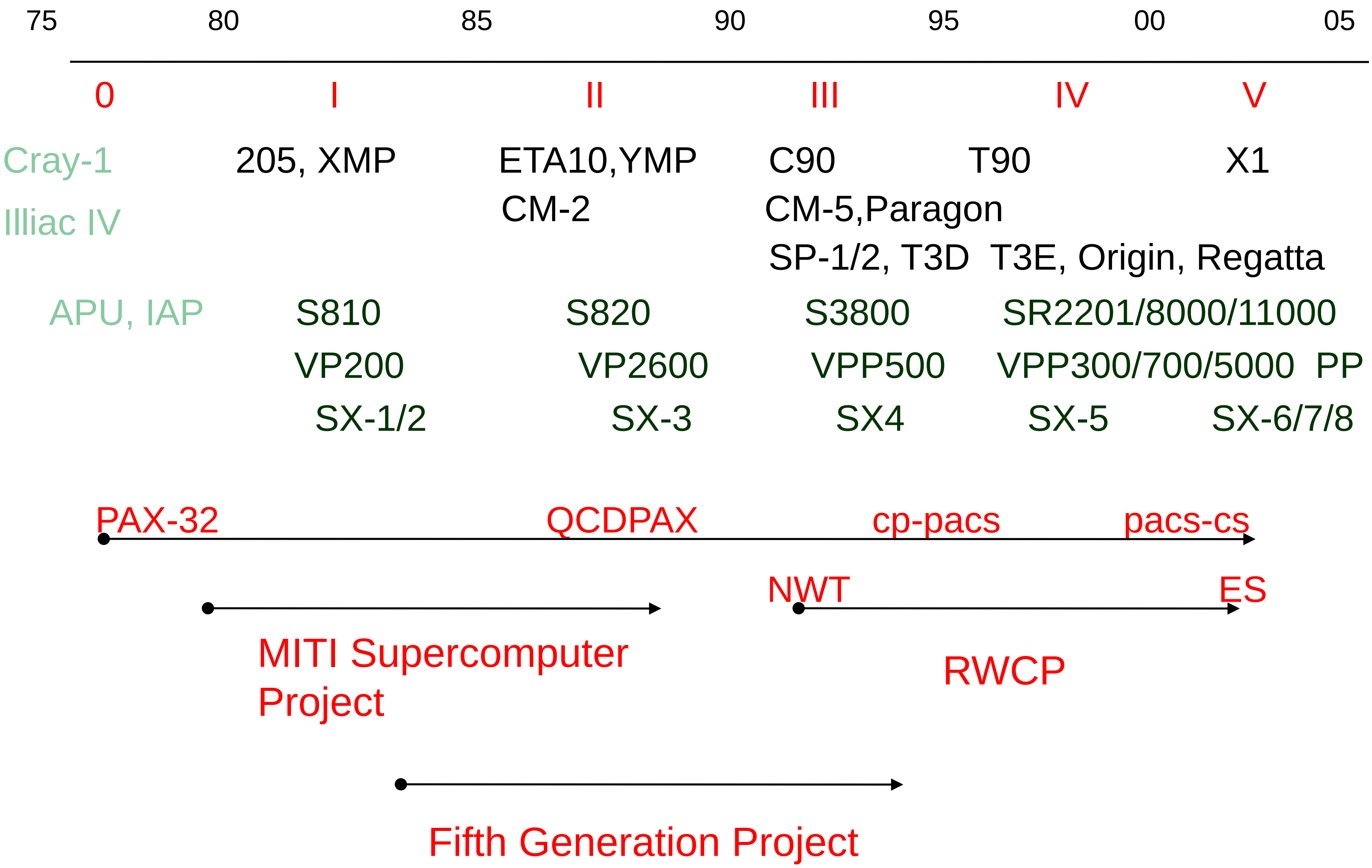
Future planning of the next generation supercomputing in Japan

- Historical overview of Japan and US HPC’s
- What is the difference between Japan and US trends?
- How can we go beyond Petaflops?

Yoshio Oyanagi
Kobe University


75
80
85
90
95
00
05
0
I
II
III
IV
V
Cray-1
Illiac IV
205, XMP
ETA10,YMP
CM-2
C90
CM-5,Paragon
SP-1/2, T3D
T90
T3E, Origin, Regatta
X1
APU, IAP
S810
VP200
SX-1/2
S820
VP2600
SX-3
S3800
VPP500
SX4
SR2201/8000/11000
VPP300/700/5000
SX-5
PP
SX-6/7/8
PAX-32
QCDPAX
cp-pacs
pacs-cs
MITI Supercomputer Project
NWT
ES
RWCP
Fifth Generation Project

Historical Overview of Supercomputers

- 1970s: Primordial age
- 1980s: Vector age, parallel started
- 1990s: Commodity parallel in USA, Japan slowly moved to parallel
- 2000s: Commodity parallel in mainstream. NEC active in vector.
- 2010s： Petaflops age
- 2020s: ??

1970's (red for vector machines)

- USA Vendors: ASC(72), STAR-100(73), ILLIAC-IV(73), Cray-1(76), HEP (79)
  - Y. Muraoka, K. Miura and others learned at ILLIAC IV.
- UK: ICL DAP (79)
- Japan. Vendors: FACOM 230/75 APU(77), HITAC M180 IAP(78)
- Kyoto U (Electric Eng.): QA-1(74), QA-2 (VLIW)
  - Signal processing, Image processing
- Kyoto U (Nuclear Eng.): PACS-9(78) (→U. Tsukuba)
  - Reactor simulation

1980's (Vectors)

- USA Vendors:
  - Cyber-205 (81), XMP-4 (84), Cray-2 (85), IBM 3090 VF (85), ETA-10 (87), YMP (88)
  - Convex C1 (85), SCS-40 (86), Convex C2 (88), Supertek S1 (89)
- Japanese Vendors:
  - Hitac S810/20 (83), S820 (87)
  - FACOM VP200 (83), VP2600 (89)
  - NEC SX-2 (85), SX-3 (90)

1980's (US Parallel)

- Parallel Ventures in US:
  BBN Butterfly (81), Cosmic Cube (83),
  Elxsi 6400 (83), Pyramid 90x (83),
  Balance 8000 (84), nCUBE/1 (85),
  Alliant FX/8 (85), Encore Multimax (86),
  FPS T-series (86), Meiko CS-1 (86),
  Thinking Machines CM-1 (86), CM-2 (87),
  Multiflow Trace/200 (87)

1980's (Japan Parallel)

- Japanese Activities (mainly for research):
  - U. Tsukuba: Pax-32 (80), Pax-128 (83), Pax-32J (84), qcdpax (89) for qcd
  - Fifth Generation (ICOT) of MITI 82-92 PIM machines for inference
  - Supercomputer Project of MITI 81-89 PHI, Sigma-1 (dataflow), CAP, VPP (GaAs)
  - Osaka U.: EVLIS (82) for LISP
  - Keio U.: SM^2 (83) for sparse matrix
  - U. Tokyo: Grape-1 (89)

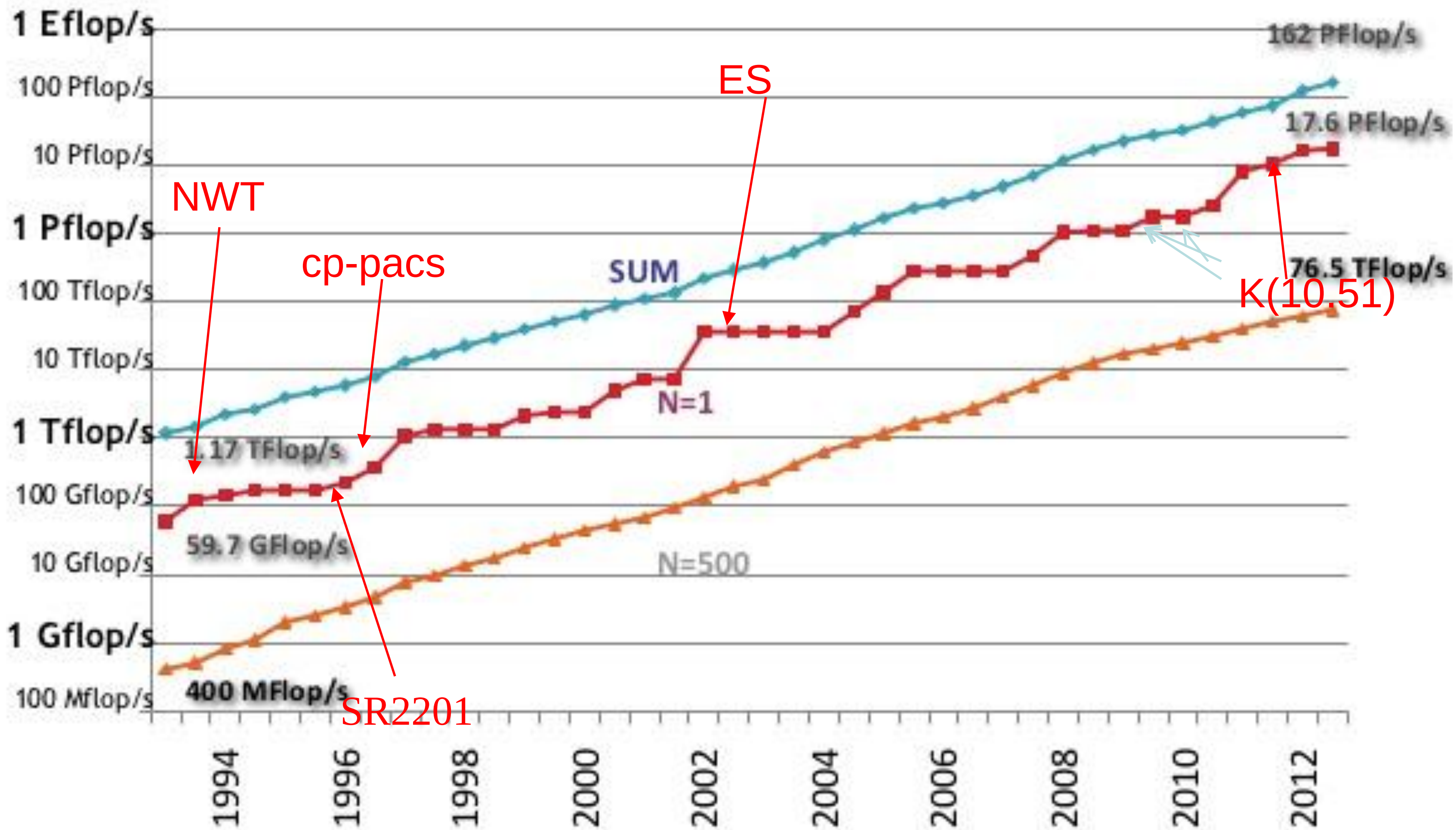
1990's (USA)

- USA Vectors: C90 (91), Cray-3 (93), T90 (95), SV1 (98)
- USA Parallel (use commodity processors):
  - CM-5 (92), KSR-2 (93, special chip), SPP (94)
  - SP1 (93), SP2 (94), ASCI Blue Pacific (97), Power 3 SP (99)
  - T3D (93), T3E (96)
  - ASCI Red (97)
  - Origin 2000 (96), ASCI Blue Mountain (98)

1990's (Japan)

- Japan. Vectors: S3800 (93), NWT (93), VPP500 (93), SX-4 (95), VPP300 (95), VPP5000 (99)
- Japan. Parallel:
  - cp-pacs (96), SR2201 (96), SR8000(98)
  - AP1000 (94), AP3000 (97)
  - Cenju-2 (93), Cenju-3 (94), Cenju-4(97)
  - Some are sold as a testbed.
- RWCP project (MITI, 92-02): Cluster connected by Myrinet. Score middleware.

Performance Development



Japanese Supercomputers in Top20

	9306	9311	9406	9411	9506	9511	9606	9611	9706	9711	9806	9811	9906	9911	0006	0011	0106	0111
1		NWT		NWT	NWT	NWT	Todai	cp-p										
2			NWT				NWT	NWT	cp-p									
3								Todai	NWT									
4									Todai	cp-p			Todai					
5	NEC		ATP	ATP	KEK	KEK								Todai	LRZ		Todai	
6	AES	NEC	Tsuku	Tsuk						Todai	cp-p				KEK			
7		AES	Riken	Riken	JAERI		KEK			NWT						LRZ		Todai
8								KEK		ECMW							Osak	
9						NEC		Kyush							Todai	KEK		
10			Hitach	Hitac			NEC	ECMW										
11			Todai	Todai		JAERI	Stutt											
12			NEC			Nagoy					AES		TAC			ECMW	LRZ	Osak
13			Toho								Toho					Todai		
14					ATP	Gene	JAERI				Todai	cp-p					KEK	
15				AES	Tsuk	ISS	Nagoy				NWT			Kyoto				
16				NEC	Riken												ECMW	
17			Toho	Toho			Gene	NEC	Kyush					TAC	ISS		Todai	LRZ
18			AES	Toho			ISS	Osaka	KEK			AES	cp-p			JMA		
19			IMS	AES		ATP		Osaka			ECMW	FZJ						KEK
20				IMS		Tsuk		Stutt										

Japanese Supercomputers in Top20

	0206	0211	0306	0311	0406	0411	0506	0511	0606	0611	0706	0711	0806	0811	0906	0911	1006	1011	1106	1111	1206	1211
1	ES	ES	ES	ES	ES														K	K		
2																					K	
3						ES																K
4							ES											TIT				
5																			TIT	TIT		
6			NAL					ES														
7					Riken				TIT													
8							AIST															
9										TIT												
10									ES													
11				JAXA																		
12								AIST													Roku	
13	Todai										TIT											
14	LRZ					Riken				ES											TIT	
15							JAERI		AIST			TIT	Todai									Roku
16																						
17									KEK													TIT
18									KEK												Todai	
19	Osak																					
20					AIST						ES		Tsuk									

Observation of Japan (1/3)

- Until late 1990's, Japanese vendors focused on vector machines.
- Users exploited the power of vectorization.
- Vendors thought parallel machines were for specialized purposes (eg. image processing). Most users dared not try to harness parallel machines in the 80's.
- Some computer scientists were interested in building parallel machines, but they were not used for practical scientific computing.

Observations of Japan (2/3)

- Practical parallel processing for scientific computing was started by application users: qcd-pax, NWT, cp-pacs, GRAPE's, ES.
- Softwares
  - Very good vectorizing compilers.
  - Users were spoiled by them.
  - Users found difficulties in using message passing.
  - HPF efforts for the Earth Simulator.
  - OpenMP
  - Score middleware from RWCP

Observation of Japan (3/3)

- Japan was at least ten years late in parallel processing for scientific computing as compared to US around 2000.
- Industry reluctant in parallel computing.
- Education in parallel processing is a urgent issue for the K computer 「京計算机」(10PF machine).
- More collaboration of computer scientists and application scientists is needed.

京(K) Supercomputer of Japan

- Started in April 2006, open to users in September 2012
- Over 10 PF with LINPACK
- Site: Kobe (Port Island)
- Run by Riken (AICS)
- Architecture
  - Octacore scalar processor (Sparc64 viiifx)
  - Tofu interconnect (6-dim torus)

History of the K Computer

- Preparation (2004－2007)
- Proposed architectures and killer applications
- Hybrid machine of vector and scalar
- Five strategic application fields defined
- Withdrawal of vector machine in 2009
- the Government Revitalization Unit proposed to shutdown the Next Generation Supercomputer Project (Nov. 13, 2009)
- HPCI started
  (High Performance Computing Infrastructure)

Preparations in Japan

- IT Strategic Headquarter (2001): e-Japan, but only network was emphasized.
  - At this stage, level up of supercomputers was to be promoted according to the needs of each field (not a national project).
- Earth Simulator attained 36 Tflops (2002)
- Information Science and Technology committee in Mext has been discussing the measures to promote computational science and technology since August 2004.

Preparations in Japan

- Recommendation to a Mext committee (2005):
  - Promote a national project to construct a leading edge supercomputer
  - Government decision (July 25, 2005)
- Riken started the project (October 2005)
- Mext funded four projects to promote "Element Technologies for Future Supercomputers" in 2005-2007. $40M per year (in total)
  - Four groups were accepted
  1. System Interconnect (Kyushu U and Fujitsu)
  2. Interconnect by IP (U of Tokyo, Keio U etc)
  3. Low Power Device and Circuits (Hitachi, U of Tokyo, U of Tsukuba)
  4. Optical Connection of CPU and Memory (NEC and Titech)

Killer Applications

- The WG in Mext identified killer applications
  - Life Science
  - Astrophysics
  - Space and Aeronautics
  - Materials
  - Atomic Energy
  - Environment
  - Disaster Prevention
  - Fluid Dynamics
  - Plasma (space/fusion)
  - Industrial Design

広汎な分野での利活用 －次世代スパコンが拓く世界－

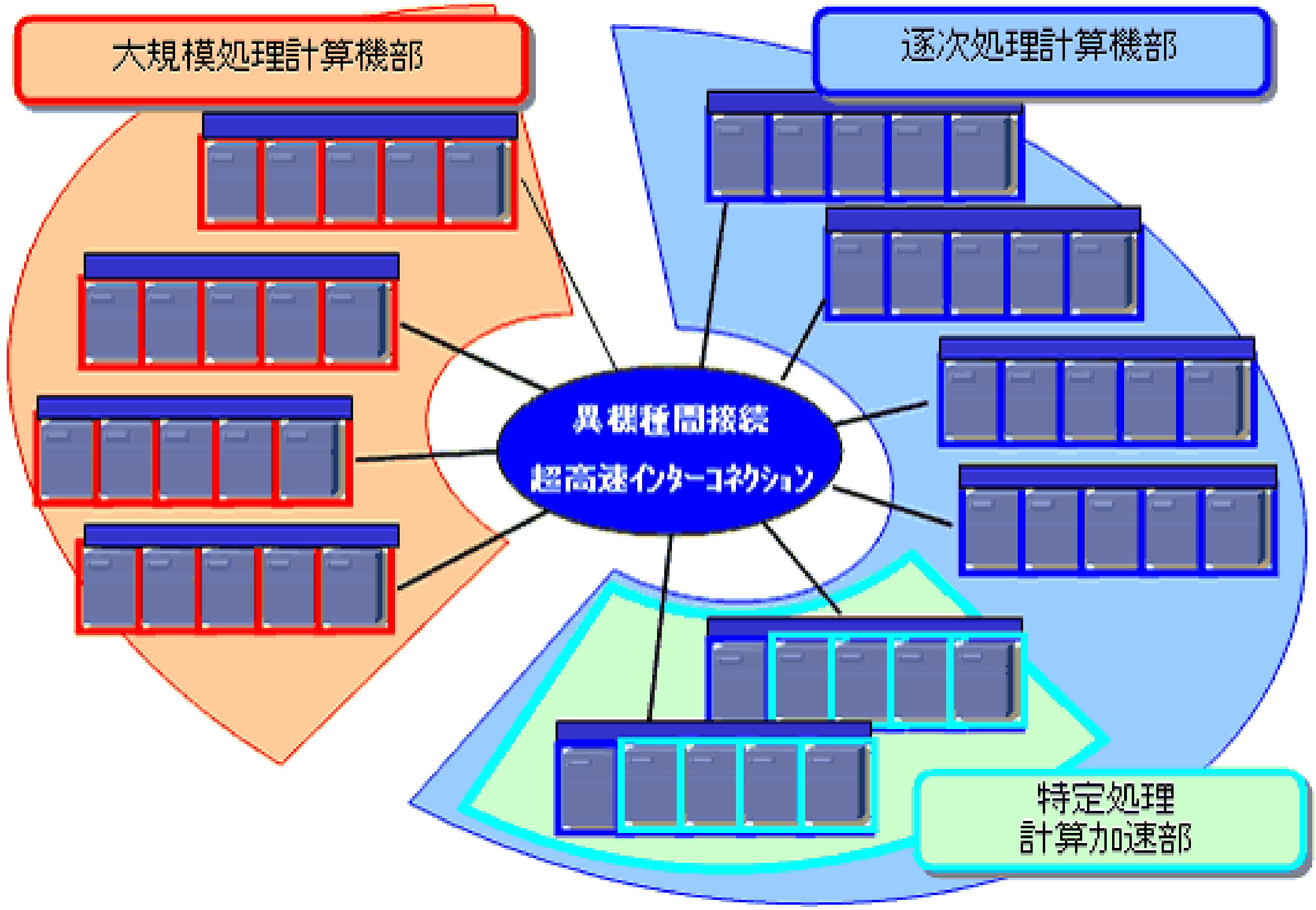
Observations of the WG

- Multiphysics-multiscale simulation will be important in those fields.
- To keep high performance in different types of computing, hybrid architecture is appropriate.
  - We estimated required performances.
- The interconnection between different architectures should have high speed.

Original Proposal

Large scale processing

Scalar computer



Special-purpose computer

Architecture

- Conceptual design of hardware (Sept. 14, 2007) The system consists of two parts:
  - Scalar processor part by Fujitsu
  - Vector processor part by NEC and Hitachi
- Site selection: Kobe (March 28, 2007)
- NEC retired from the project (May 13, 2009)
- Riken decided to continue the joint development with Fujitsu to build 10 Pflops machine (May 14, 2009).

SPIRE (Strategic Programs for Innovative REserch)

- Mext identified five strategic fields (July 22, 2009):
  1. Predictive bioscience, medicare and drug design
  2. New materials and new energy
  3. Earth environmental prediction for disaster prevention and mitigation
  4. Next-generation manufacturing
  5. Origin and structure of material and universe

Friday, November 13, 2009

The 3rd WG of the Government Revitalization Unit

"Why should it be No.1 in the world?"

"Is No.2 not enough?"

Vote：
Abolish 1
Postpone 6
Budget shrink 5

Conclusion
Freeze the project!

←村田(謝) 蓮舫(Hsieh Lien Fang)

Revival

- Strong reactions from academia and industry.
- Government decided to overturn the conclusion (Dec. 2009)
- Leading organizations for five strategic fields announced (Jan. 2010)
- HPCI started (March 2010)
  - High Performance Computing Infrastructure

Leading Organizations for SPIRE

1. Predictive bioscience, medicare and drug design: Riken
2. New materials and new energy: Institute for Solid State Physics, Univ. of Tokyo
3. Earth environmental prediction for disaster prevention and mitigation: JAMSTEC
4. Next-generation manufacturing: Institute for Industrial Sciences, Univ. of Tokyo
5. Origin and structure of material and universe: Center for Computational Sciences, Univ. of Tsukuba

Nickname 京

- Proposals were solicitated from public.
- Final decision (July 5, 2010):
  - 「京(Kei)」(The K Computer)
  - 「京」＝10^{16}＝「億億(亿亿)」(in Chinese)
  - Capital or big city（北京、南京、東京、京都）
  - Originally means “big gate”
- Fujitsu disclosed SPARC64 viiifx chip (July 9, 2010)
- The first eight racks were shipped to Riken, Kobe (Sept. 28, 2010)

AICS, Riken

航空写真

研究棟(南側)

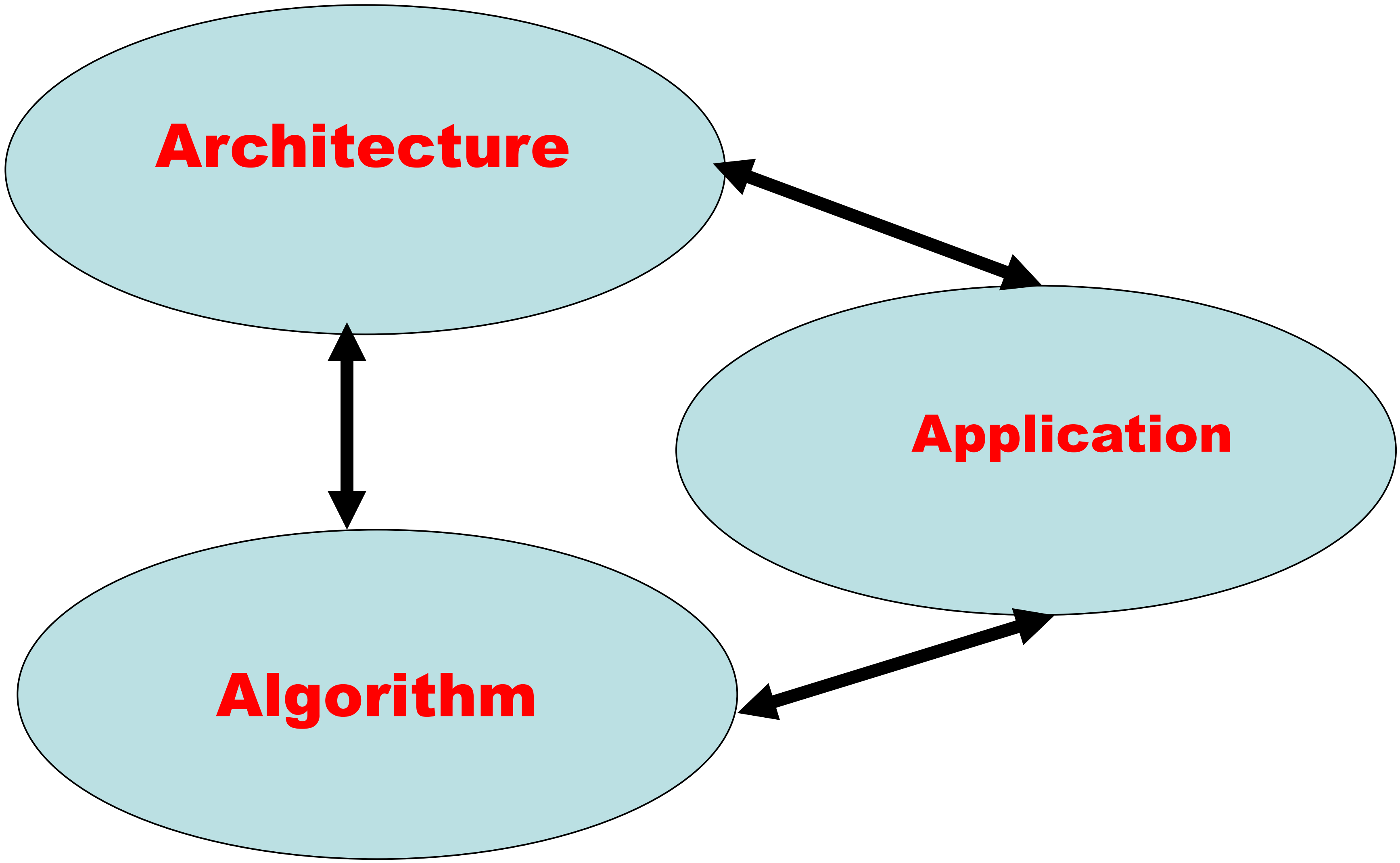
理研webより

No.1 in the world

- 8.162 Pflops attained using 80% of the system—No.1 in the Top500 (ISC2011) June 20, 2011
- 10.51 Pflops in SC2011 (Seattle)
- Now tuning the system using strategic test programs in various fields.
- Open to users: September 2012
- RIST is to manage K Computer users
  (Research Organization for Information Science & Technology)

What's next? Toward EXA FLOPS!!

- Preliminary consideration among scientists
- Mext started a WG for future HPC (April 2011)
  - Hardware-System Software-Application co-design is important.
  - Should be science-driven.
    - We identified possible break throughts in science and technology. Social needs are also considered.
    - Linpack Exaflops is not our target.
  - Limitation by budget, power, foot print ….
  - Several different architectures are considered.


Architecture
Application
Algorithm

Two subgroups worked

- Application Subgroup
  - Application
  - Numerical library
  - Algorithm
  - Automatic tuning
- System Subgroup
  - CPU and architecture
  - Compiler
  - System software

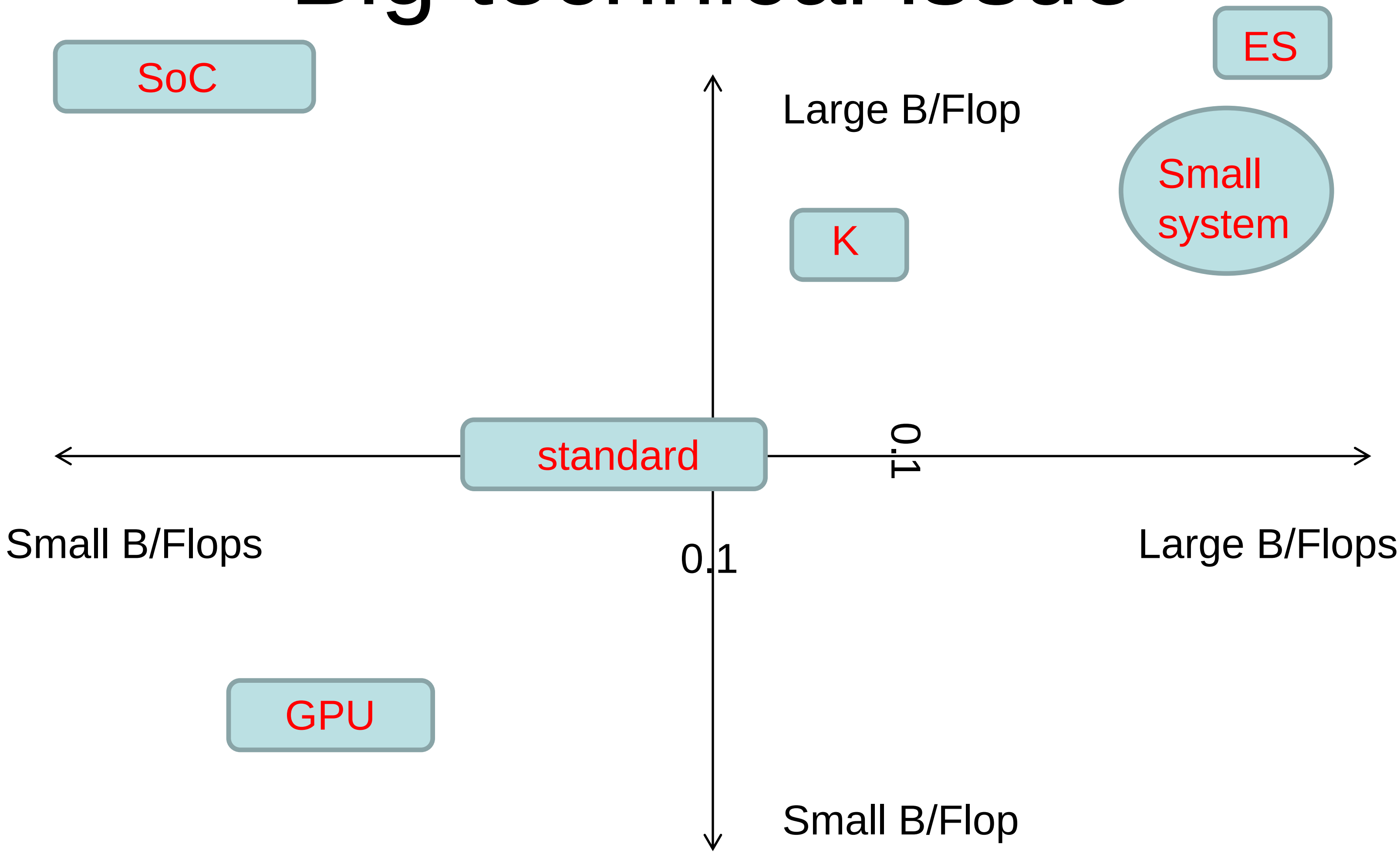
Final Report in March 2012

--Executive summary "Report on the development of future HPCI technology“ (28 slides)
http://www.open-supercomputer.org/workshop/report/FutureHPCI-Report.pdf

--Roadmap of computational sciences (158 pages)
http://www.open-supercomputer.org/workshop/report/science-roadmap.pdf

--Roadmap of HPCI technology (108 pages)
http://www.open-supercomputer.org/workshop/report/hpci-roadmap.pdf

Memory and memory bandwidth
Big technical issue



Memory and memory B/W

- The K Computer (single node)
  - 64 GB/s for 128 Gflops---0.5 B/Flop
  - 16 GB for 128 Gflops---0.125 B/Flops
- Standard EXA
  - 0.1 EB/s for 1 Eflops---0.1 B/Flops
  - 10-100PB for 1 Eflops---0.01-0.1 B/Flops
- Limitation
  - Cost, power, floor, fault tolerance, …..
  - Programmability

Efforts in Japan

- Feasibility studies of future HPCI systems (2012-3):
  - One application team
  - Three system design teams
- Working group to consider future HPCI policy (chair: oyanagi):
  - National and international computer technology
  - User needs to HPCI
  - Possible scientific and social outcome of HPCI
  - Necessary computing resources in the future
  - Possible HPCI systems
  - Necessary cost and benefits

Efforts in Japan

- Working group
  - Interim Report (May 2013)
  - Public Review
  - Budget Request for 2014fy (August 2013)
  - Final Report (March 2014)
- Recommendation of the WG
  - We should have a computer 100x faster than K in practical applications
  - We have to build one with our own technology
  - We should not aim at winning No.1 in Top500

Conclusion (1/2)

- In Japan, due to the success of vector computers in the 1980’s, parallel processing was behind US and Europe.
- Practical parallel computers (NWT, cp-pacs, ES) were built in collaboration with application users.
- Strong head wind to the K Computer
- By the success of the K Computer, we caught up in parallel processing.
- The K is very stable and used extensively.

Conclusion (2/2)

- We hope to build exascale supercomputers around 2020
- Different applications require different architecture in terms of B/Flop and B/Flops
- They should be science-driven, not Linpack-driven
- Support of taxpayers is important
- Very hard to program on Exascale due to memory hierarchy
- Applications strongly demand such machines

Thank you for your attention