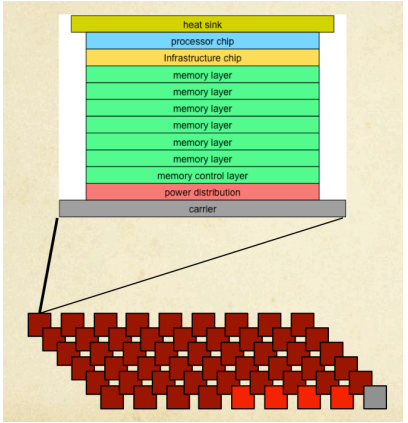
Exascale (10^{18} Flop/s) Systems: Two Possible Swim Lanes

- **Light weight processors (think BG/P)**
  - ~1 GHz processor (10^9)
  - ~1 Kilo cores/socket (10^3)
  - ~1 Mega sockets/system (10^6)



Socket Level
Cores scale-out for planar geometry

- **Hybrid system (think GPU based)**
  - ~1 GHz processor (10^9)
  - ~10 Kilo FPUs/socket (10^4)
  - ~100 Kilo sockets/system (10^5)



Node Level
3D packaging

Summary

- Major Challenges are ahead for extreme computing
  - Parallelism
  - Hybrid
  - Fault Tolerance
  - Power
  - … and many others not discussed here
- Not just a programming assignment.
- This opens up many new opportunities for applied mathematicians and computer scientists

Conclusions

- For the last decade or more, the research investment strategy has been overwhelmingly biased in favor of hardware.
- This strategy needs to be rebalanced - barriers to progress are increasingly on the software side.
- High Performance Ecosystem out of balance
  - Hardware, OS, Compilers, Software, Algorithms, Applications
    - No Moore's Law for software, algorithms and applications

- Our community is needed and has a great deal to offer and contribute.
- "The golden age of numerical analysis has not yet started!" - Volker Mehrmann