SX-9 の導入と大規模計算機システムについて – SX-9 and Beyond –

2008/10/24

Manabu Higashida
manabu@cmc.osaka-u.ac.jp

Installation of SX-9

Latest Vector Supercomputer in Production: NEC SX-9

- Cyber Science Center, Tohoku University
  - In 2008/03
  - 16-nodes
    - 26.2TFLOPS, 16TB Memory
- Cybermedia Center, Osaka University
  - In 2008/07
  - 10-nodes
    - 16.4TFLOPS, 10TB Memory
- Earth Simulator Center, JAMSTEC
  - In 2009/03
  - 160-nodes
    - 131.1TFLOPS, 20TB Memory

世界最高の単体CPU性能

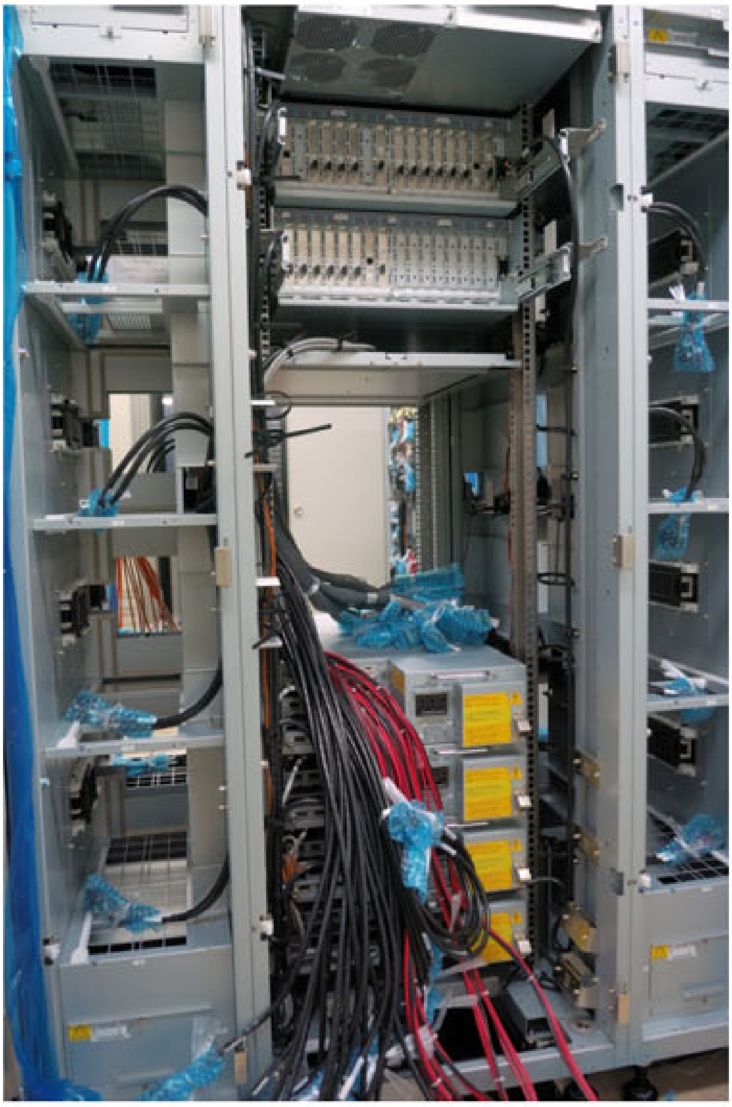
- 102.4GFLOPS
- 2.5Bytes/FLOP (256GB/s) のデータ転送

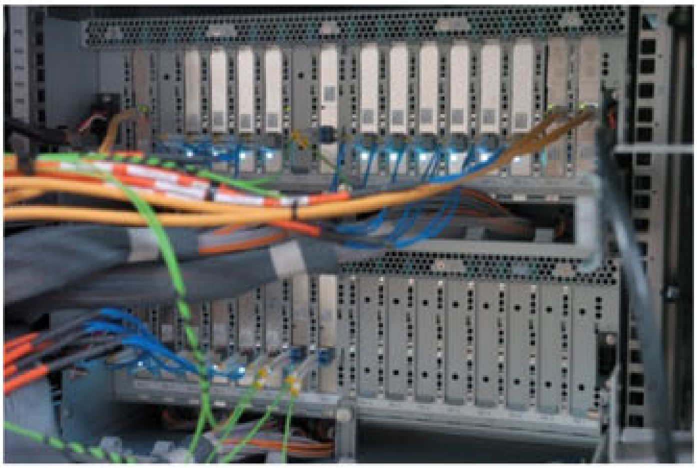
大容量共有メモリ

- 16CPU (1.6TFLOPS) で1TBの主記憶を共有

10ノード (IXS: 8ノード)

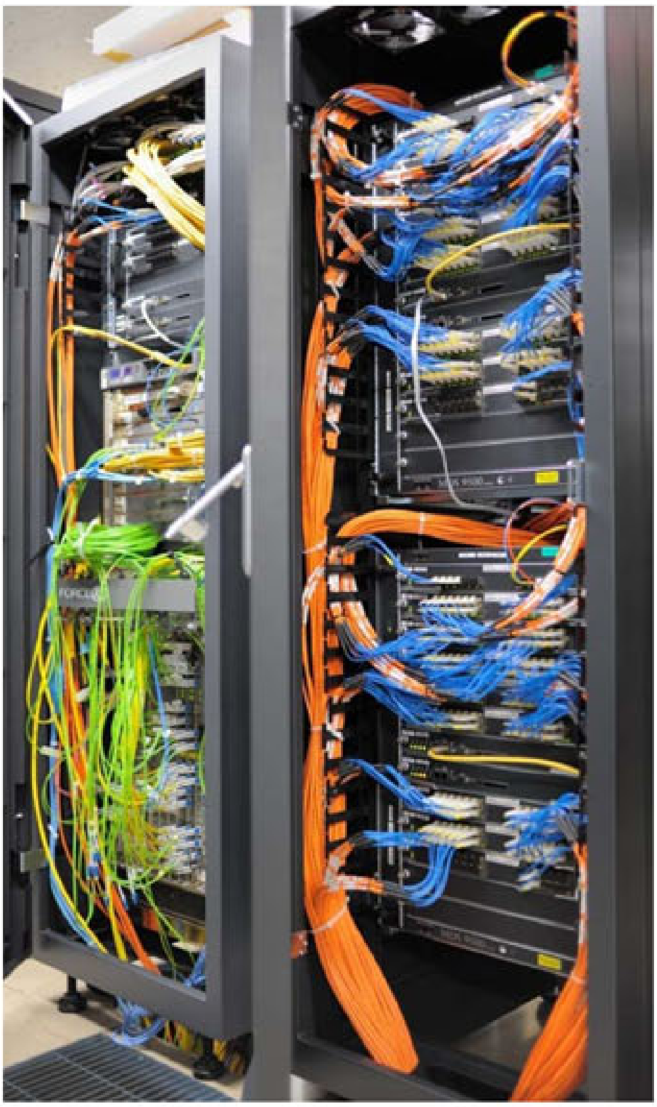
- 16.4TFLOPS (IXS: 13.1TFLOPS)
- 10TB (IXS: 8TB)





電源とI/Oラック

- I/Oは、汎用のPCI Express拡張カードを使う
- 1ノードあたり10Gbps Ethernetが3枚 (30Gbps) でODINSやSINET3に接続している
- 1ノードあたり4Gbps FibreChannelが16枚 (64Gbps) で、1PBのファイル・システムに接続している

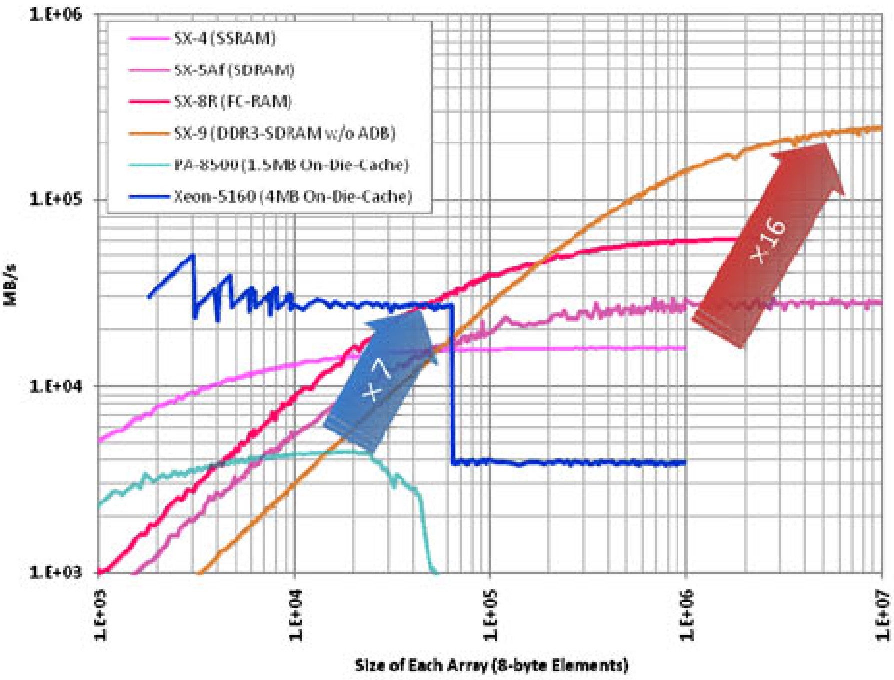


SXシリーズにおけるCPU毎のByte/FLOP値の推移

	SX-4	SX-5A	SX-5Af	SX-6	SX-7	SX-8	SX-8R	SX-9
	1994年	1998年	2000年	2001年	2002年	2004年	2006年	2007年
駆動 周波数	125MHz	250MHz	312.5MHz	1GHz	1.1GHz	2GHz	2.2GHz	3.2GHz
ピーク 演算性能値	2GF	8GF	10GF	8GF	8.83GF	16GF	35.2GF	102.4GF
メモリ バンド幅	16GB/s	64GB/s	40GB/s	64GB/s	35.3GB/s	64GB/s	70.4GB/s	256GB/s
浮動小数演算 あたりの 転送バイト数	8B/F	8B/F	4B/F	8B/F	4B/F	4B/F	2B/F	2.5B/F
ADB バンド幅	-	-	-	-	-	-	-	409.6GB/s
浮動小数演算 あたりの 転送バイト数	-	-	-	-	-	-	-	4B/F

STREAM Triad: Single Processor

- SX-4 (1994)
  - Synchronous SRAM
  - 15.7GB/s @ 100K
    - 2MB of 8GB (0.025%)
  - 7.98Byte/FLOP
- SX-9 (2007)
  - DDR3 SDRAM
  - 260GB/s @ 40M
    - 1GB of 1TB (0.1%)
  - 2.57Byte/FLOP



CPU毎の実効転送性能 (STREAM Triad)

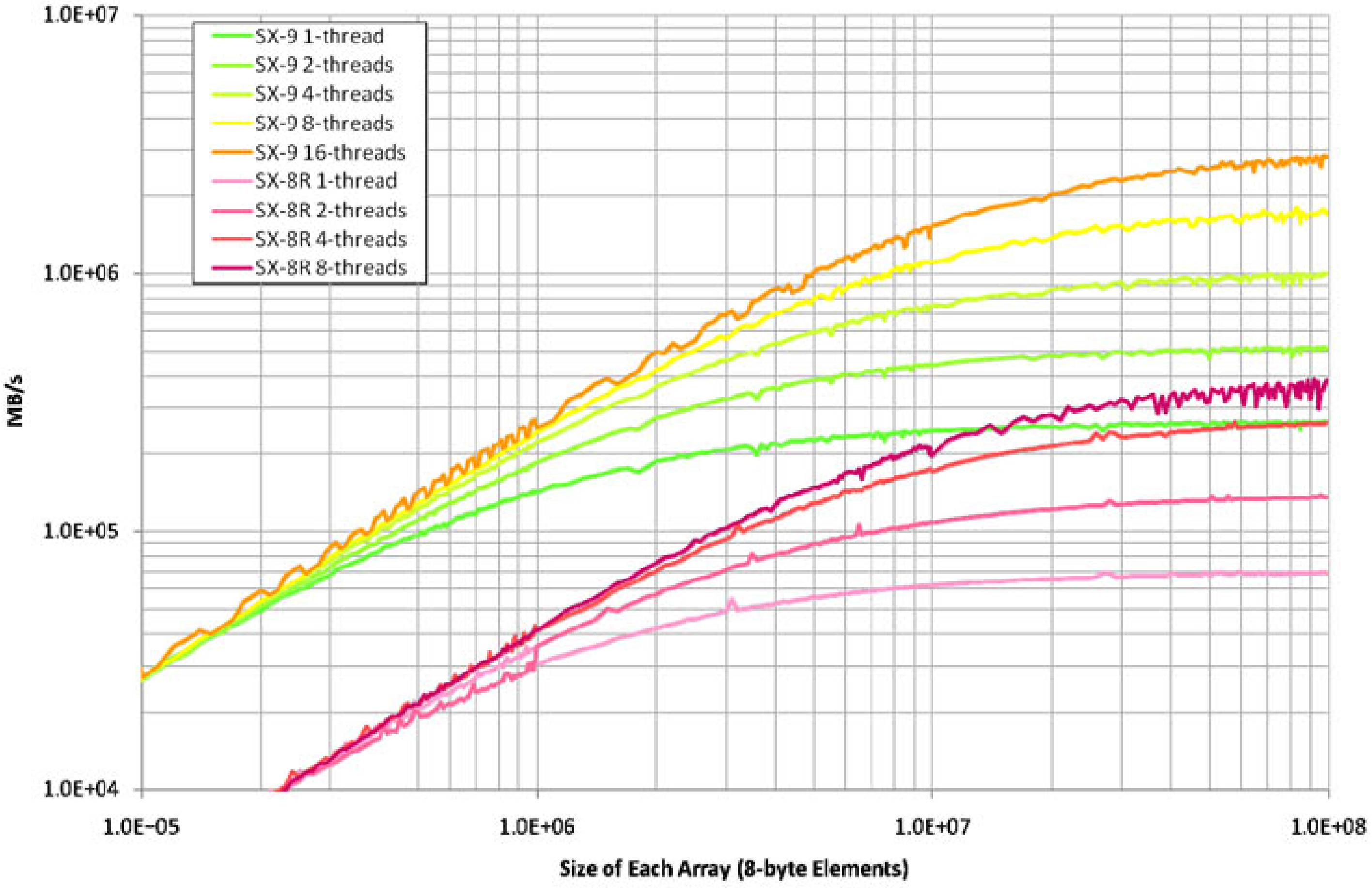
	SX-4	SX-5A	SX-5Af	SX-7	SX-8	SX-8R	SX-9
ピーク 演算性能値	2GF	8GF	10GF	8.83GF	16GF	35.2GF	102.4GF
メモリ バンド幅	16GB/s	64GB/s	40GB/s	35.3GB/s	64GB/s	70.4GB/s	256GB/s
実効性能値	15.9GB/s	47.8GB/s	28.4GB/s	35.3GB/s	-	69.0GB/s	261.6GB/s
ピーク 性能比	99.4%	74.7%	70.9%	100%	-	98.0%	102.4%
浮動小数演算 あたりの 転送バイト数の 実効性能値	7.95B/F	5.97B/F	2.84B/F	4B/F	-	1.96B/F	2.55B/F

8B/F

4B/F

2〜2.5B/F

STREAM Triad: Multi Processor



ノード毎の実効転送性能 (STREAM Triad)

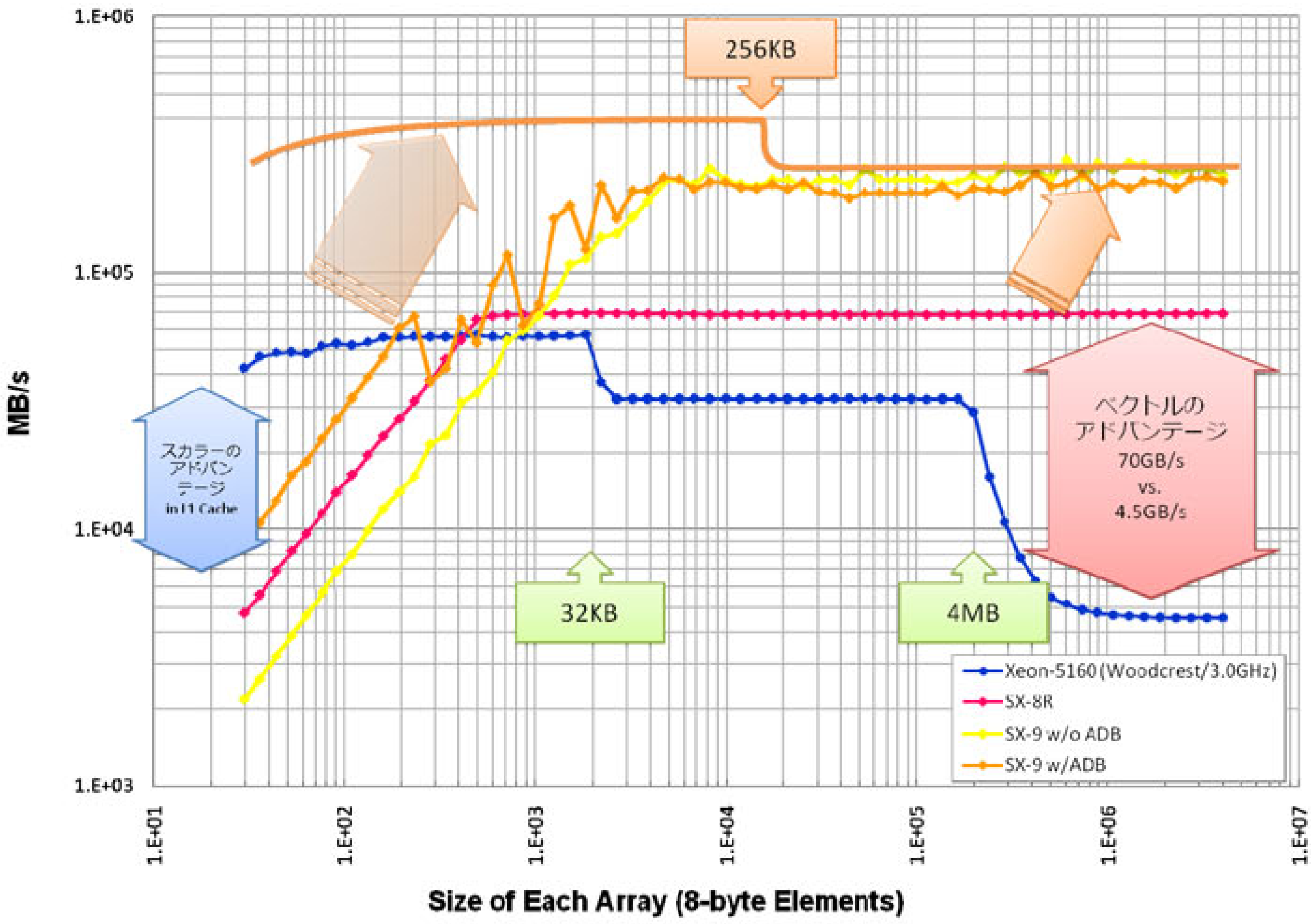
	SX-4	SX-5	SX-5Af	SX-7	SX-8	SX-8R	SX-9
ピーク演算性能値	64GF	128GF	160GF	282.56GF	128GF	281.6GF	1,638.4GF
メモリバンド幅	512GB/s	1,024GB/s	640GB/s	1,129.6GB/s	512GB/s	563.2GB/s	4,096GB/s
実効性能値	437.0GB/s	583.1GB/s	340.0GB/s	872.3GB/s	320GB/s (参考)	381.6GB/s	2,700GB/s (参考)
ピーク性能比	85.3%	56.9%	53.1%	77.2%	62.5%	67.8%	65.9%
浮動小数演算あたりの転送バイト数の実効性能値	6.83B/F	4.56B/F	2.13B/F	3.09B/F	2.50B/F	1.36B/F	1.65B/F

8B/F

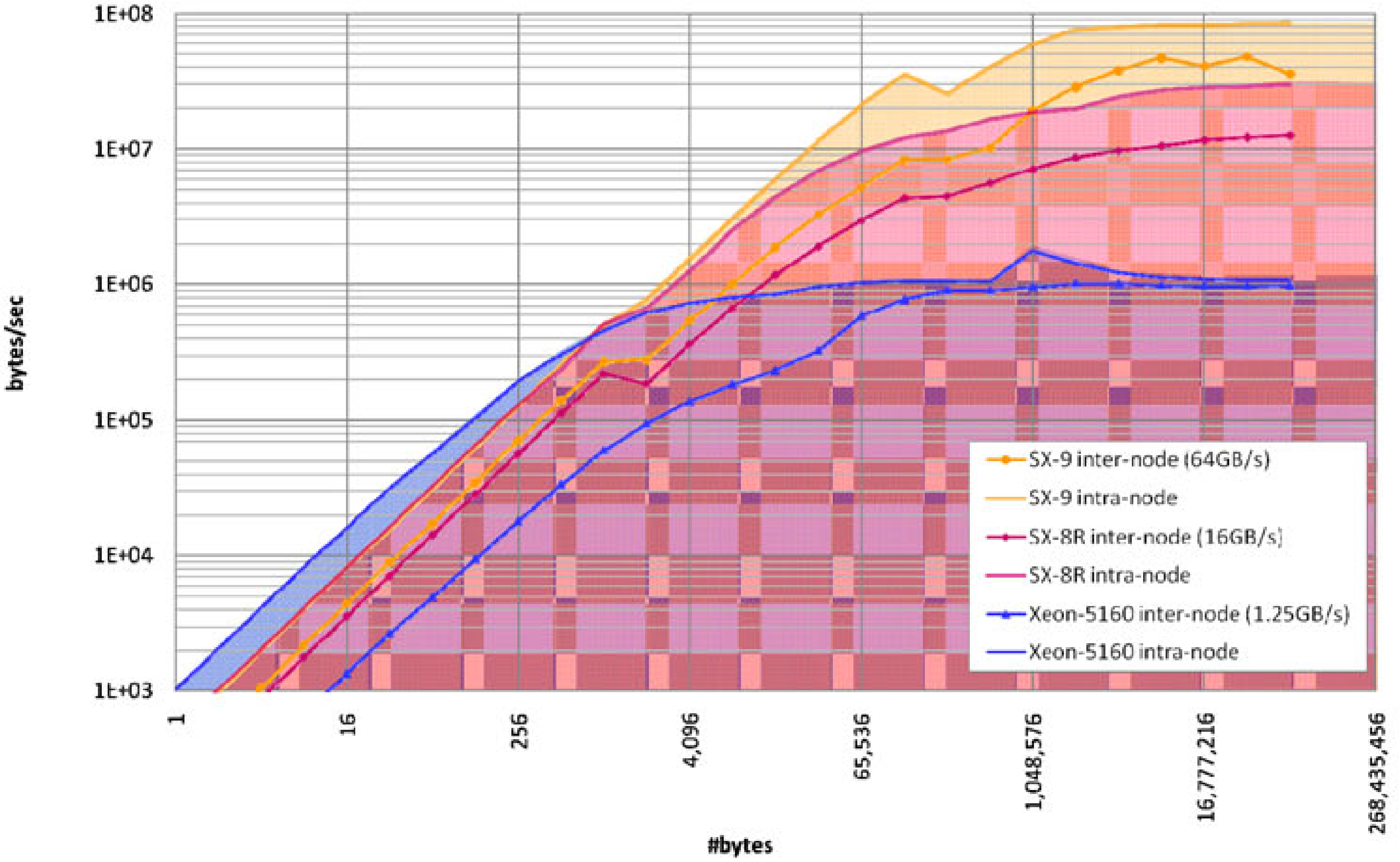
4B/F

2～2.5B/F

STREAM2 DAXPY

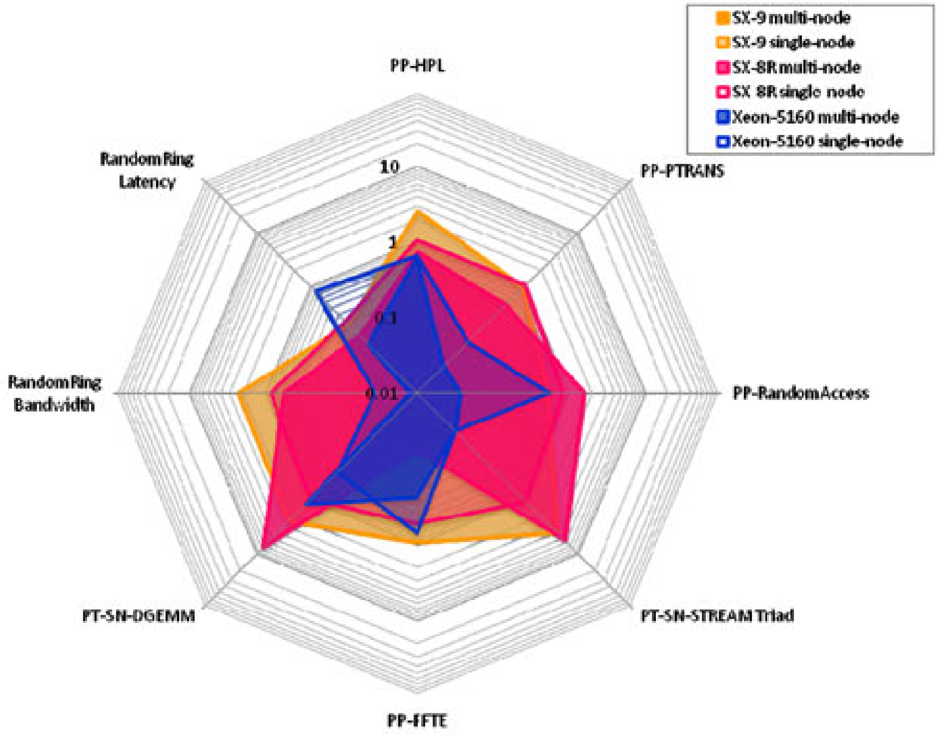


Performance of Interconnecting Network by Intel MPI Benchmark



HPCC Benchmark per Process Performance (Base Run)

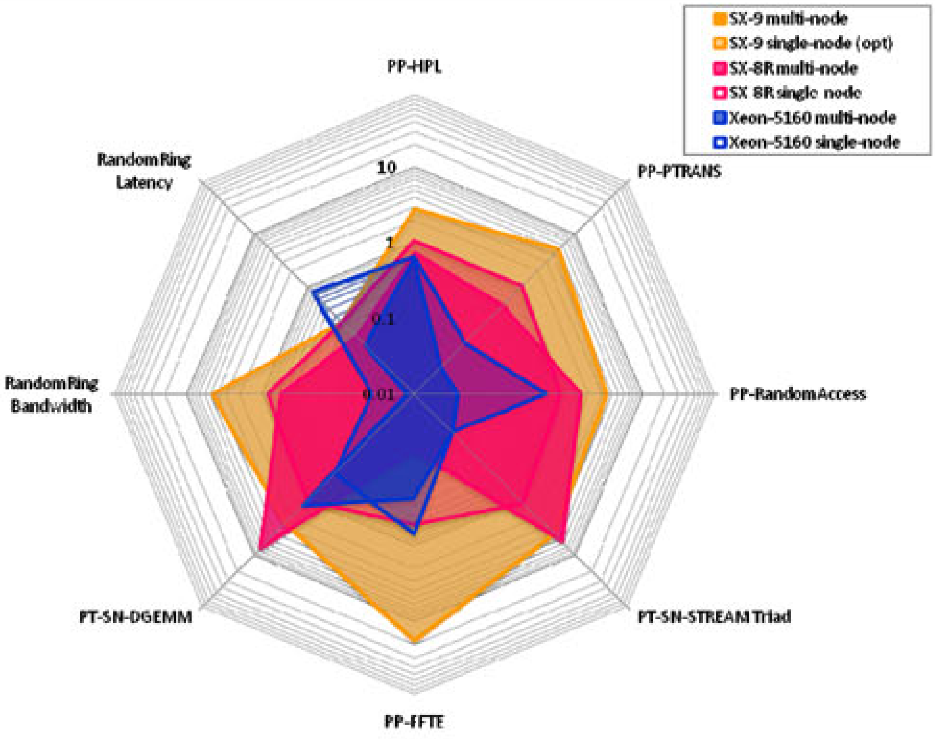
- SMP機にとって不利な条件
  - HPL (High Performance Linpack) の1プロセスが32-bitアドレッシングのため、共有メモリ型で高並列処理を行うとアクセス競合が起こる
  - HPLのパラメータを元にして他の試験項目のパラメータを導出しているため、上述のアドレス制限が他の試験のデータサイズに波及して性能が出ない (特にFFTE)



System - Processor - Speed - Processors Count – OpenMP Threads – MPI Processes					G-HPL	G-PTRANS	G-Random Access	G-FFTE	EP-STREAM Sys	EP-STREAM Triad	EP-DGEMM	RandomRing Bandwidth	RandomRing Latency
					Tflop/s	GB/s	Gup/s	Gflop/s	GB/s	GB/s	Gflop/s	GB/s	usec
NEC SX-9	3.2GHz	16	1	16	1.24832	63.4479	0.062697	12.7494	2,880.512	180.032	50.8616	31.0694	4.08536
		32	1	2	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.
NEC SX-8R	2.2GHz	8	1	8	0.25778	35.1761	0.050399	3.64893	373.924	46.7405	34.21	11.4181	3.63782
		16	1	2	0.36489	29.9111	0.183769	0.77579	721.54	360.77	278.191	7.7624	6.56754
NEC Express 120Rg-1 Intel Xeon 5160	3.0GHz	2	1	4	0.04034	0.7431	0.008019	1.23380	5.50	1.37	10.80	0.4898	0.97902
		32	1	16	0.61089	3.2971	0.008736	6.45151	51.81	3.24	43.10	0.1685	9.67466

HPCC Benchmark per Process Performance (Opt. Run)

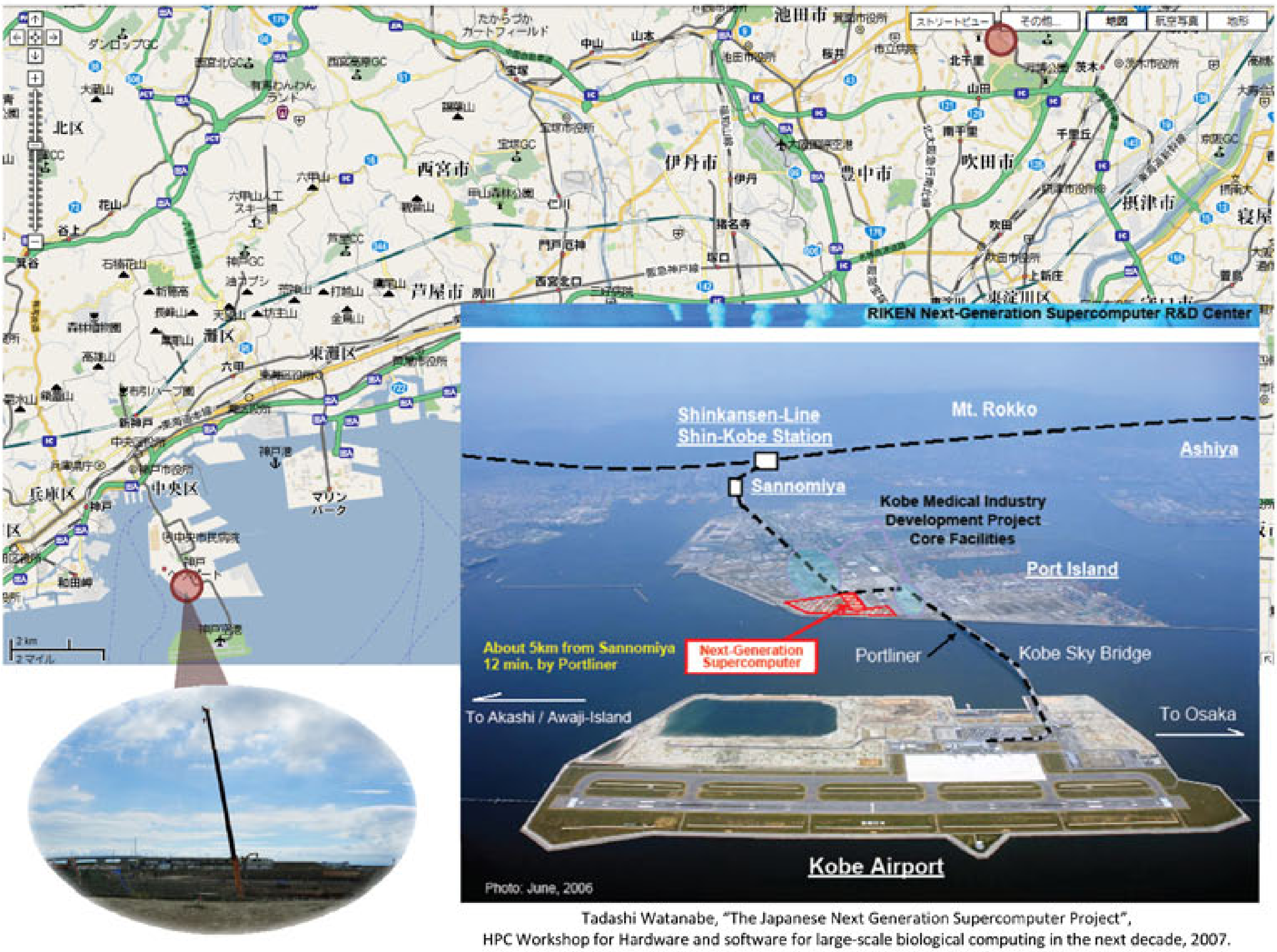
- アドレッシングとバリア同期のタイミング改善によって劇的に性能向上



System - Processor - Speed - Processors Count – OpenMP Threads – MPI Processes					G-HPL	G-PTRANS	G-Random Access	G-FFTE	EP-STREAM Sys	EP-STREAM Triad	EP-DGEMM	RandomRing Bandwidth	RandomRing Latency
					Tflop/s	GB/s	Gup/s	Gflop/s	GB/s	GB/s	Gflop/s	GB/s	usec
NEC SX-9	3.2GHz	16	1	16	1.38215	334.561	0.402194	241.713	2723.904	170.244	88.725	62.8236	3.53181
		32	1	2	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.	T.B.D.
NEC SX-8R	2.2GHz	8	1	8	0.25778	35.1761	0.050399	3.64893	373.924	46.7405	34.21	11.4181	3.63782
		16	1	2	0.36489	29.9111	0.183769	0.77579	721.54	360.77	278.191	7.7624	6.56754
NEC Express 120Rg-1 Intel Xeon 5160	3.0GHz	2	1	4	0.04034	0.7431	0.008019	1.23380	5.50	1.37	10.80	0.4898	0.97902
		32	1	16	0.61089	3.2971	0.008736	6.45151	51.81	3.24	43.10	0.1685	9.67466

Newly-build construction area of the facility of the Japanese Next Generation Supercomputer at 2008/10/07: Under ground improvement…

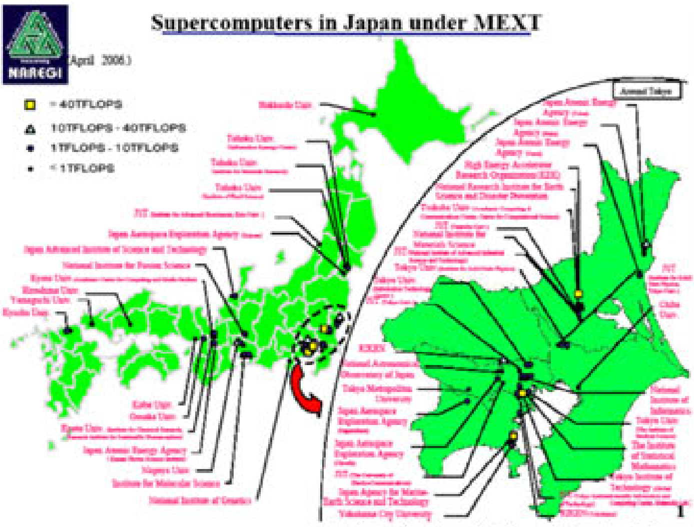
次世代スパコン施設 新築工事
"Sky Bridge" to Kobe Airport
Control Tower of Kobe Airport

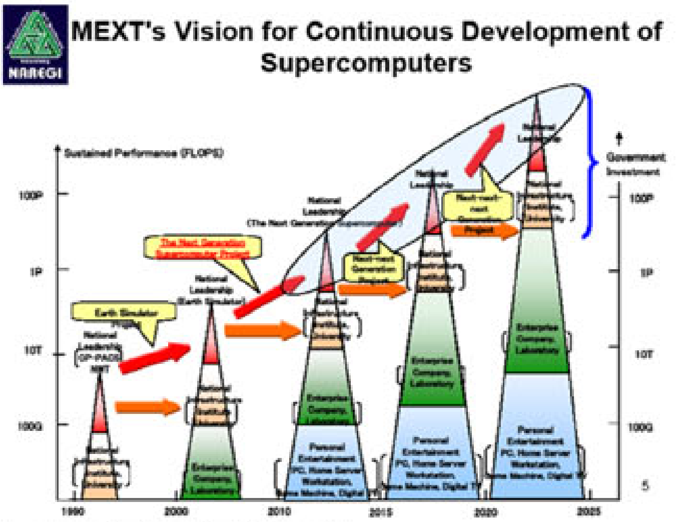


Tadashi Watanabe, "The Japanese Next Generation Supercomputer Project", HPC Workshop for Hardware and software for large-scale biological computing in the next decade, 2007.

Hierarchy of Japanese Supercomputing Systems

- NLS (National Leadership System)
  - "Earth Simulator" (under replacement to "Earth Simulator 2")
- NIS (National Infrastructure System)
  - 7+2 Academic Supercomputer Site including the Cybermedia Center
- LLS (Laboratory Level System)

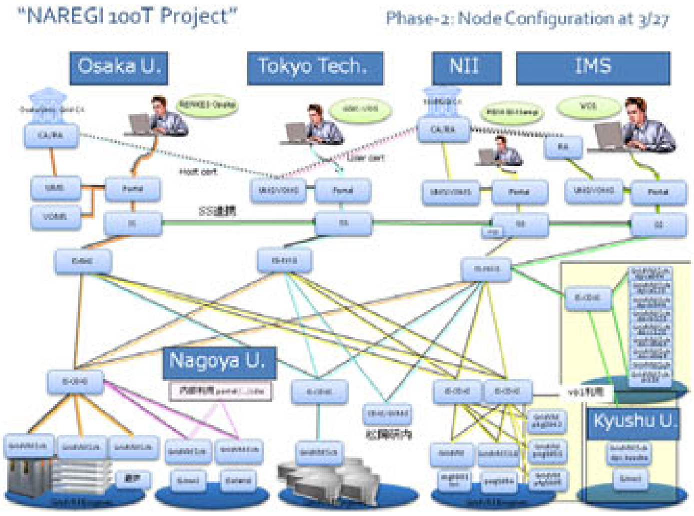

Supercomputers in Japan under MEXT
NAREGI
(April 2006)
> 40TFLOPS
10TFLOPS - 40TFLOPS
1TFLOPS - 10TFLOPS
< 1TFLOPS
Around Tokyo
Hokkaido Univ.
Tohoku Univ.
JST
Japan Aerospace Exploration Agency
Japan Advanced Institute of Science and Technology
National Institute for Fusion Science
Kyoto Univ.
Hiroshima Univ.
Yamaguchi Univ.
Kyushu Univ.
Kobe Univ.
Osaka Univ.
Japan Atomic Energy Agency
Nagoya Univ.
Institute for Molecular Science
National Institute of Genetics
Japan Atomic Energy Agency
High Energy Accelerator Research Organization (KEK)
National Research Institute for Earth Science and Disaster Prevention
Tsukuba Univ.
National Institute for Materials Science
Tokyo Univ.
Chiba Univ.
RIKEN
National Astronomical Observatory of Japan
Tokyo Metropolitan University
Japan Aerospace Exploration Agency
National Institute of Informatics
The Institute of Statistical Mathematics
Tokyo Institute of Technology
Japan Agency for Marine-Earth Science and Technology
Yokohama City University

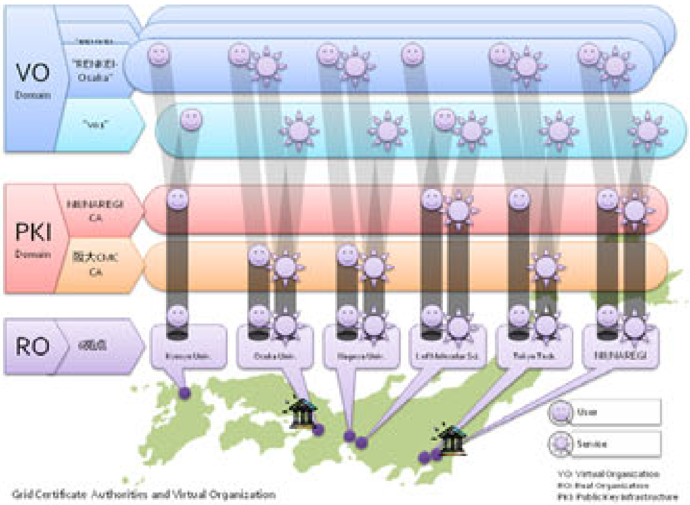


Kenichi Miura, "Outline of the Next Generation Supercomputer System Project in Japan", DEISA Symposium, 2007.

NAREGI (National Research Grid Initiative)

- Originally started as an R&D project funded by MEXT (FY2003-FY2007)
- Collaboration of National Labs. Universities and Industry in the R&D activities (IT and Nano-science Apps.)
- Project redirected as a part of the Next Generation Supercomputer Development

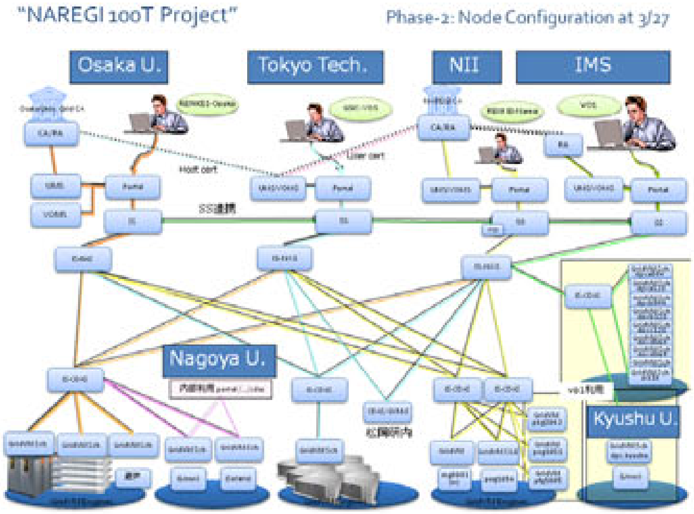


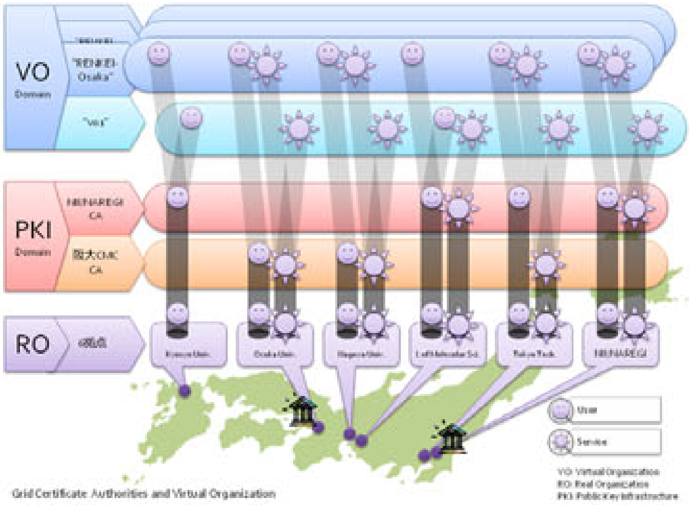


Grid Certificate Authorities and Virtual Organization

NAREGI Project Goals

- To develop a Grid Software System as the prototype of future Grid Infrastructure in scientific research in Japan
- To provide a Testbed to prove that the High-end Grid Computing Environment can be practically utilized in the Nano-science Applications over SINET (Japanese Academic Information Network)
  - "NAREGI-6" in 2007 and "NAREGI-10" in 2008





Grid Certificate Authorities and Virtual Organization