HPC Trends in Japan
—Toward Exaflops—

- 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

Kogakuin University, Tokyo

Historical Overview

- 1970s: premordial 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.

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)
- Japan. 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),
  CM-1 (86), CM-2 (87),
  Multiflow Trace/200 (87)

1980's (Japan Parallel)

- Japan. 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), 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)
  - Except SR's, they are sold as a testbed.
- RWCP project (MITI, 92-02): Cluster connected by Myrinet. Score middleware.

Observation (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 (2/3)

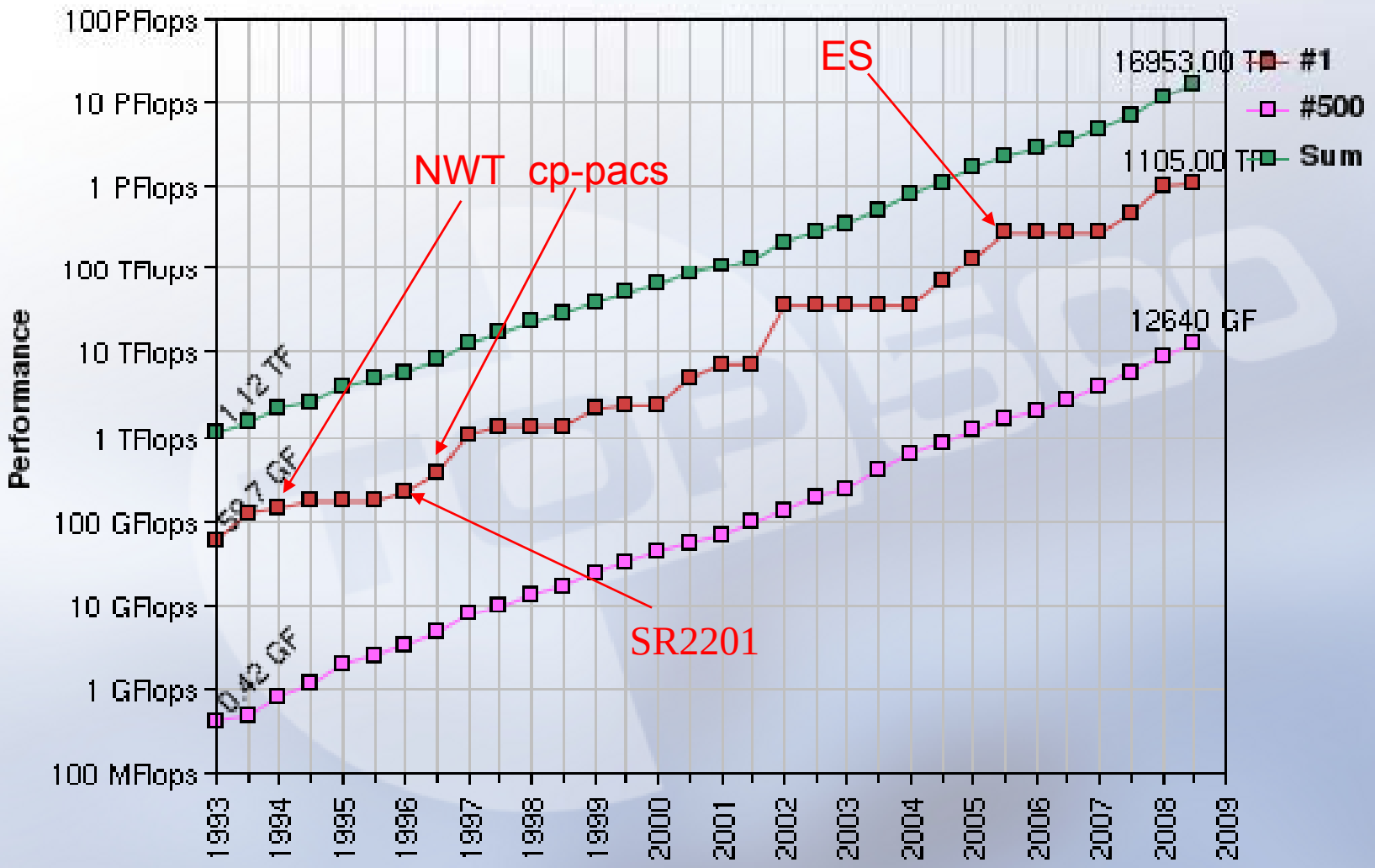
- Practical parallel processing for scientific computing was started by application users: qcd-pax, NWT, cp-pacs, GRAPE's, ES.
- Softwares
  - Very good vectorizing compilers.
  - Users were spoiled by them.
  - Users found difficulties in using message passing.
  - HPF efforts for the Earth Simulator.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

Observation (3/3)

- Japan was at least ten years late in parallel processing for scientific computing as compared to US.
- Education in parallel processing is a urgent issue for the Next Generation Super (10PF machine).
- More collaboration of computer scientists and application scientists is needed.

As a result

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
1	ES	ES	ES	ES	ES									
2														
3						ES								
4							ES							
5														
6			NAL					ES						
7					Riken				TIT					
8							AIST							
9										TIT				
10									ES					
11				JAXA										
12								AIST						
13	Todai										TIT			
14	LRZ					Riken				ES				
15							JAERI		AIST			TIT	Todai	
16														
17									KEK					
18									KEK					
19	Osak													
20					AIST						ES		Tsuk	

Future of Japan

- Japan has to provide supercomputer infrastructure to promote scientific and engineering research.
- Japan started a seven year project to develop supercomputer and its applications.

Next-Generation Supercomputer of Japan

- Started in April 2006, fully operational in March 2012
- Over 10 PF with LINPACK
- Site: Kobe (Port Island)
- Architecture (two combined)
  - NEC-Hitachi: low power vector
  - Fujitsu: multicore scalar processor
- Detailed design and its evaluation

兵庫県神戸市

所在地	兵庫県神戸市中央区港島南町7丁目(ポートアイランド第2期内) ・ポートアイランド南駅より徒歩約1分(JR新神戸駅から25分)

広さ	候補地　40,000㎡(拡張用地含む) (準工業地域)	所有者	神戸市

Technology Trend of High End Computing

- Moore's Law
- Multicore Technologies

Moore's Law

- So called Moore's Law: Number of transistors on a chip doubles every 18 months.
- More than enough for a single processor.
- New Moore's Law: Number of cores on a chip doubles every 18 months.

Multicore technologies

- Many identical CPU's on a chip (multicore)
- Many FPU's on a chip (manycore)
  - Attached SIMD like MMX/SSE
  - GPGPU
  - Accelerator chip like ClearSpeed
  - Cell Processor (CPU＋8 cores)
  - Special Purpose chip: Grape/MD-Grape/Grape-DR
  - Intel, AMD, ……

Technological challenge

- Manycore-Multicore processors
- Hierarchical memory structure
- Insufficient memory bandwidth
- Load balance among nodes
- Optimizing communications (latency hiding, pipelining)
- Debugging parallel codes

Architecture and Algorithm

- Computing Complexity
- Memory Hierarchy
- Architecture and Application

Complexity

- Time complexity: number of arithmetic operations
  - Matrix multiplication $O(n^3)$
  - Strassen $O(n^{\log 7})$
- Computing time used to be nearly proportional to the number of operations
- Now, FLOPS is much cheaper than Bytes/s

Memory Hierarchy

- FPU – registers - L1 cache - L2 cache – L3 cache – memory – paging disk
- Computing time depends on memory access
- Parallel processing: communication time
- Multicore: on-chip memory

Application Users in Japan

- Either
  - Buy a package and use it, or
  - Proprietary software developed by seniors
- Usually they do not dare to change even a line of their software.
- “Please speed up my program as it is!!”

Vector Supercomputers

- In 1980's, vector computers became available in Japanese universities.
- They provided very good vector compiler as compared to U.S.
- With minimal tuning, we could get 10 to 30 times speed-up.
- Not the case for parallel computers!

Parallel Computer research in Japan

- Large number of parallel computer researches in universities and laboratories since 1970s
- Applications users were not interested in parallel computing. “It is too difficult to use it.”
- No commercialization before 1990

In designing the next generation supercomputer

- Application users were not interested in the architecture.
  - They at most tune their code for a given computer.
- Architects tends to build a LINPACK machine.
  - They believe general-purpose machine should fit to all.

Dream of “Tensor Processor”

A possible paradigm beyond Peta?

Why vector was fast?

- Operation issue at every clock cycle in pipelining
- Parallel pipes
  - Element parallel
  - Chaining
- Fast vector registers（Except CDC）
  - Latency hiding、data reuse
- Bank memory（256～1024 banks）
  - Bandwidth: 0.5 W / Flop

BLAS (Basic Linear Algebraic Subroutines)

- BLAS
  - Basic subroutinges in LINPACK, LAPACK
  - Linear kernels optimized for a machine
- Level-1 BLAS
  - Vector operation : DAXPY $y = a \times x + y$
- Level-2 BLAS
  - Matrix-vector operation : $y = Ax, \quad x = T^{-1}y$
- Level-3 BLAS
  - Matrix-Matrix operation : $A = BC$

Why higher level BLAS?

- Level-1 BLAS: eg. DAXPY
  - Op: $2n$ I/O: $3n$
- Level-2 BLAS: Matrix-vector prod.
  - Op: $2n^2$ I/O: n^2+2n
- Level-3 BLAS: Matrix-matrix prod.
  - Op: $2n^3$ I/O: $3n^2$
- Op-I/O ratio is essential
- In Level-2, if you fix the matrix, I/O count is 2n

Beyond Vector Processing

- FP is cheap, but I/O with memory is not.
  - Optical connection is faster but power consuming
- How to reduce the bandwidth below 0.5 W/Flops?
  - Level-3 BLAS on a chip
    - B: $n \times n$、A and C: $n \times m$ ($n \ll m$)、B: fixed
  - Level-2 with fixed matrix
- I will call it “Tensor Processing”.

Tensor Processing

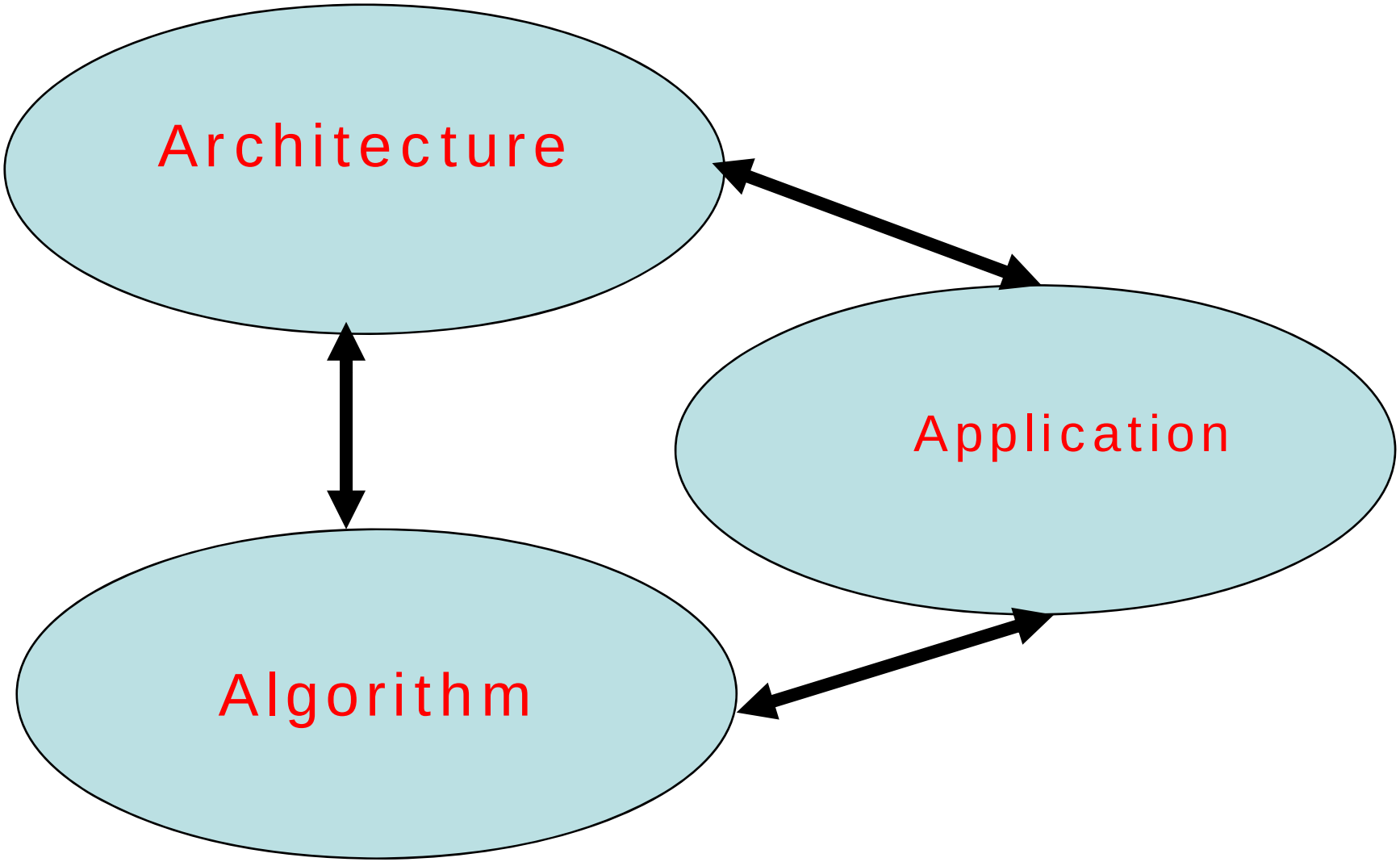
- Tensor register: memorize a fixed matrix ($n \times n'$)
- For one word from the memory, it performs *n* FP operations and output should be one (or two) word.
- Or, $O(n^2)$ operations for n words.

Is Tensor Processing practical?

- Not all computations can be processed by tensor processor.
- Core of the LINPACK is a matrix- matrix product called DGEMM
- New algorithms for the tensor processing?
- Best architecture for the algorithms?

Conclusion

- Next-Generation Supercomputer may not be a simple extension of current machines.
- We should be architecture conscious.
- New algorithms suited to new architectures are needed.
- Alliance of 3 A's is very important especially beyond Peta


Architecture
Application
Algorithm