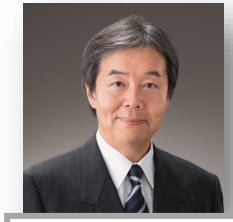
Keynote at Int'l Conf. on Image Processing and Robotics (ICIPRoB2022) in Sri Lanka (online) on March 12, 2022

Green Multicore Computing for Scientific, Image and Deep Learning Computation



Hironori Kasahara, Ph.D., IEEE Fellow, IPSJ Fellow

Senior Executive Vice President, Waseda University

IEEE Computer Society President 2018

1980 BS, 82 MS, 85 Ph.D. , Dept. EE, Waseda Univ.
1985 Visiting Scholar: U. of California, Berkeley,
1986 Assistant Prof., 1988 Associate Prof., 1989-90 Research Scholar: U. of Illinois, Urbana-Champaign, Center for Supercomputing R&D, 1997 Prof.,
2004 Director, Advanced Multicore Research Institute,
2017 Member of the Science Council of Japan
2020 Board Mem of the Engineering Academy of Japan
and 2018 IEEE Computer Society President
Nov. Senior Vice President, Waseda Univ.

<AWARD> 1987 IFAC World Congress Young Author Prize
1997 IPSJ Sakai Special Research Award,
2005 STARC Academia-Industry Research Award,
2008 LSI of the Year Second Prize,
2008 Intel Asia Academic Forum Best Research Award,
2010 IEEE CS Golden Core Member Award
2014 Minister of Japan Edu., Sci. & Tech. Research Prize
2015 IPSJ Fellow, 2017 IEEE Fellow, Eta Kappa Nu
2019 Spirit of IEEE Computer Society Award,
2020 IPSJ Contribution Award, 2021 SCAT President Grand Award

Reviewed Papers: 231, Invited Talks: 223,
Granted Patents: 66 (Japan, US, GB, China),
Articles in News Papers, Web News, TV etc.: 688

Committees in Societies and Government 284
IEEE Computer Society: President 2018, Executive Committee(2017-2019), BoG(2009-14), Strategic Planning Committee Chair 2018, Multicore STC Chair (2012-), Japan Chair(2005-07),
IPSJ Chair: HG for Magazine. & J. Edit, Sig. on ARC.
【METI/NEDO】 Project Leaders: Multicore for Consumer Electronics, Advanced Parallelizing Compiler, Chair: Computer Strategy Committee
【Cabinet Office】 CSTP Supercomputer Strategic ICT PT, Japan Prize Selection Committees, etc.
【MEXT】Info. Sci. & Tech. Committee,
Supercomputers (Earth Simulator, HPCI Promo., Next Gen. Supercomputer K) Committees
JST Moonshot Project G3 Robot & AI Vice Chair,
【COCN】Board Member in Council of Competitiveness Nippon, etc.

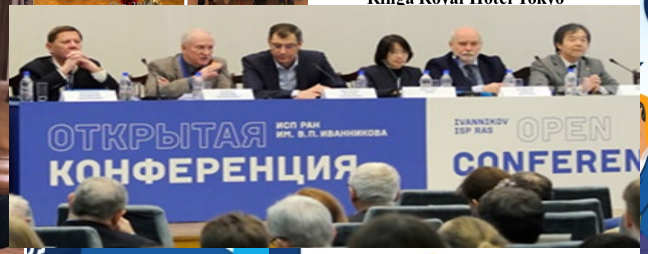
IEEE Computer Society

Prof. Hironori Kasahara is the first President from outside North America in 75 years IEEE Computer Society History

Bjarne Stroustrup: Morgan Stanley & Columbia Univ.
2018 IEEE Computer Society Computer Pioneer Award
IEEE COMPSAC2018 Keynote & Award Ceremony

July 26, 2018, Keynote, Hitotsubashi Hall

July 25, 2018 Award Ceremony Rihga Royal Hotel Tokyo



215 International Conferences

12 Magazines

35 Journals

47 Total Publications

12,000+ Volunteers

615 Committees/ Boards

847,000+ Articles in CSDL

2,352+ Meetings/ Teleconferences

6 New Standards

230 Active Standards

IEEE754, 802

373,100+ Community Members

168 Countries with CS Members

634 Chapters

ACM/IEEE SC (SuperComputing) 19, Denver, Nov.17-22, 2019

Cornel Univ. Prof. Steven Squyres: Mars Exploration, Caltech. Dr. Katie Bouman:Visualization of Blackhole

USA President:
Bill Clinton

USA Ambassador:
Caroline Kennedy

Chinese President:
Hu Jintao

WASEDA University

Tokyo, Japan

Microsoft:
Dr. Bill Gates

British Prime Minister:
Boris Johnson

1882 》
Okuma Shigenobu
founded Tokyo
Senmon Gakko
(College)

5

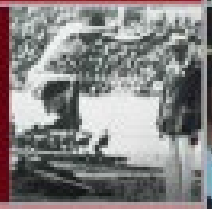
The "Group of Four" who contributed to
the development of Waseda University

1903 》
Start of the Waseda-Keio baseball match (Jubilee)

早稲田大学

《 1922

Archaeological excavation of
the Amarna site



1928 》
Japan's first gold medalist

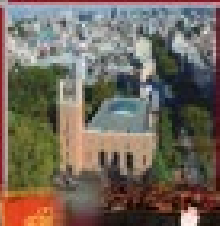
1993 》
Visit to Waseda University
by then U.S. President
Bill Clinton

《 1940
"Visas for life"
from diplomat Chiune Sugihara

1956 》
The beginnings of the Ishibashi Cabinet, first
alumnus of Waseda to become Prime Minister

2007 》

《 1962



《 2012

Alibaba
Mr. Jack Ma

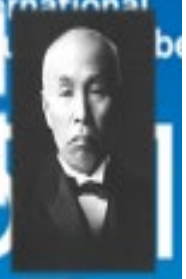
WASEDA UNIVERSITY

Waseda University
早稲田大学

Number of International Students
7,942 from 125 countries and territories
(Undergraduate and Graduate)

Alumni CEOs in Japan
10,606

9 Prime Ministers
Founder Shigenobu OKUMA



Graduate Employability
#1 in private university of Japan
(#2 in Japan, #27 in the world)
QS Graduate Employability Rankings 2019

0% Schools
School Ranking (2018)

ALUMNI [卒業生]	PARTNER INSTITUTIONS [協定大学・機関]
630,000	848 (93 countries)

Hiroshi YAMAUCHI | Masaru IBUKA | Tadashi YANAI

https://www.kantei.go.jp/jp/rekidainaikaku/index.html

mercari | HEROZ JAPAN

FACULTY [教員]	ENROLLMENT [学生数]	UNDERGRADUATE STUDENTS [学部生]	GRADUATE STUDENTS [大学院生]
5,468	49,436	41,051	8,385

* Entire Year 2018

NUMBER OF BOOKS [図書館蔵書]
5,800,000

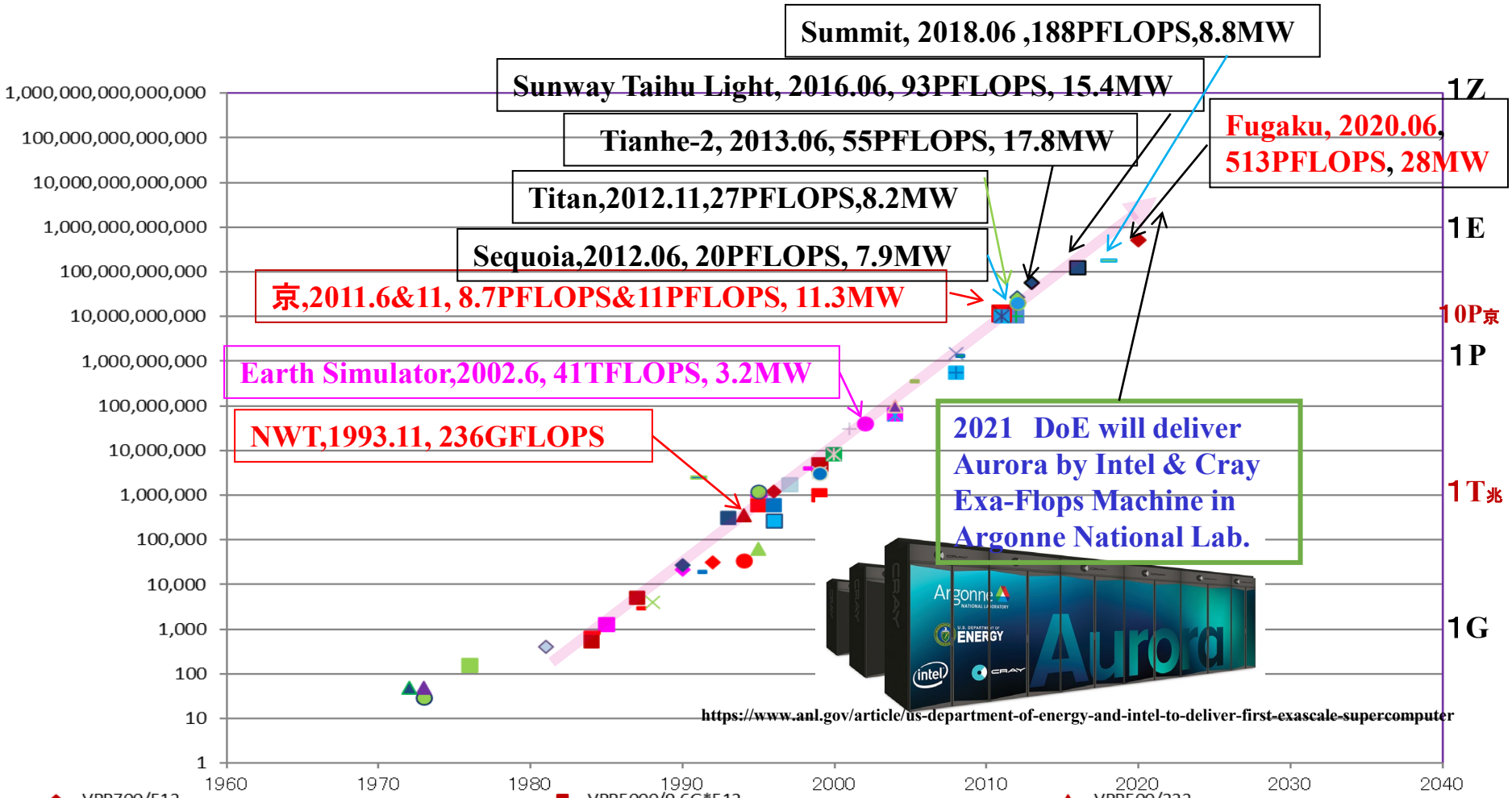
Prime Ministers	Business Leaders	Business Leaders
	Founders of global companies	CEOs of global companies
8th Shigenobu Okuma		ANA (All Nippon Airways)
17th Shigenobu Okuma	Sony	HONDA
66th Tanzan Ishibashi	Samsung	Nintendo
74th Noboru Takeshita	Casio	UNIQLO
76th Toshiki Kaifu	LOTTE	Shiseido
84th Keizo Obuchi		Nomura Securities Co., Ltd.
85th Yoshiro Mori		Tokio Marine & Nichido Fire Insurance Co., Ltd.
91st Yasuo Fukuda		Olympus Corporation
96th Yoshihiko Noda		
100th Fumio Kishida		

Aiji TANAKA
President International Political Science Association (IPSA) President 2016

Hironori KASAHARA
Senior Executive Vice President IEEE Computer Society President 2018. The first president from outside USA and Canada in 72 years CS history. CS has 84,000 members from 168 countries.

Toshio FUKUDA
The University Professor Waseda. Waseda Alumnus. Prof. Emeritus Nagoya Univ., Prof. Meijo Univ. IEEE President 2020. The first from Asia in 135 years history. IEEE has 420,000 members.

Trend of Peak Performances of Supercomputers



https://www.top500.org/

Nov. 2020

TOP 500 The List.

PRESENTED BY ICL INNOVATIVE COMPUTING LABORATORY; BERKELEY LAB Lawrence Berkeley National Laboratory; ISC GROUP Moving Forward.

FIND OUT MORE AT top500.org

NOVEMBER 2020	SYSTEM	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
2	Summit	IBM POWER9 (22C, 3.07GHz), NVIDIA Volta GV100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/SC/ORNL	USA	2,414,592	148.6	10.1
3	Sierra	IBM POWER9 (22C, 3.1GHz), NVIDIA Tesla V100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/NNSA/LLNL	USA	1,572,480	94.6	7.44
4	Sunway TaihuLight	Shenwei SW26010 (260C, 1.45 GHz) Custom Interconnect	NSCC in Wuxi	China	10,649,600	93.0	15.4
5	Selene	NVIDIA DGX A100, AMD EPYC 7742 (64C, 2.25GHz), NVIDIA A100, Mellanox HDR Infiniband	NVIDIA Corporation	USA	555,520	63.4	2.65

https://japanese.engadget.com/arm-super-computer-fugaku-top-500-034015910.html

No. 1 since June 2020

Supercomputer Fugaku: A64FX 48C 2.2GHz, Tofu interconnect D,
RIKEN Center for Computational Science, Fujitsu
Cores:7,299,072; Memory:4,866,048GB;
Processor:A64FX 48C 2.2GHz （arm based）
Interconnect: Tofu interconnect D
Linpack (Rmax)415,530 TFlop/s;
Theoretical Peak (Rpeak)513,855 TFlop/s : 513PFLOPS
HPCG [TFlop/s]13,366.4; Power: 28,334.50 kW(Submitted)

A Strategic Initiative of Computing: Systems and Applications (SISA)- Integrating HPC, Big Data, AI and Beyond, Jan.18-19, 2017 @Waseda Univ. Green Computing Center

A Strategic Initiative of Computing:Systems and Applications

(SISA) --Integrating HPC, Big Data, AI and Beyond-- Jan. 18-19, 2017

Opening: Prof. Gao, Prof. Kasahara

Waseda VP Shuji Hashimoto

I. Architecture and Applications

Keynote: William J. Dally,
NVIDIA and Stanford University, USA

- Kimihiko Hirao, RIKEN, Japan
- G. W. Yang, Tsinghua Univ. China
- J. Sexton, IBM, USA

II . System Software and Applications

Keynote : Rick. Stevens ANL. USA

- S. Mikhail Smelyanskiy Intel USA
- Fred. Streitz, LLNL USA
- R. Govind, IIS, India
- H. Hironori Kasahara, Waseda Univ,

III. Extreme Scale and Beyond

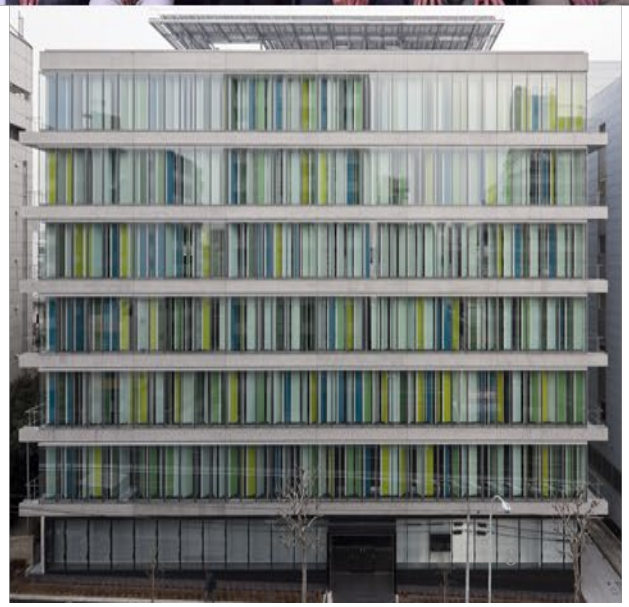
Keynote: Paul Messina ANL, USA

- Motoaki Saito, PEZY, Japan
- Eiji Ishida, MEXT, Japan
- Depei Qian, BUAA, China
- Toshiyuki Shimizu, Fujitsu, Japan

IV. Integration of HPC, Big Data, and AI

Keynote: Thomas Sterling, Indiana Univ., USA

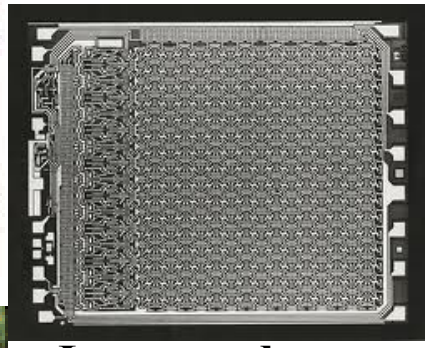
- Masaru Kitsuregawa, NII and Univ. of Tokyo, Japan
- Thomas Schulthess, ETH, Swiss
- Moriyuki Takamura/Toshiaki Kitamura, Oscar Tech, Japan



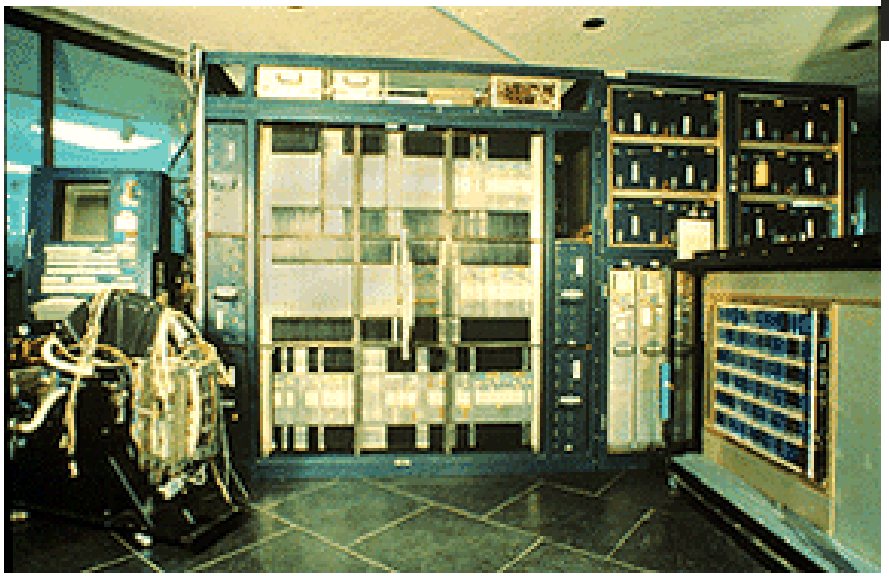
ILLIAC IV, Univ. Illinois at Urbana-Champaign & Burroughs

SIMD
64 Processor Element, Processor Array

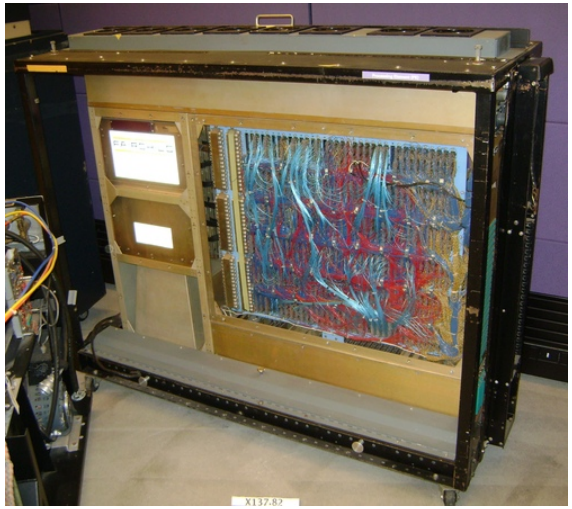
1972-3, NASA
200 MIPS,
300 MOPS,
1 billion bits per second of I/O transfer



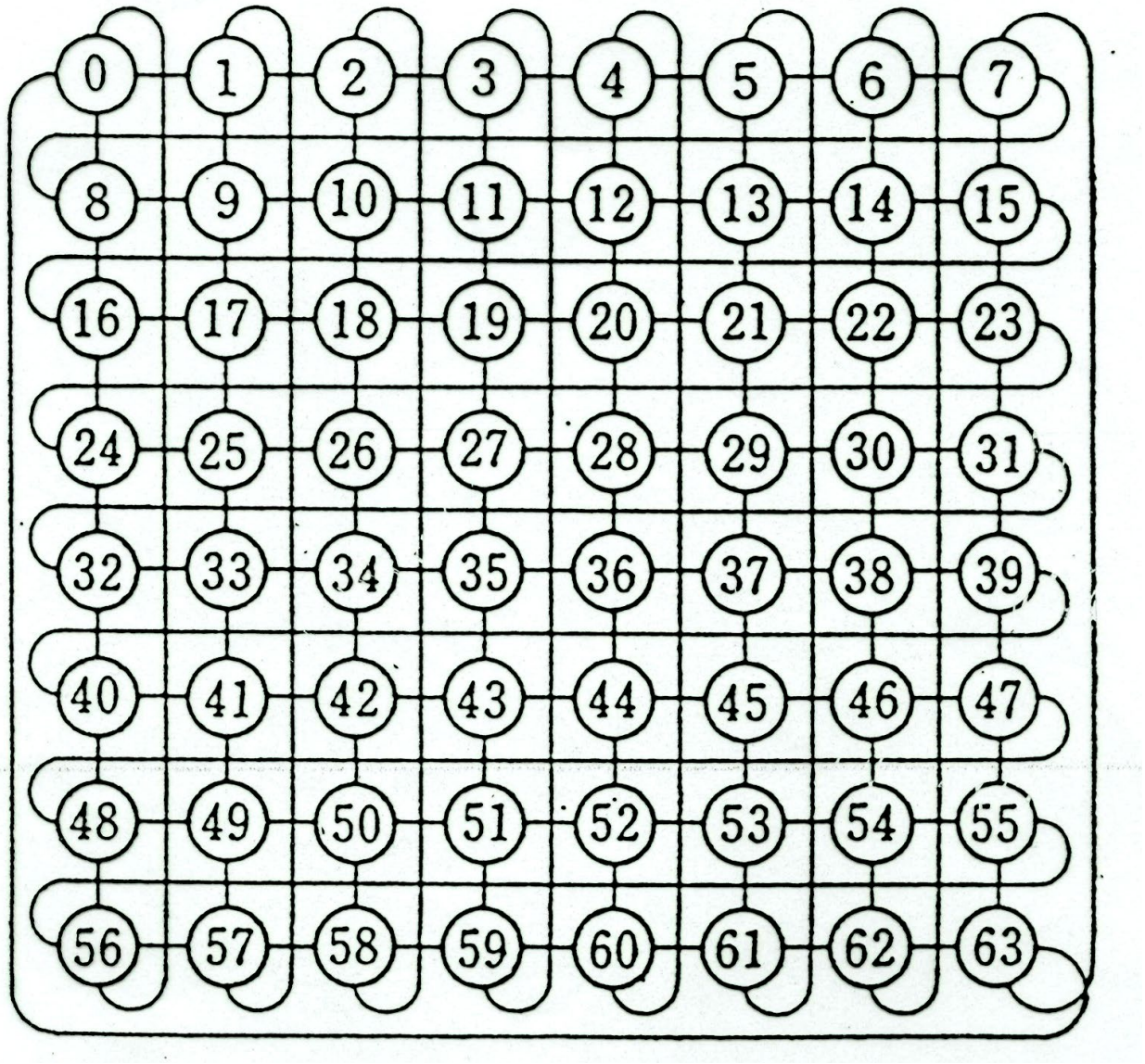
Integrated circuits



Processing Element

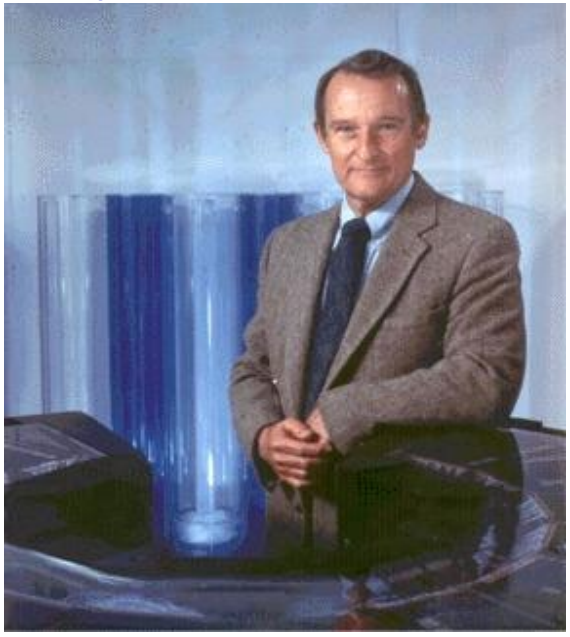


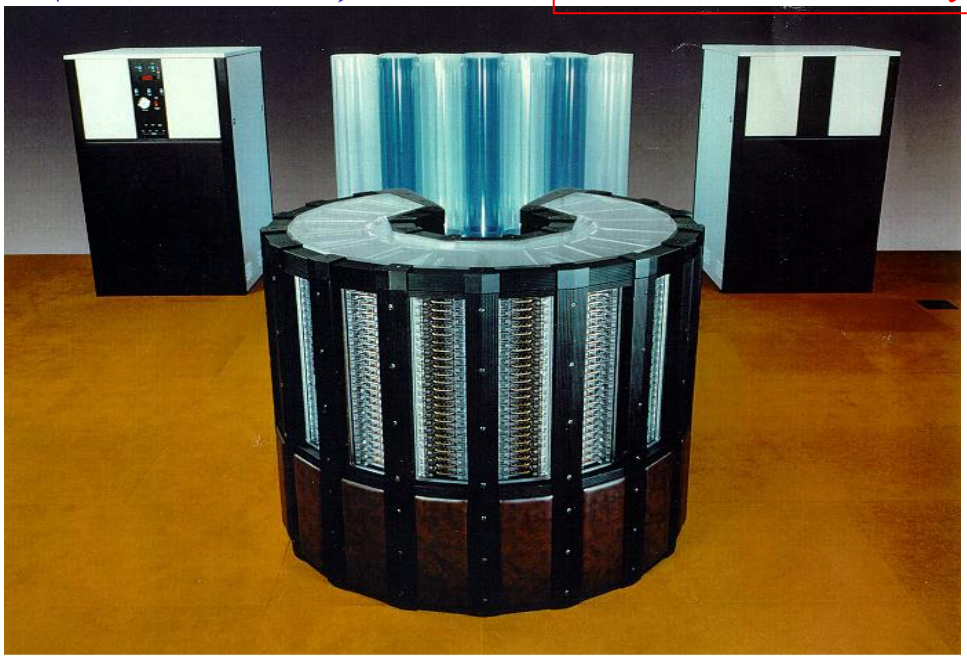
ILLIAC IV Inter-PE Connection Network



Seymour Cray (1925-1996)

IEEE Computer Pioneer Award,
IEEE Eckert-Mauchly Award



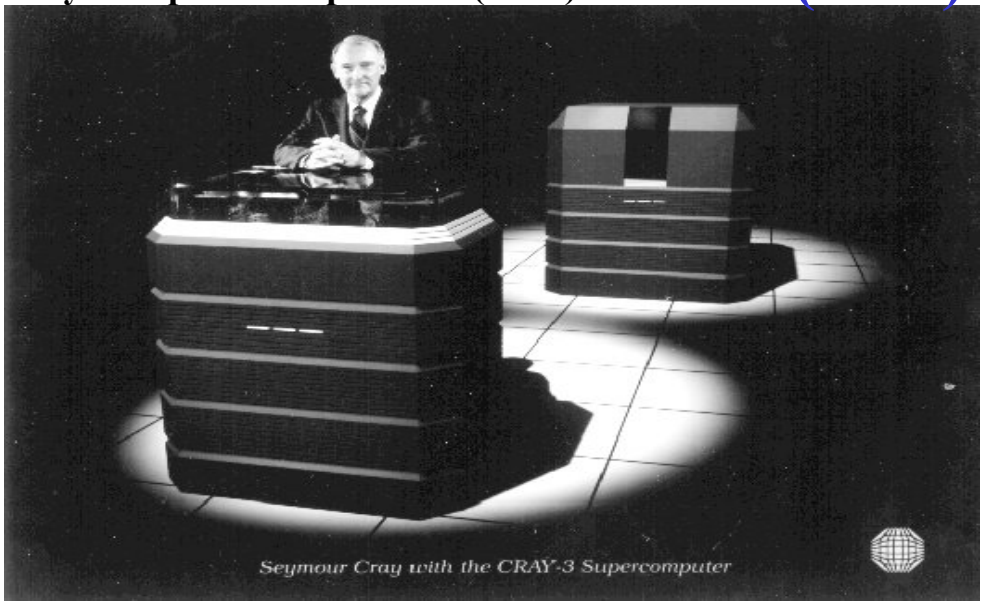


CRAY-2 (1985) 1.9 GFLOPS

CRAY-1 (1975) 160 MFLOPS

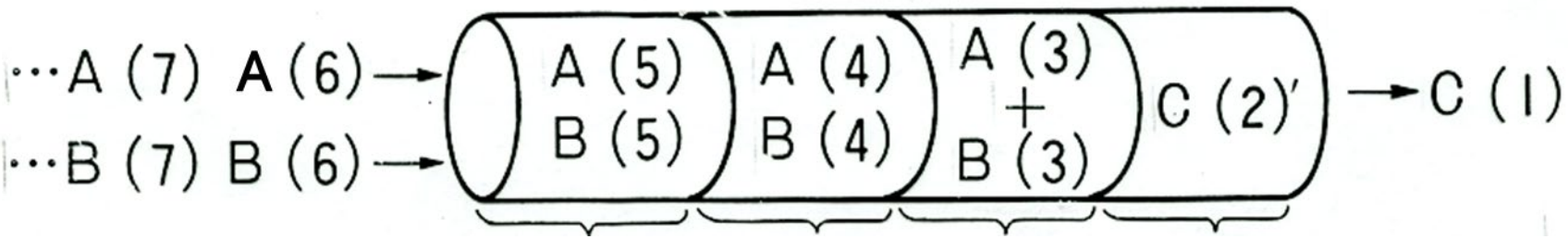
Cray Computer Corporation (CCC) 1988-1995

CRAY-3 (1993)



Vector Pipeline

Floating Point Addition



Comparison of exponent	Sifting of mantissa	Addition	Normalization

Some of papers in and just after Ph.D. Course in Waseda U.

IEEE TRANSACTIONS ON COMPUTERS, VOL. C-33, NO. 11, NOVEMBER 1984 1023

Practical Multiprocessor Scheduling Algorithms for Efficient Parallel Processing

HIRONORI KASAHARA, MEMBER, IEEE, AND SEINOSUKE NARITA, SENIOR MEMBER, IEEE

Courtesy of dexchao - Fotolia.com

104 IEEE JOURNAL OF ROBOTICS AND AUTOMATION, VOL. RA-1, NO. 2, JUNE 1985

Parallel Processing of Robot-Arm Control Computation on a Multimicroprocessor System

HIRONORI KASAHARA MEMBER, IEEE, AND SEINOSUKE NARITA, SENIOR MEMBER, IEEE

2nd International Conference on Superecomputing
Santa Clara, CA, USA May 3-8,1987

A PARALLEL PROCESSING SCHEME FOR THE SOLUTION OF SPARSE LINEAR EQUATIONS USING STATIC OPTIMAL-MULTIPROCESSOR-SCHEDULING ALGORITHMS

H.Kasahara*, T.Fujii*, H.Nakayama*, S.Narita*, and Leon O.Chua**

* Dept. of Electrical Eng., Waseda University, Tokyo, 160, Japan
** Dept. of Electrical Eng. and Computer Sciences, University of California, Berkeley, CA 94720, U.S.A.

PARALLEL PROCESSING OF ROBOT MOTION SIMULATION

Young Author Prize (YAP)
This prize, created in 1986, is awarded at IFAC World

H. Kasahara, H. Fujii and M. Iwata

Department of Electrical Engineering, Waseda University, 3-4-1 Ohkubo Shinjuku-ku, Tokyo 160, Japan

Research on OSCAR Parallelizing Compiler & Co-designed Hardware Since 1984

66 international patents in USA, UK, China, Japan to improve effective performance, cost-performance, software productivity and power efficiency

High Performance & Low Power

1) Multigrain Parallelization for Embedded to HPC Homogeneous and Heterogeneous Multicores

Coarse-grain task parallelism among loops, subroutines & basic blocks in addition to the loop parallelism

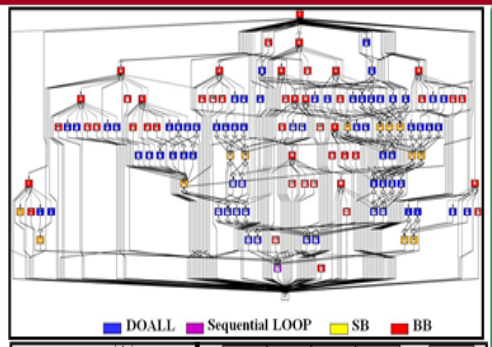
2) Data Localization:
Optimization of Cache & Local Memory Usage

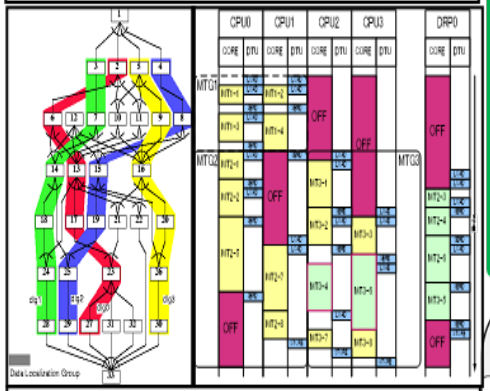
- Automatic data decomposition and data reuse control for Distributed shared memory, Cache and Local memory
- Data Transfer Control
  - Overlapping Data Transfer using Data Transfer Unit, or DMAC

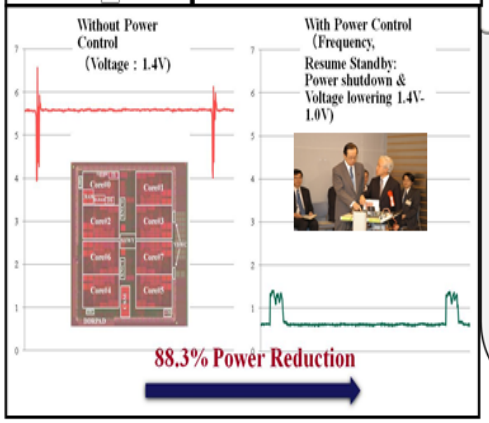
3) Automatic Power Reduction

- OSCAR Compiler can reduce power consumption by using DVFS and Clock- & Power- gating with hardware supports.

4) Codesigned Accelerator: See right figure



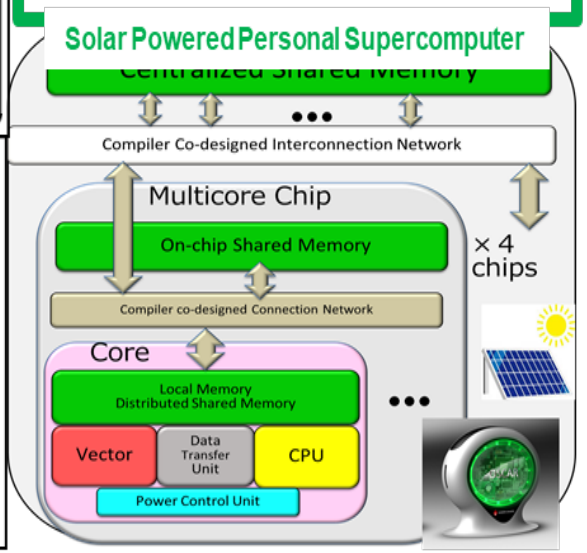




Green Accelerator

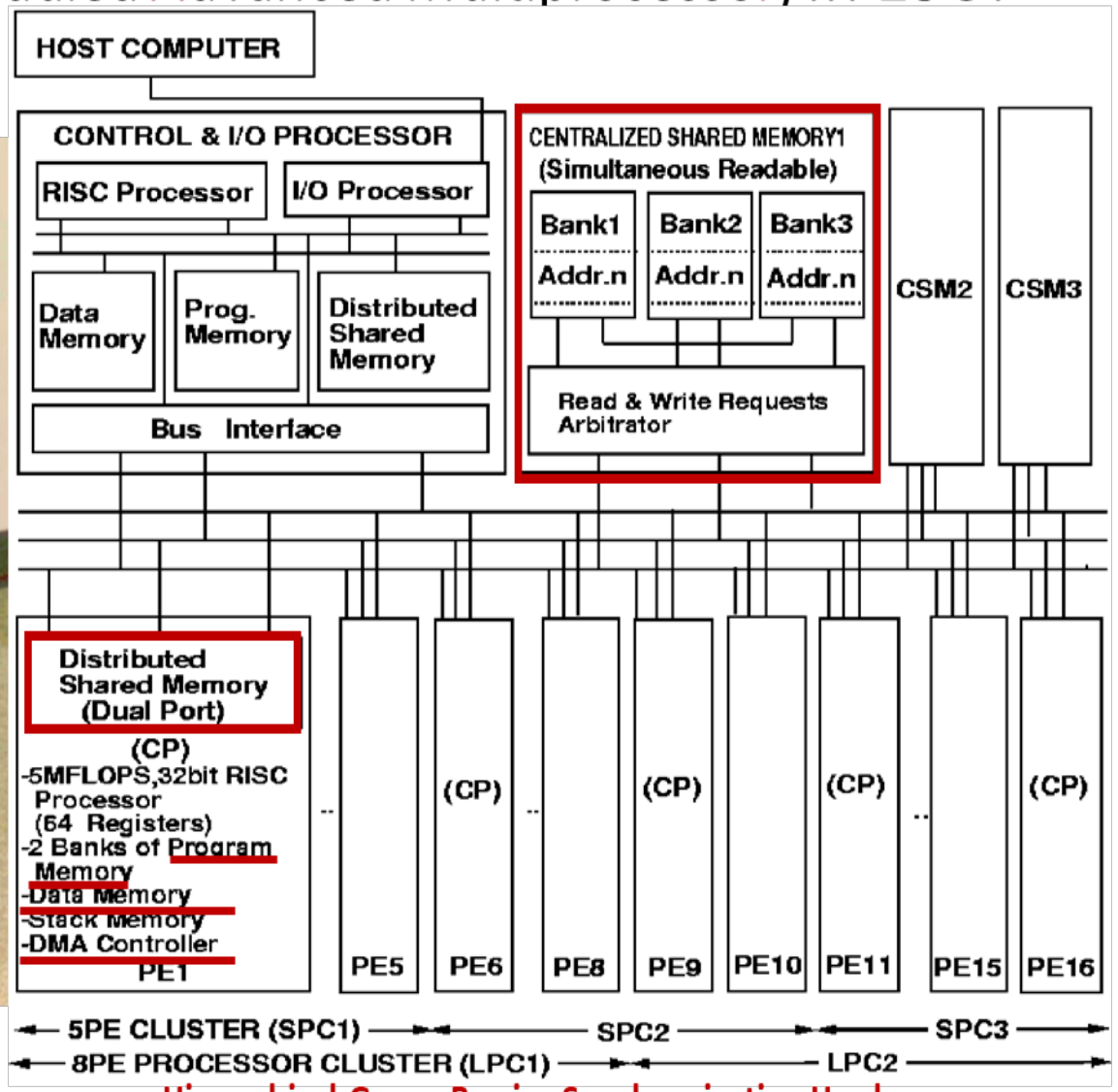
can be attached to any processor cores, RISCV, arm, Infineon Renesas, AMD, Intel, etc, without instruction extensions.

➢It works with automatic local memory management and power reduction control by OSCAR Automatic vectorizing & parallelizing compiler



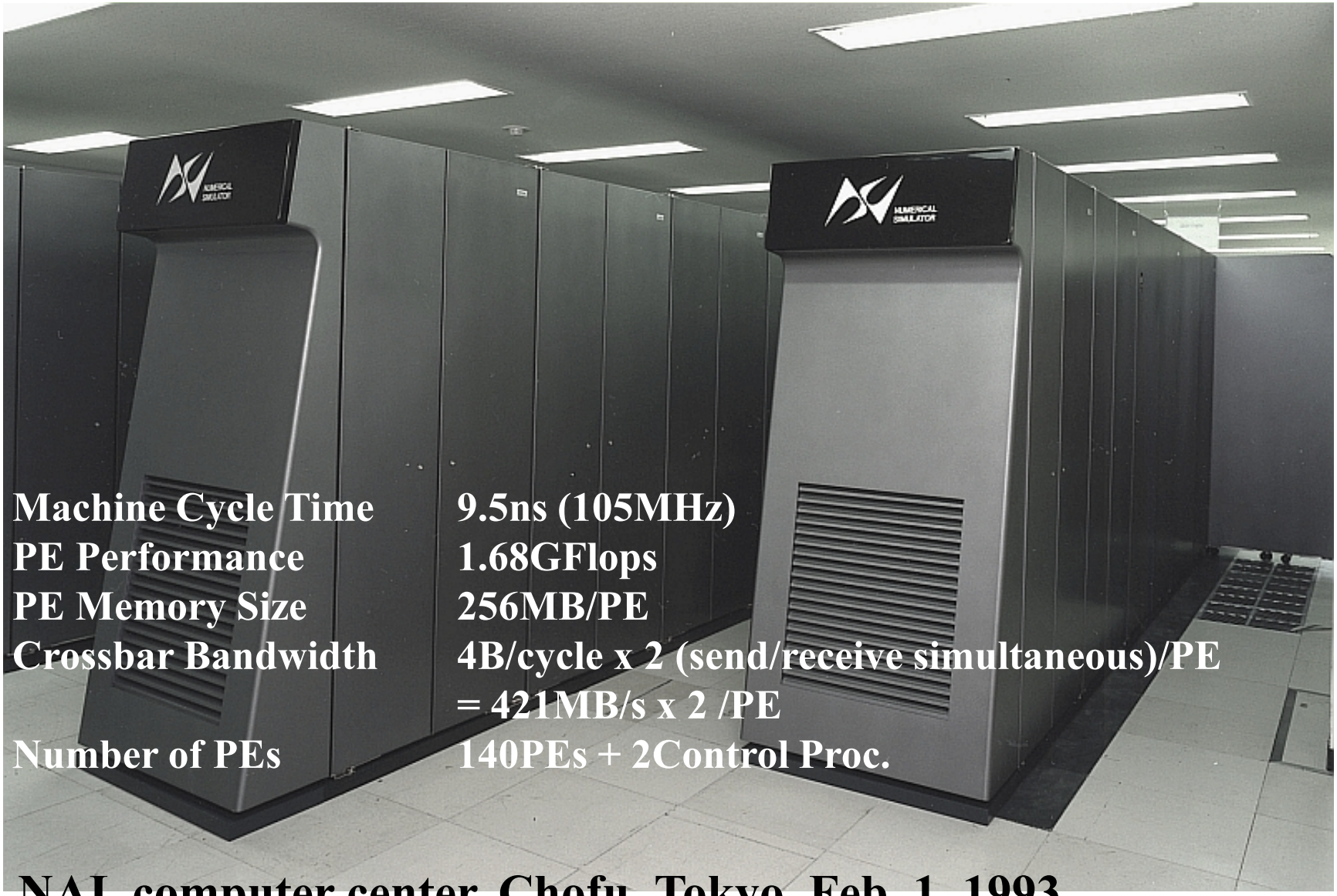
The First Codesigned Architecture with Compiler

OSCAR (Optimally Scheduled Advanced Multiprocessor) in 1987



Hierarchical Group Barrier Synchronization Hardware

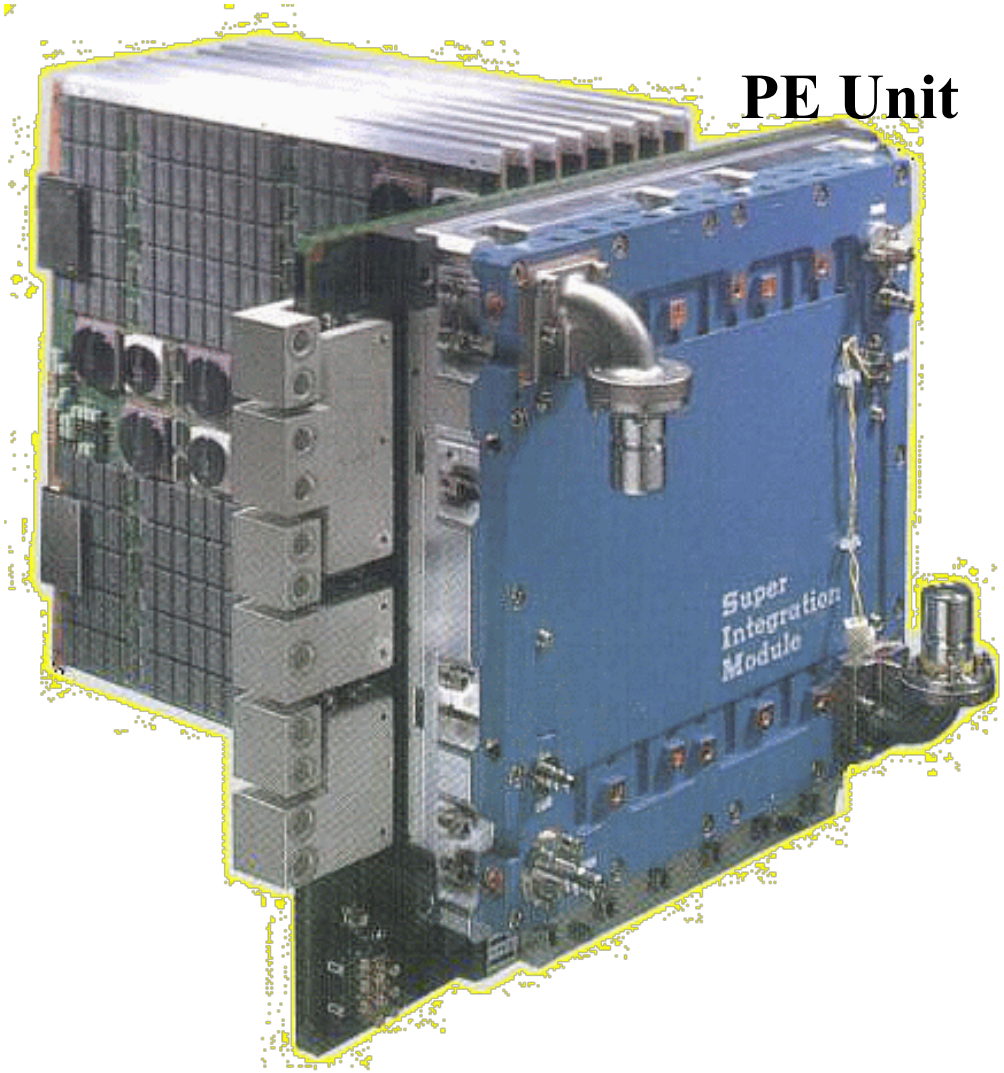
NWT (Numerical Window Tunnel) in 1993



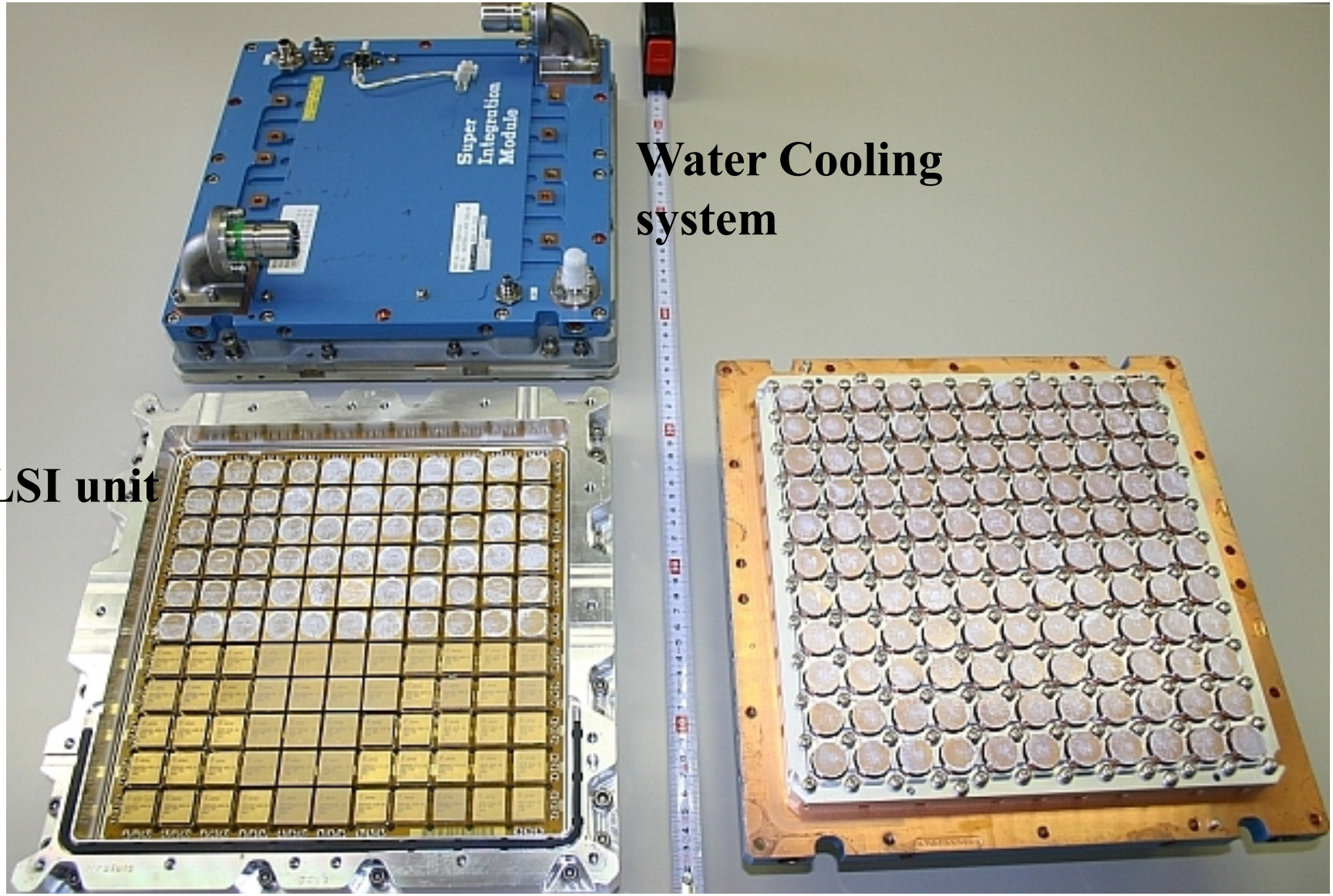
Machine Cycle Time	9.5ns (105MHz)
PE Performance	1.68GFlops
PE Memory Size	256MB/PE
Crossbar Bandwidth	4B/cycle x 2 (send/receive simultaneous)/PE = 421MB/s x 2 /PE
Number of PEs	140PEs + 2Control Proc.

NAL computer center, Chofu, Tokyo, Feb. 1, 1993

VPP500/NWT



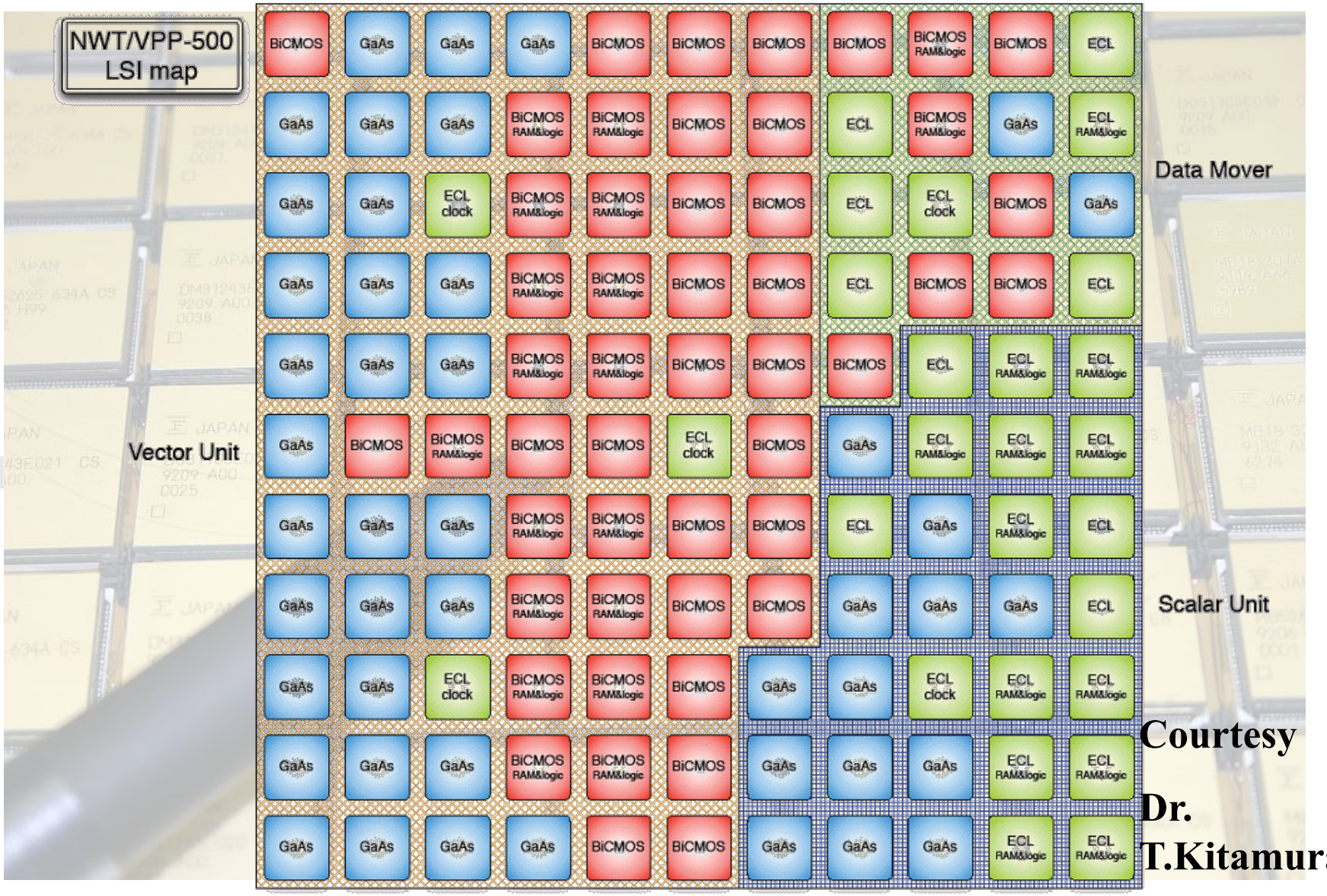
VPP500/NWT



Water Cooling system

PE LSI unit

VPP500/NWT


NWT/VPP-500
LSI map
Vector Unit
Data Mover
Scalar Unit
Courtesy
Dr.
T.Kitamura

Earth Simulator

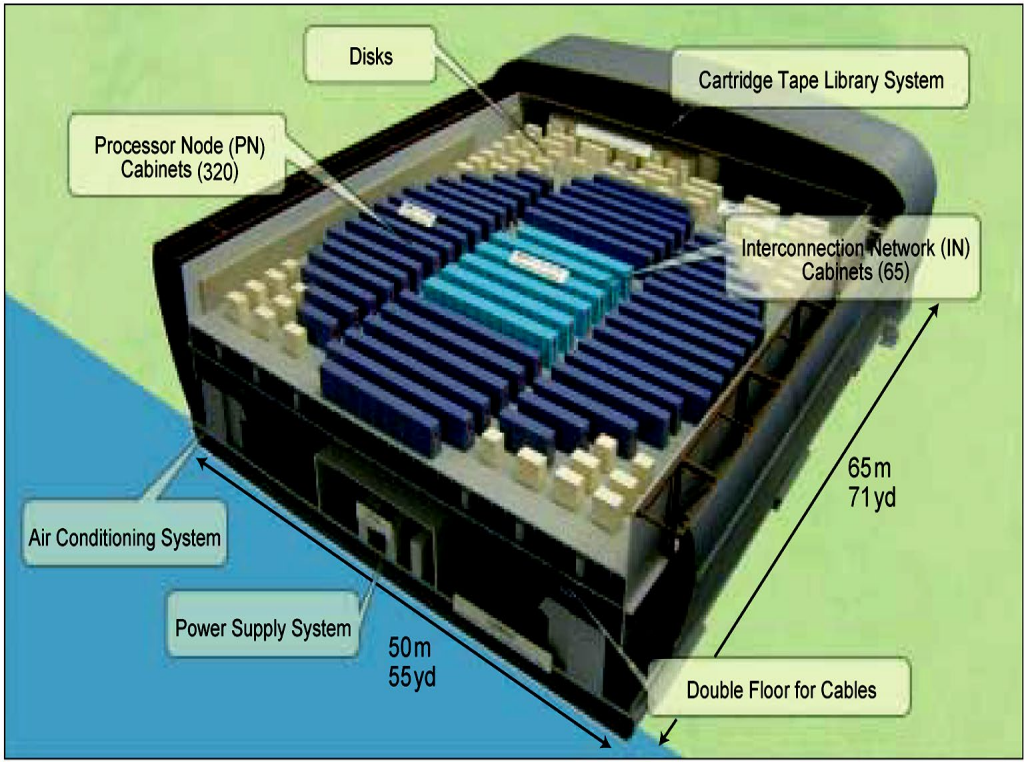
(http://www.es.jamstec.go.jp/)

- Earth Environmental simulation like Global Warming, El Nino, PlateMovement for the all lives onr this planet.
- Developed in Mar. 2002 by STA (MEXT) and NEC with 400 M$ investment under Dr. Miyoshi's direction.
  (Dr.Miyoshi: Passed away in Nov.2001. NWT, VPP500, SX6)

Mr. Hajime Miyoshi

Image of Earth Simulator

4 Tennis Courts

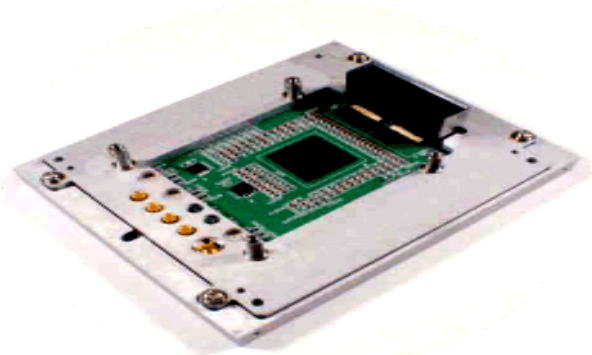


40 TFLOPS Peak ($40*10^{12}$)

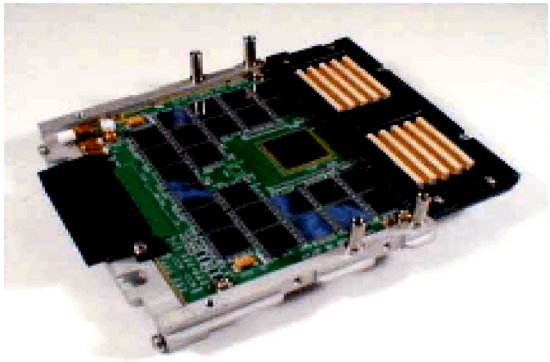
35.6 TFLOPS Linpack

Architecture of Earth Simulator

- 5120 Single Chip Vector Processors: 8processor SMP node * 640
- Inter node crossbar : Bi-Sect. Band Width: 16GB/s*2*320=10TB/s, 1800 miles of copper cable
- Distributed shared memory

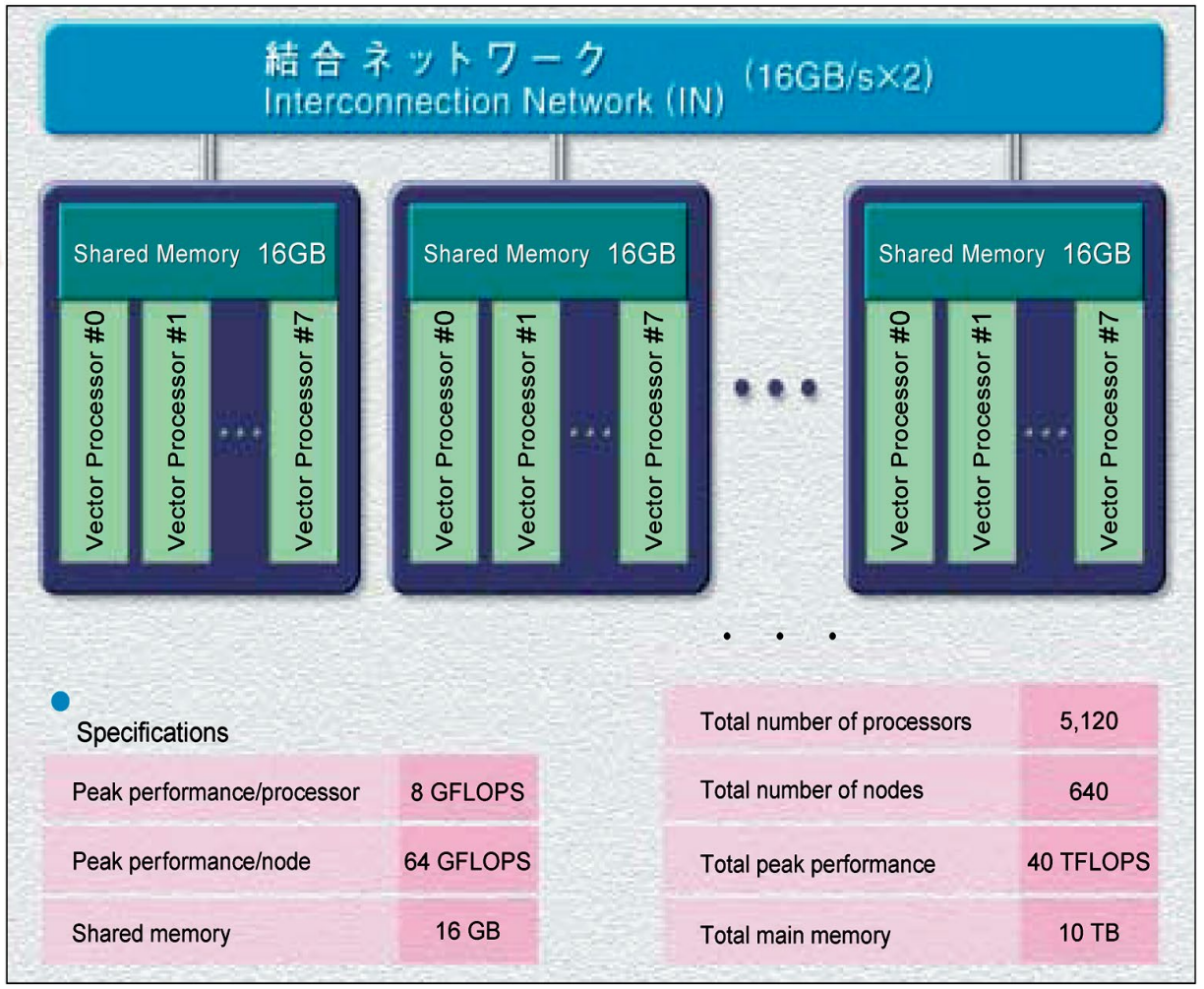


Processor Board (0.15um CMOS)

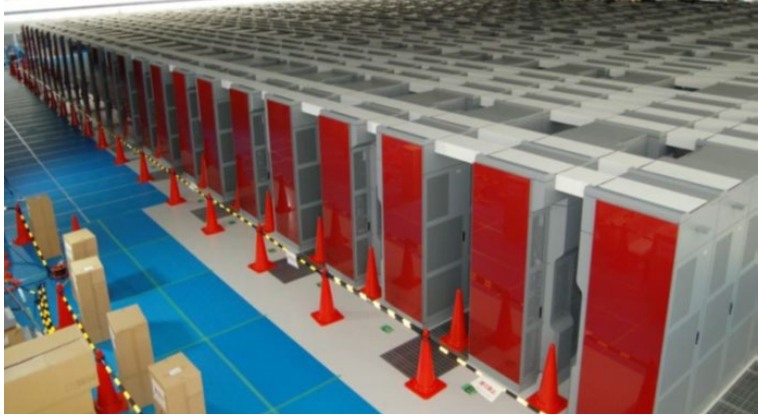


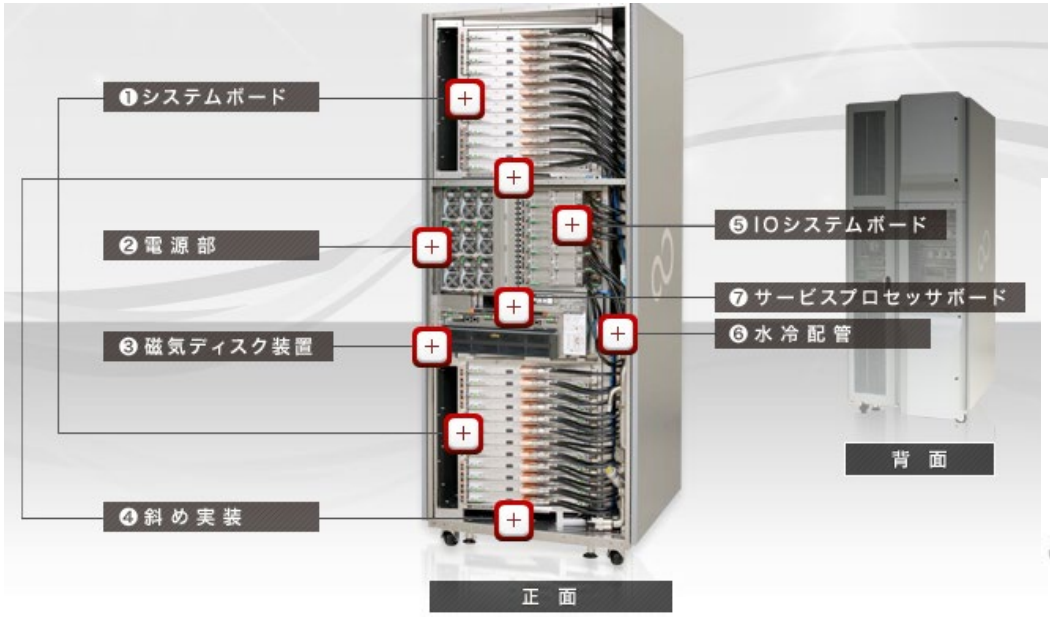
Memory Board (512MB FPLRAM)

Configuration of Earth Simulator

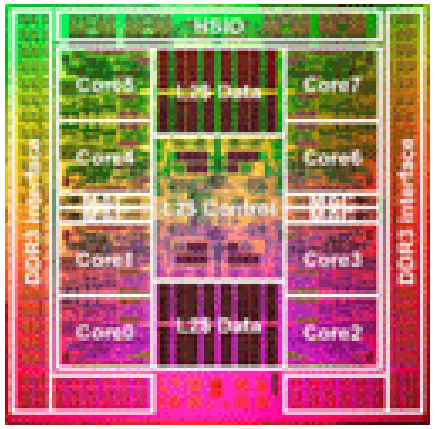


Riken K Supercomputer in 2011 (10PFLOPS: In Japanese 10P pronounced “K”)

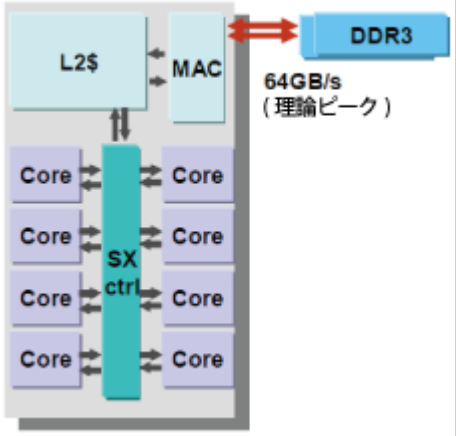


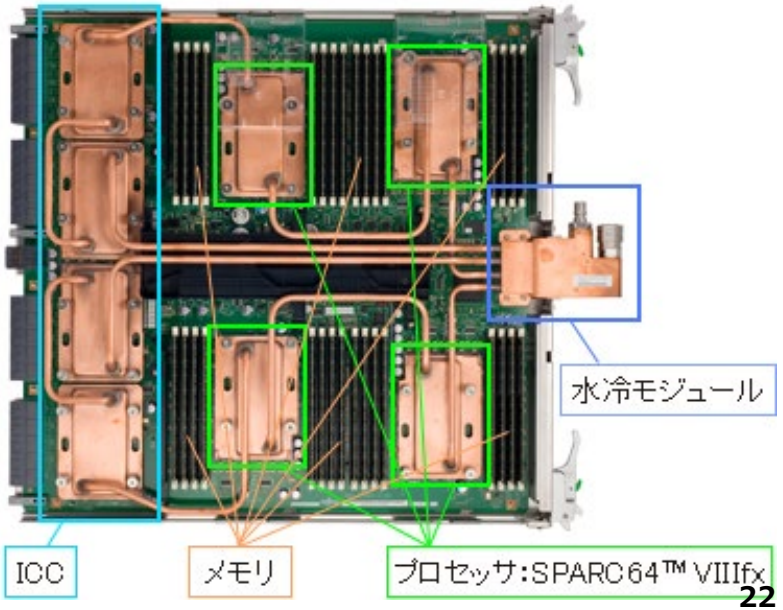


TOFUNetwork (6D Torus)



SPARC64™ VIIIfx
（提供：富士通(株)）





https://www.top500.org/

Nov. 2020

TOP500 The List.

PRESENTED BY: ICL INNOVATIVE COMPUTING LABORATORY; BERKELEY LAB Lawrence Berkeley National Laboratory; ISC GROUP Moving Forward.

FIND OUT MORE AT: top500.org

NOVEMBER 2020	SYSTEM	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
2	Summit	IBM POWER9 (22C, 3.07GHz), NVIDIA Volta GV100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/SC/ORNL	USA	2,414,592	148.6	10.1
3	Sierra	IBM POWER9 (22C, 3.1GHz), NVIDIA Tesla V100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/NNSA/LLNL	USA	1,572,480	94.6	7.44
4	Sunway TaihuLight	Shenwei SW26010 (260C, 1.45 GHz) Custom Interconnect	NSCC in Wuxi	China	10,649,600	93.0	15.4
5	Selene	NVIDIA DGX A100, AMD EPYC 7742 (64C, 2.25GHz), NVIDIA A100, Mellanox HDR Infiniband	NVIDIA Corporation	USA	555,520	63.4	2.65

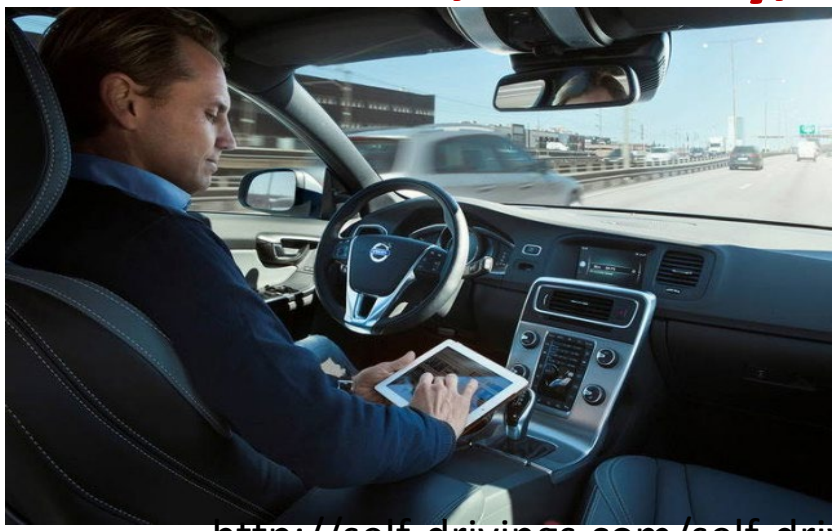
https://japanese.engadget.com/arm-super-computer-fugaku-top-500-034015910.html

No. 1 since June 2020

Supercomputer Fugaku: A64FX 48C 2.2GHz, Tofu interconnect D,
RIKEN Center for Computational Science, Fujitsu
Cores:7,299,072; Memory:4,866,048GB;
Processor:A64FX 48C 2.2GHz （arm based）
Interconnect: Tofu interconnect D
Linpack (Rmax)415,530 TFlop/s;
Theoretical Peak (Rpeak)513,855 TFlop/s : 513PFLOPS
HPCG [TFlop/s]13,366.4; Power: 28,334.50 kW(Submitted)

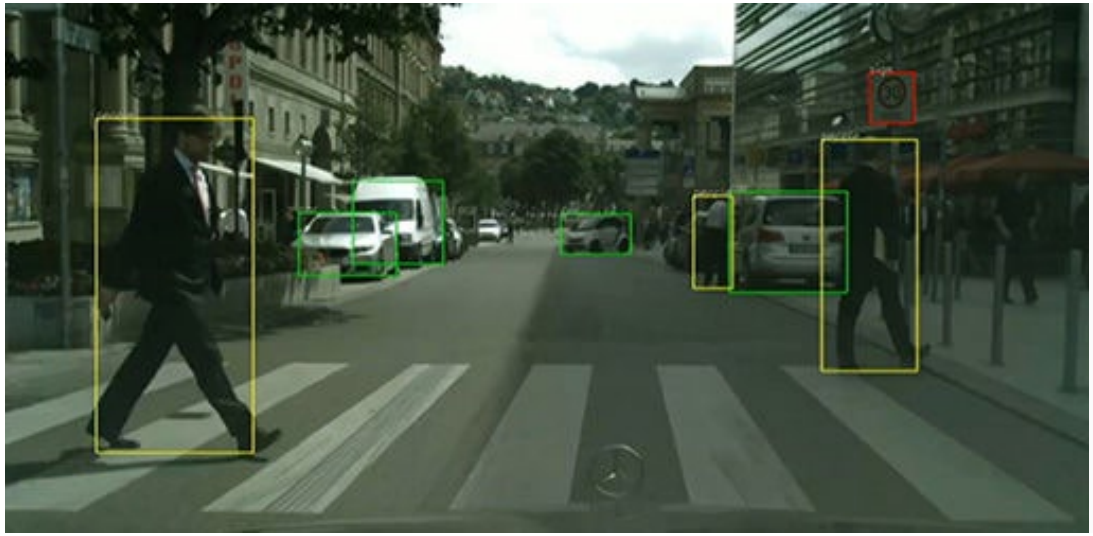
Self Driving Cars (自動運転)

Connected, Security, Big Data, Traffic Cloud



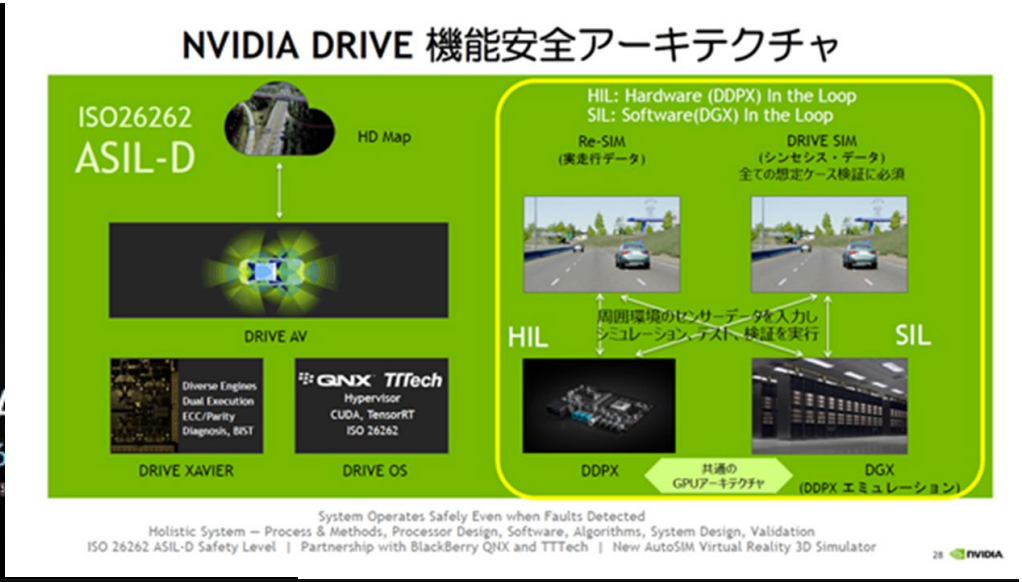
http://self-drivings.com/self-driving-cars-updated-market-analysis/

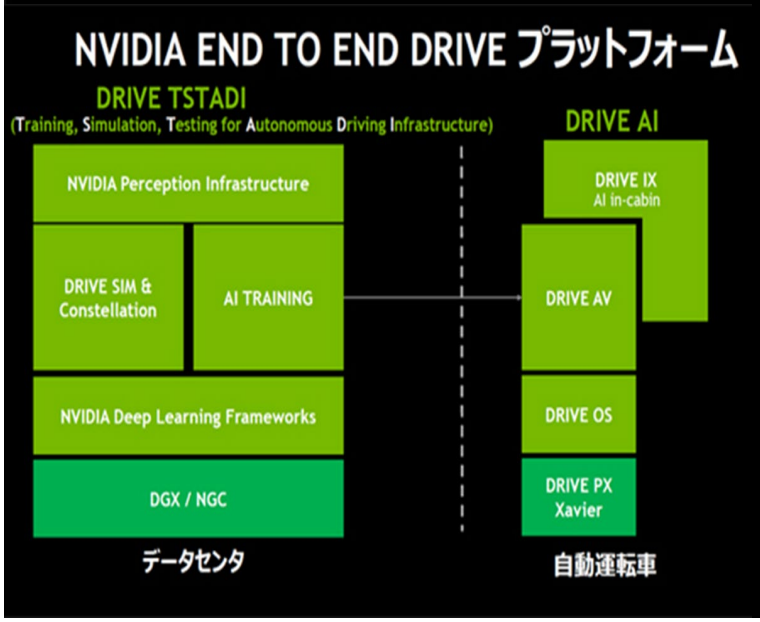
Deep Leaning （多層ニューラルネット）により画像認識

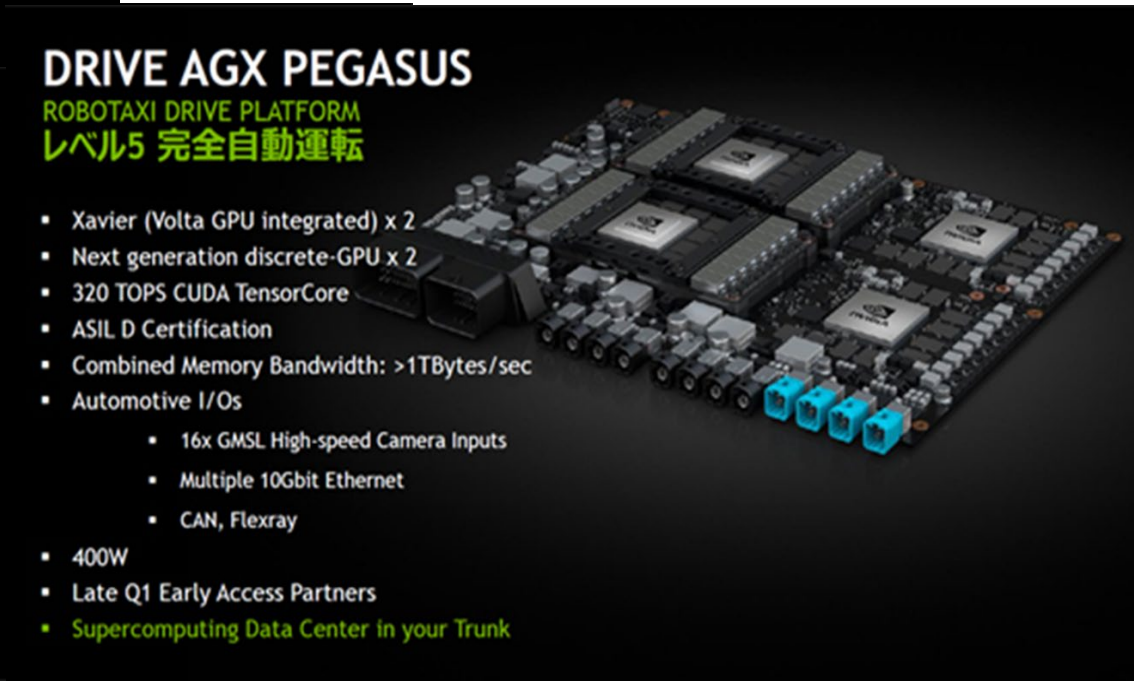


http://www.digitalartsonline.co.uk/news/creative-hardware/nvidias-water-cooled-supercomputer-helps-cars-drive-themselves/

NVIDIA Drive Accelerator　DRIVE PX XAVIER







https://news.mynavi.jp/article/20181003-700546/

Demo of NEDO Green Multicore Processor for Real Time Consumer Electronics at Council of Science and Engineering Policy on April 10, 2008

http://www8.cao.go.jp/cstp/gaiyo/honkaigi/74index.html

第74回総合科学技術会議【平成20年4月10日】


第74回総合科学技術会議の様子(1)

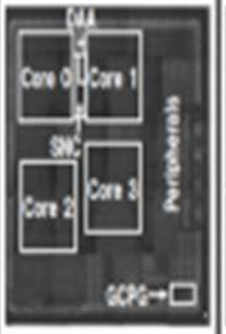
第74回総合科学技術会議の様子(2)


第74回総合科学技術会議の様子(3)


第74回総合科学技術会議の様子(4)

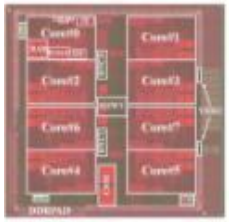
Codesign of Compiler and Multiprocessor Architecture since 1985

4 core multicore RP1 (2007), 8 core multicore RP2 (2008) and 15 core Heterogeneous multicore RPX (2010) developed in NEDO Projects with Hitachi and Renesas

RP-1 (ISSCC2007 #5.3)	RP-2 (ISSCC2008 #4.5)	RP-X (ISSCC2010 #5.3)
		
90nm, 8-layer, triple-Vth, CMOS	90nm, 8-layer, triple-Vth, CMOS	45nm, 8-layer, triple-Vth, CMOS
97.5 mm² (9.88 x 9.88 mm)	104.8 mm² (10.61 x 9.88 mm)	153.8 mm² (12.4 x 12.4 mm)
1.0V (internal), 1.8/3.3V (IO)	1.0-1.4V (internal), 1.8/3.3V (I/O)	1.0-1.2V (internal), 1.2-3.3V (I/O)
600MHz, 4.32 GIPS, 16.8 GFLOPS	600MHz, 8.64 GIPS, 33.6 GFLOPS	648MHz, 13.7GIPS, 115GOPS, 36.2GFLOPS
11.4 GOPS/W (32b換算)	16.3 GOPS/W (32b換算)	37.3 GOPS/W (32b換算)

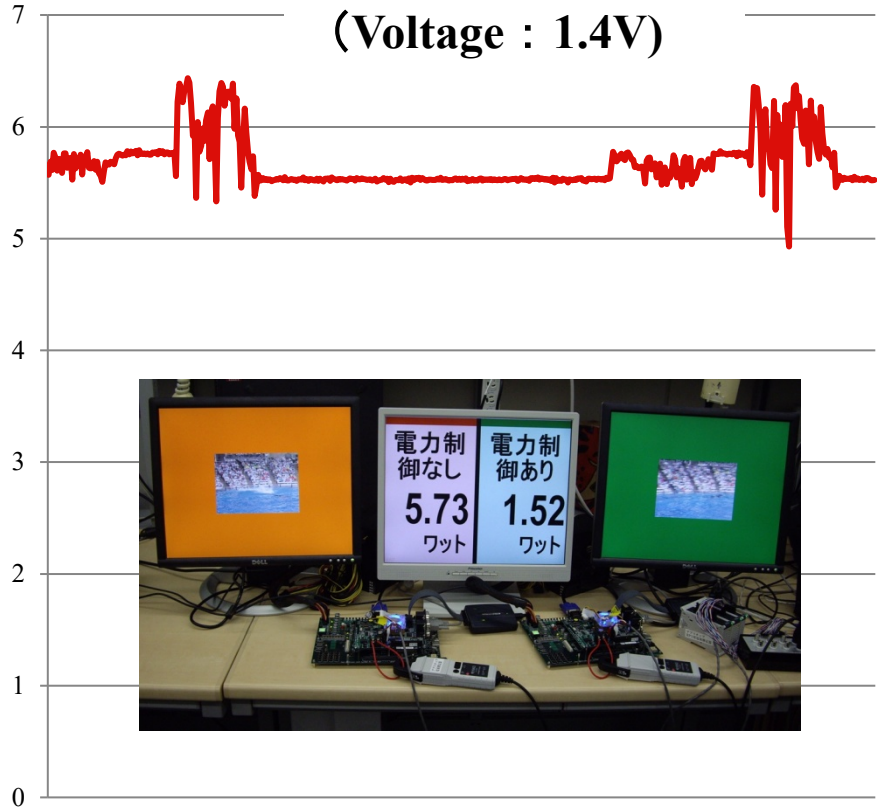
Prime Minister FUKUDA is touching our multicore chip during execution.

Power Reduction of MPEG2 Decoding to 1/4 on 8 Core Homogeneous Multicore RP-2 by OSCAR Parallelizing Compiler

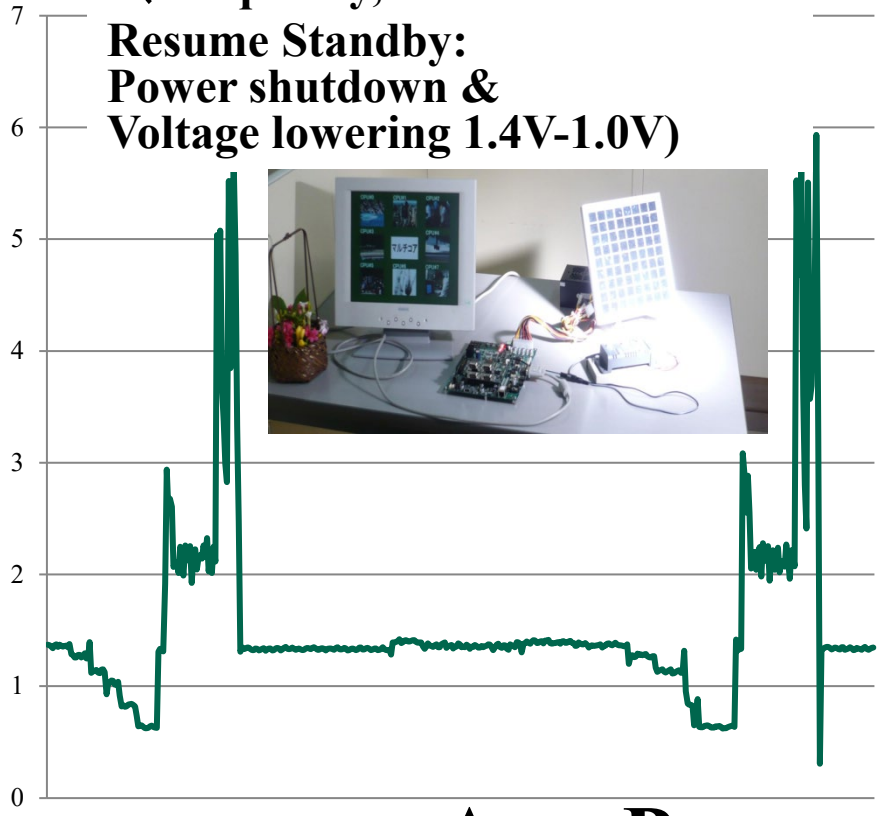


MPEG2 Decoding with 8 CPU cores

Without Power Control (Voltage : 1.4V)



With Power Control (Frequency, Resume Standby: Power shutdown & Voltage lowering 1.4V-1.0V)



Avg. Power 5.73 [W]

73.5% Power Reduction

Avg. Power 1.52 [W]

Green Computing Systems R&D Center

Waseda University

Established by Prof. Kasahara supported by METI (Mar. 2011)

＜R & D Target＞

Hardware, Software, Application for Super Low-Power Manycore

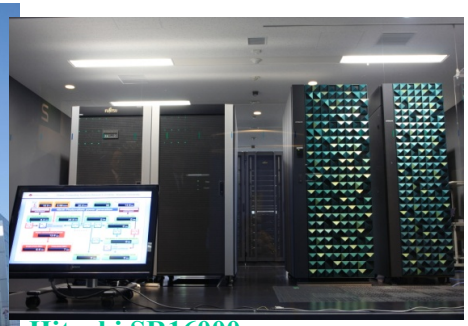
- **More than 64 cores**
- **Natural air cooling (No fan) Cool, Compact, Clear, Quiet**
- **Operational by Solar Panel**

<Industry, Government, Academia>

Hitachi, Fujitsu, NEC, Renesas, Olympus, Toyota, Denso, Mitsubishi, Toshiba, OSCAR Technology, etc

＜Ripple Effect＞

- **Low CO_2 (Carbon Dioxide) Emissions**
- **Creation Value Added Products**
- **Automobiles, Medical, IoT, Servers**



Hitachi SR16000:
Power7 128coreSMP
Fujitsu M9000
SPARC VII 256 core SMP

Beside Subway Waseda Station, Near Waseda Univ. Main Campus

OSCAR Parallelizing Compiler

To improve effective performance, cost-performance and software productivity and reduce power

Multigrain Parallelization(LCPC1991,2001,04)

coarse-grain parallelism among loops and subroutines (2000 on SMP), near fine grain parallelism among statements (1992) in addition to loop parallelism

Data Localization

Automatic data management for distributed shared memory, cache and local memory
(Local Memory 1995, 2016 on RP2,Cache2001,03)
Software Coherent Control (2017)

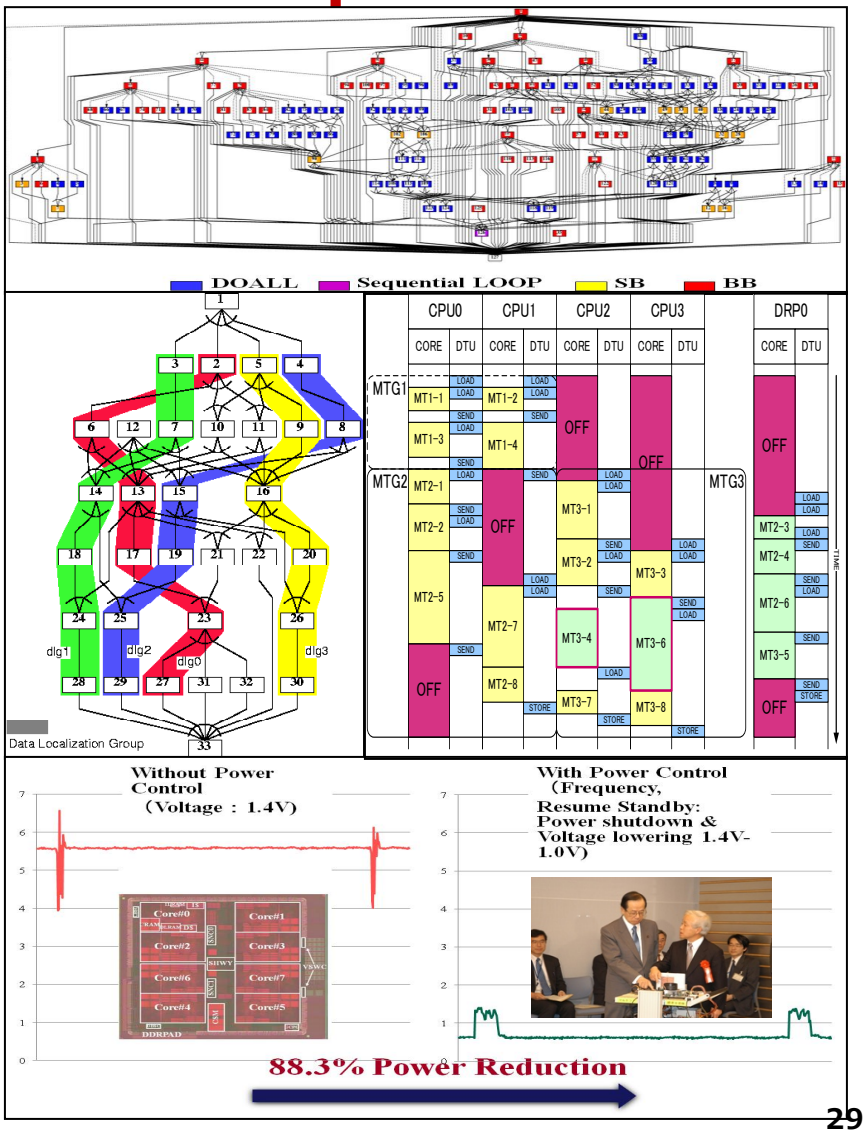
Data Transfer Overlapping(2016 partially)

Data transfer overlapping using Data Transfer Controllers (DMAs)

Power Reduction

(2005 for Multicore, 2011 Multi-processes, 2013 on ARM)

Reduction of consumed power by compiler control DVFS and Power gating with hardware supports.



Heterogeneous Multicore Programing with OSCAR API V2.0

Sequential Application Program in Fortran or C
(Consumer Electronics, Automobiles, Medical, Scientific computation, etc.)

Homogeneous

Hetero

Manual parallelization / power reduction

Accelerator Compiler/ User
Add "hint" directives before a loop or a function to specify it is executable by the accelerator with how many clocks

Waseda OSCAR Parallelizing Compiler
- Multigrain Parallelization including Coarse grain task and Loop parallelization
- Data Localization: Data reuse
- DMAC data transfer
- Power reduction using DVFS, Clock/ Power gating

Hitachi, Renesas, NEC, Fujitsu, Toshiba, Denso, Olympus, Mitsubishi, Esol, Cats, Gaio, 3 univ.

OSCAR API for Homogeneous and/or Heterogeneous Multicores and manycores
Directives for thread generation, memory, data transfer using DMA, power managements

Parallelized API F or C program

Proc0
Code with directives
Thread 0

Proc1
Code with directives
Thread 1

Accelerator 1
Code

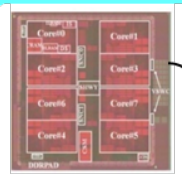
Accelerator 2
Code

Low Power Homogeneous Multicore Code Generation
API Analyzer | Existing sequential compiler

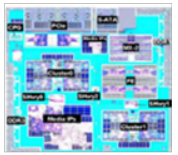
Low Power Heterogeneous Multicore Code Generation
API Analyzer (Available from Waseda) | Sequential compiler for CPU and Acc. or linking ACC Library

Server Code Generation
OpenMP Compiler

Generation of parallel machine codes using sequential compilers



Homegeneous Multicore s from Vendor A (SMP servers)



Heterogeneous Multicores from Vendor B

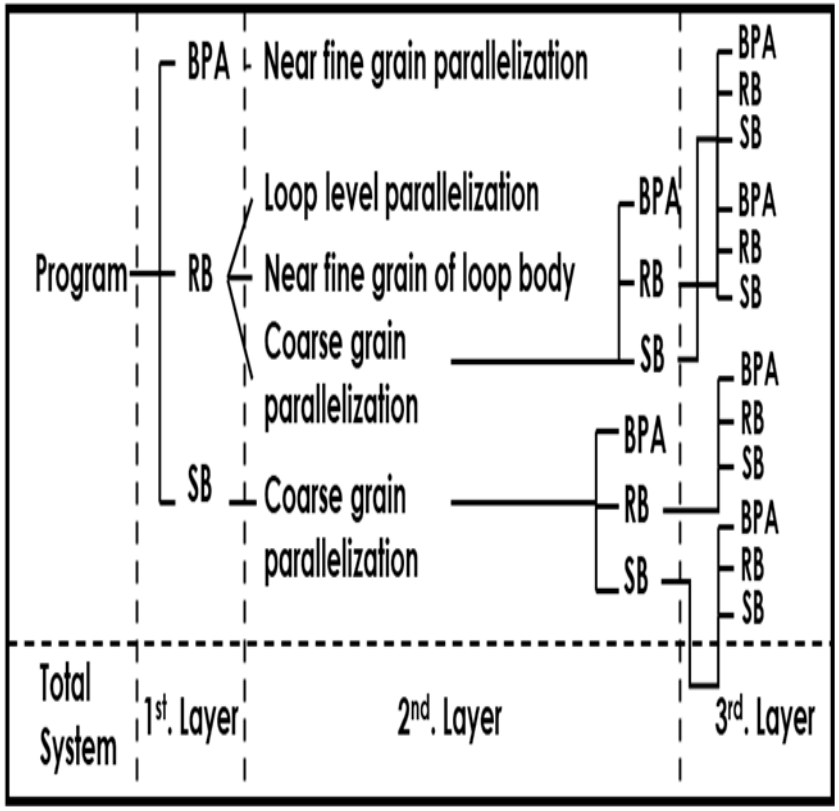
CC-NUMA servers

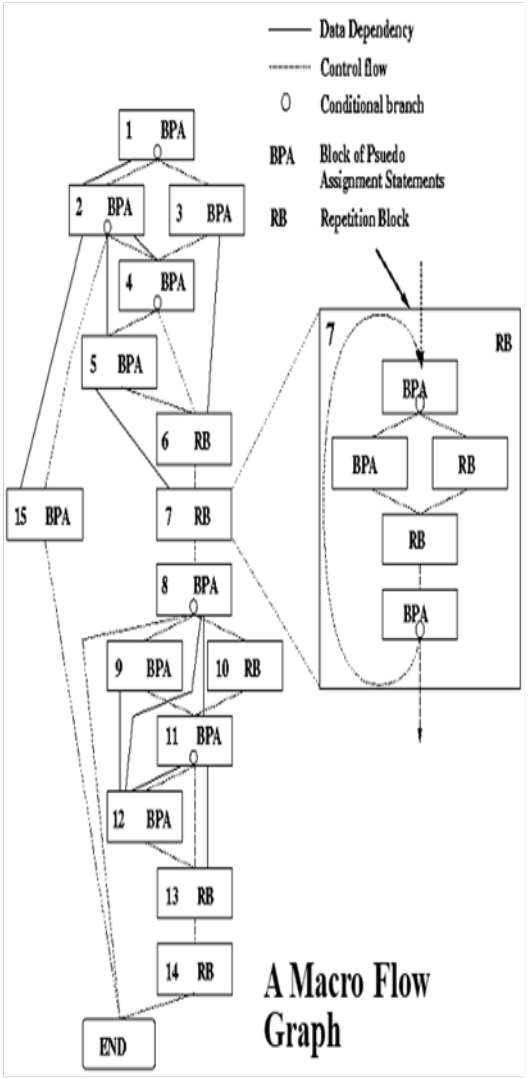
Executable on various multicores

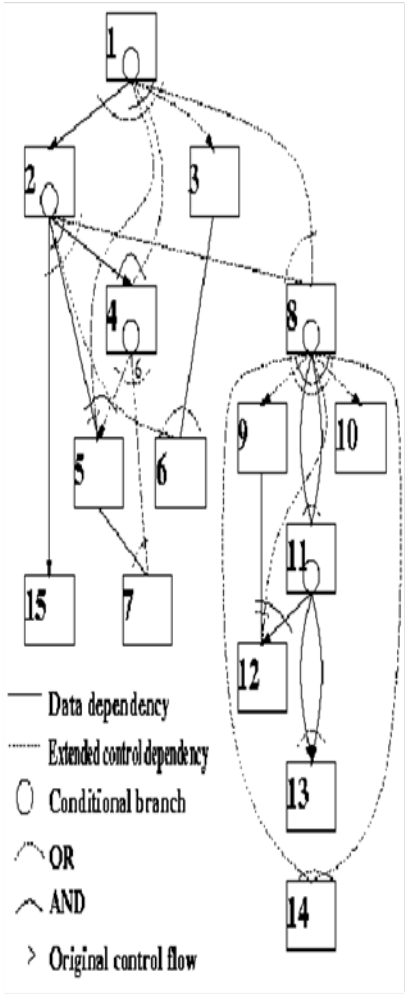
Generation of Coarse Grain Tasks

Earliest Executable Condition Analysis for Coarse Grain Tasks (Macro-tasks)

- Macro-tasks (MTs)
  - Block of Pseudo Assignments (BPA): Basic Block (BB)
  - Repetition Block (RB) : natural loop
  - Subroutine Block (SB): subroutine



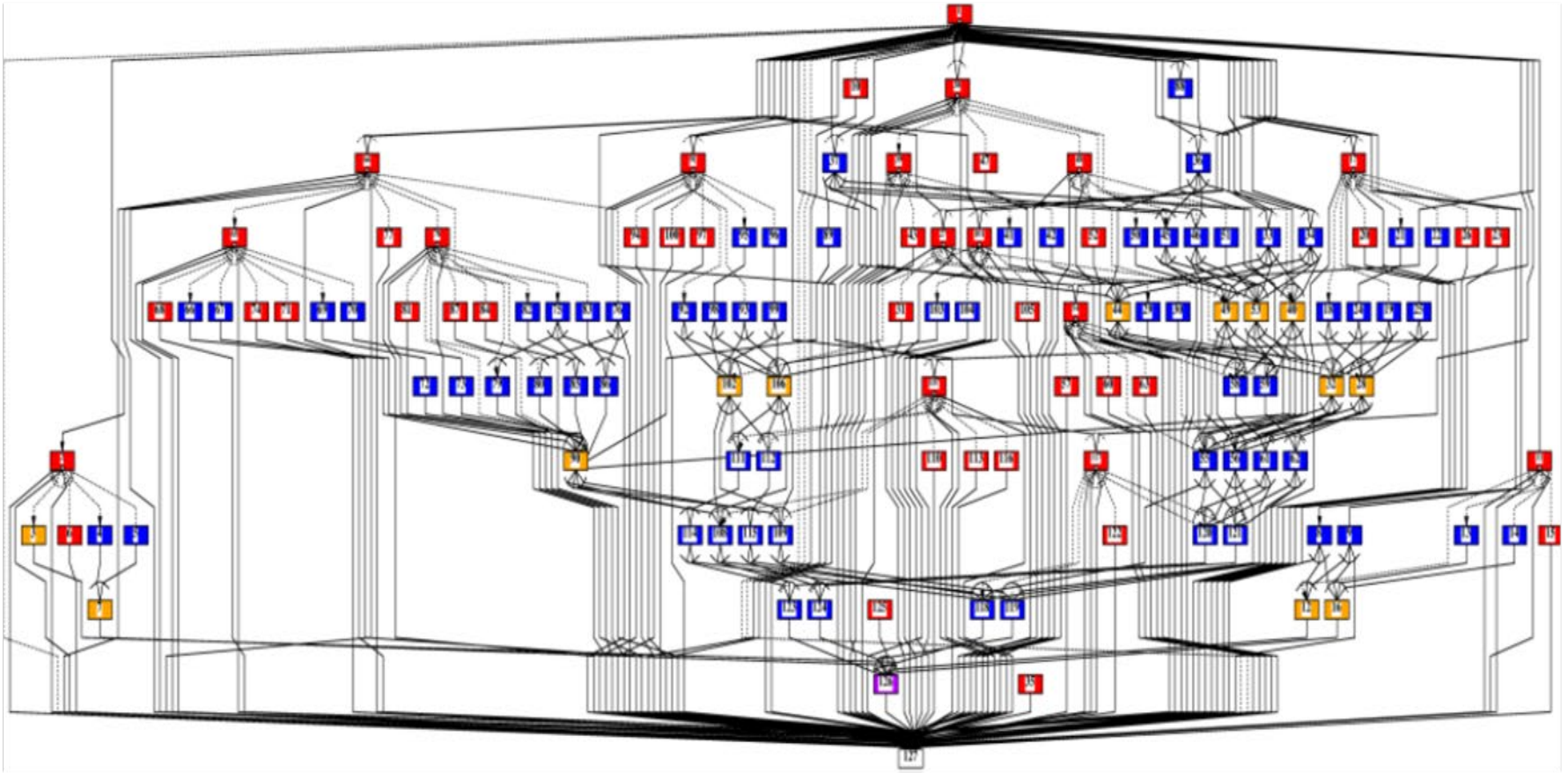




A Macrotask Graph Showing Parallelism among Parallel loops, Sequential loops, Subroutines, and Basic Blocks by th Earliest Executable Condition Analysis in the OSCAR Compiler

MTG of Su2cor-LOOPS-DO400

- **Coarse grain parallelism PARA_ALD = 4.3**

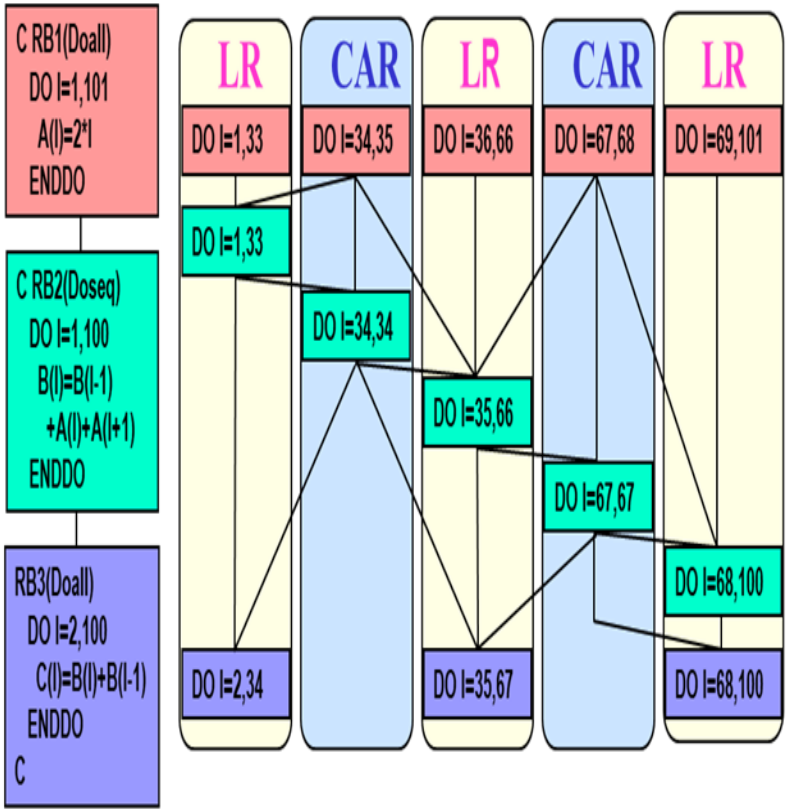


DOALL **Sequential LOOP** **SB** **BB**

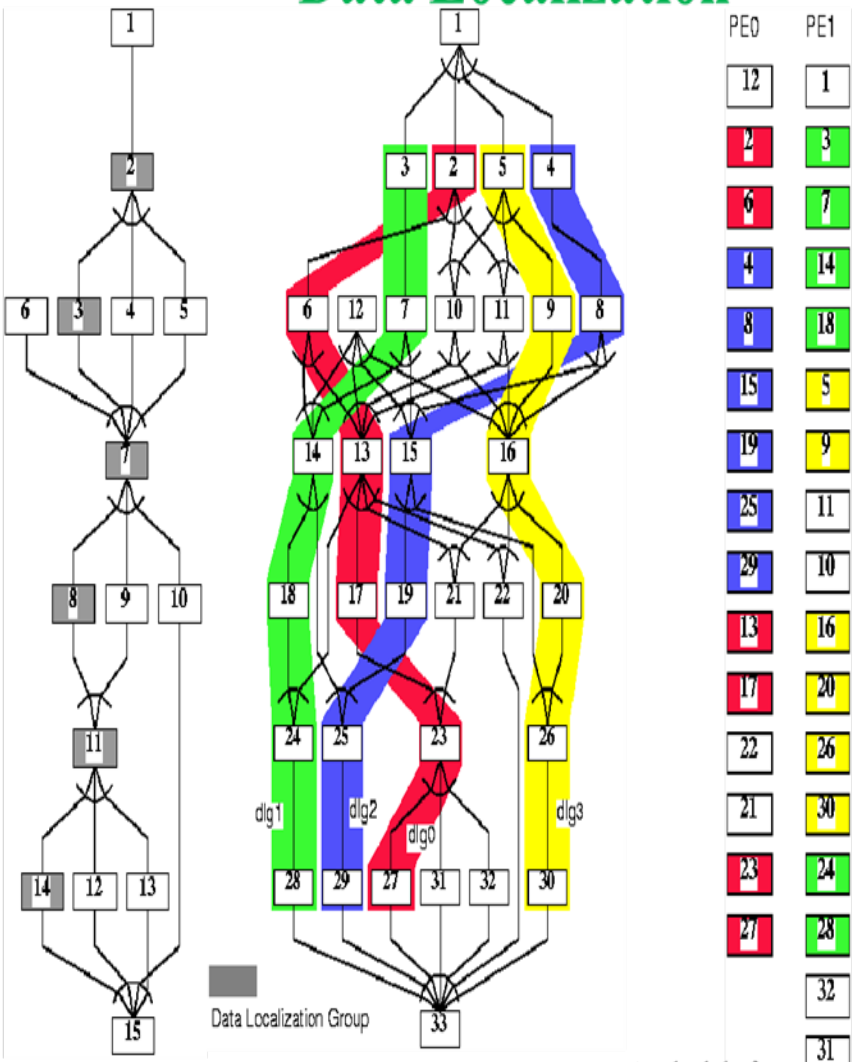
Data Localization for Data Reuse Optimization of Cache & Local Memory

Data-Localization: Loop Aligned Decomposition

- Decompose multiple loop (Doall and Seq) into CARs and LRs considering inter-loop data dependence.
  - Most data in LR can be passed through LM.
  - LR: Localizable Region, CAR: Commonly Accessed Region

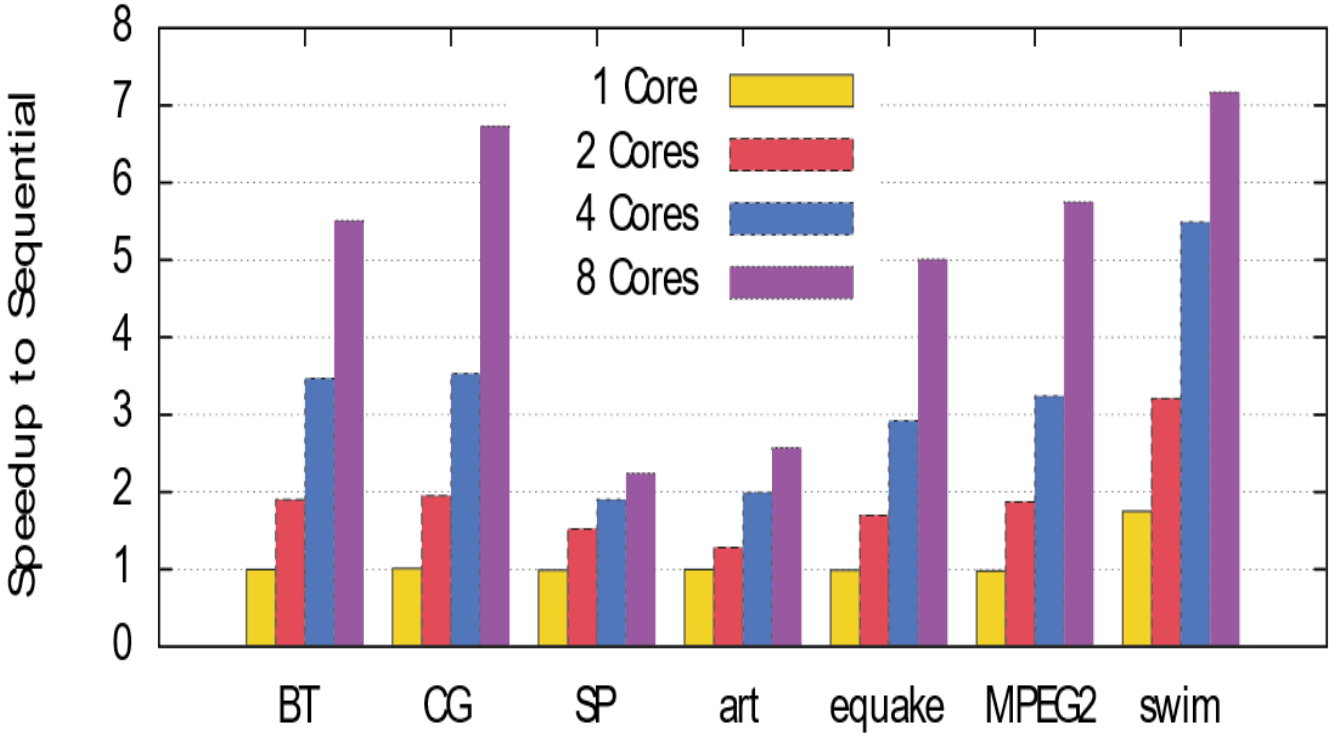


Data Localization



Speedups by OSCAR Compiler on Intel Xeon E5-2650v4

- speedup to sequential version (higher is better)
- gcc as backend



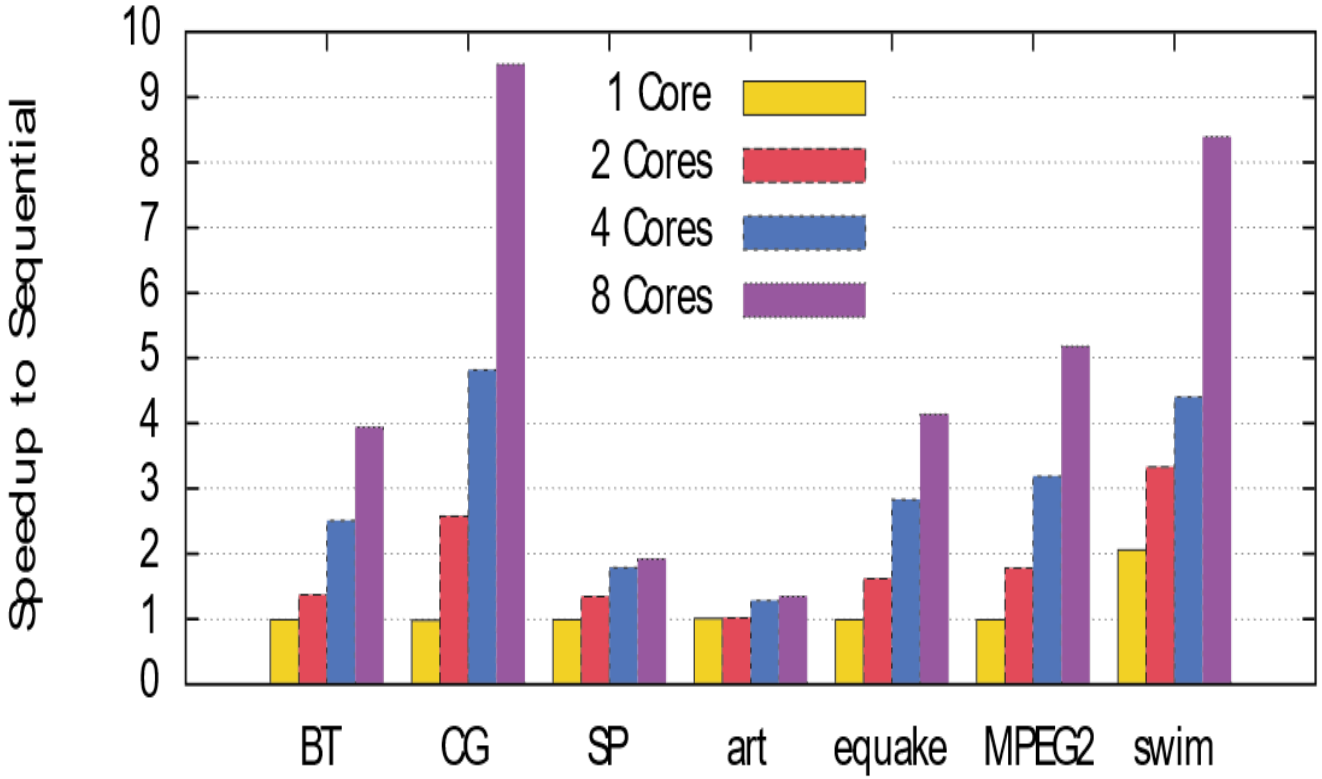
Intel Xeon E5-2650v4

- x86-64 based Architecture
- 12 Cores
- 2.2 GHz – 2.9 GHz
- 30 MiB shared L3 cache
  - L3 Cache: Shared by all cores

- swim shows superlinear speedup and 1 core speedup
  - seq.: 58.1 s, 1 core OSCAR: 33.2 s, 4 core OSCAR: 10.5 s

Speedups by OSCAR Compiler on AMD EPYC 7702P

- speedup to sequential version (higher is better)
- gcc as backend



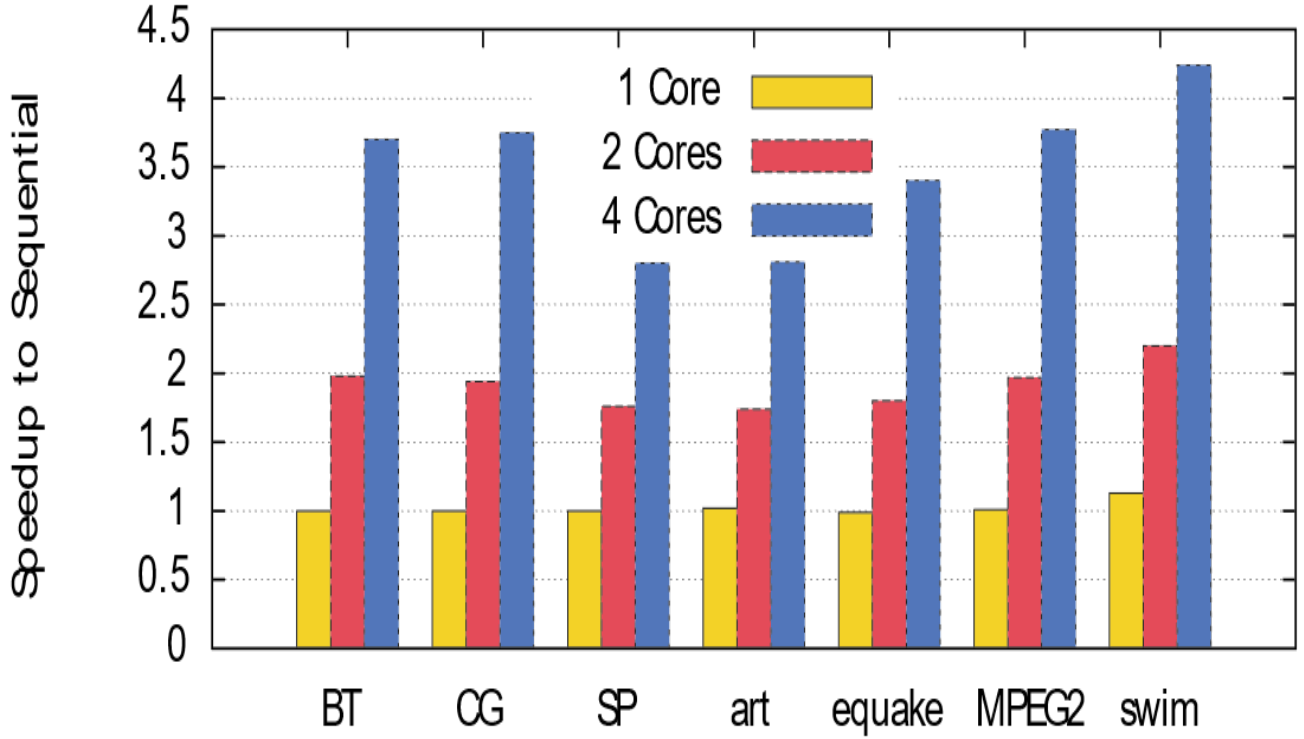
AMD EPYC 7702P

- x86-64 based Architecture
- 64 Cores
- 2.0 GHz – 3.35 GHz
- 16 MiB L3 cache per 4 core cluster
  - shared within the cluster

- CG and swim show superlinear speedup
  - CG: seq.: 0.86 s, 8 core OSCAR: 0.09 s

Speedups by OSCAR Compiler on SiFive Freedom U740

- speedup to sequential version (higher is better)
- gcc as backend



SiFive Freedom U740

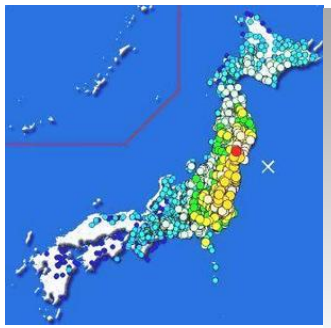
- RISC-V based Architecture
- 4 Cores
- 1.2 GHz
- 2 MiB shared L2 cache
  - L2 Cache: Shared across all cores

- overall good speedup is observed, swim superlinear
  - BT: seq.: 2041 s, 4 core OSCAR: 551 s

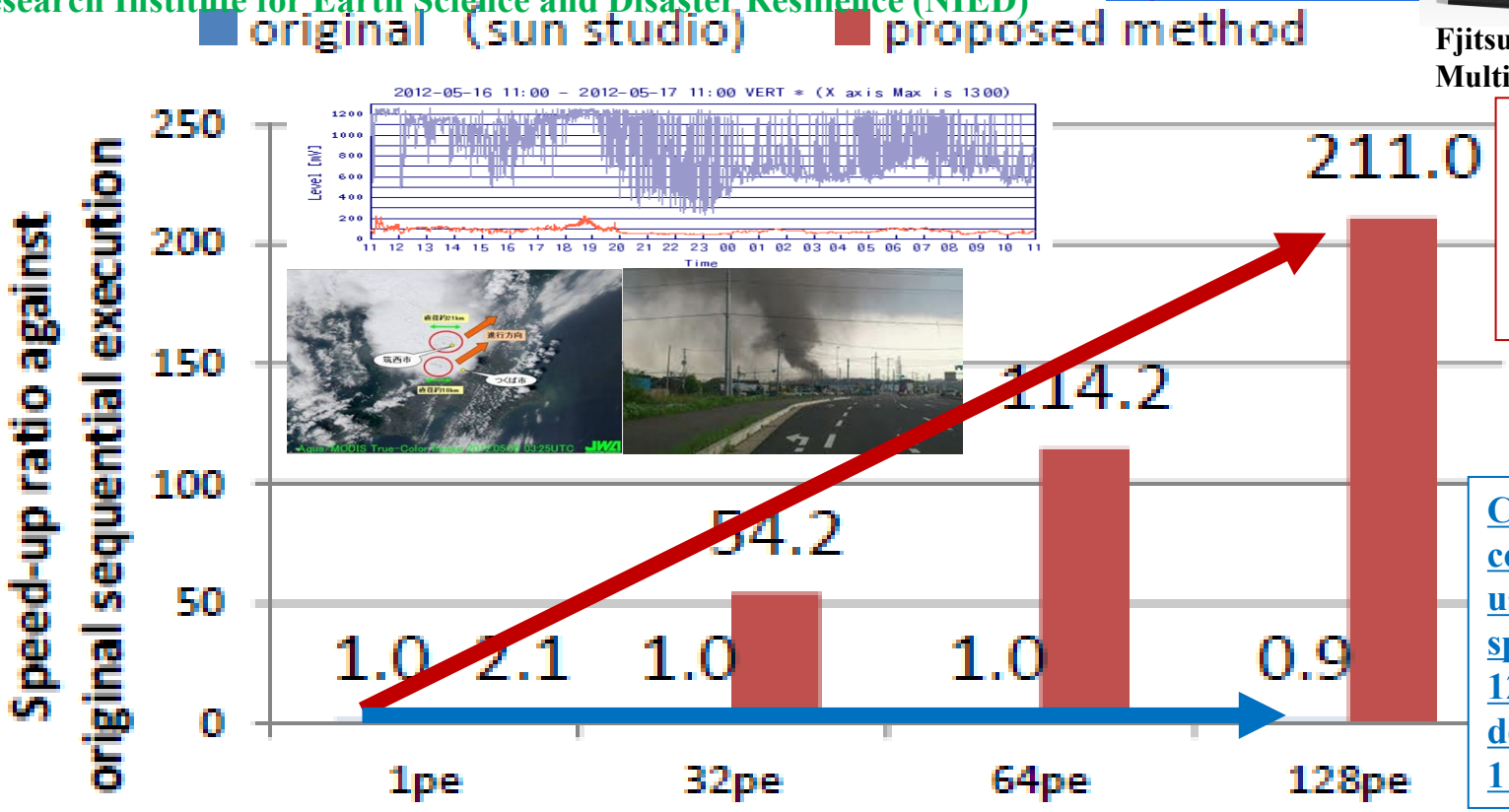
Parallel Soft is important for scalable performance of multicore (LCPC2015)

- Just more cores don't give us speedup
- Development cost and period of parallel software are getting a bottleneck of development of embedded systems, eg. IoT, Automobile

Earthquake wave propagation simulation GMS developed by National Research Institute for Earth Science and Disaster Resilience (NIED)



Fjitsu M9000 SPARC Multicore Server



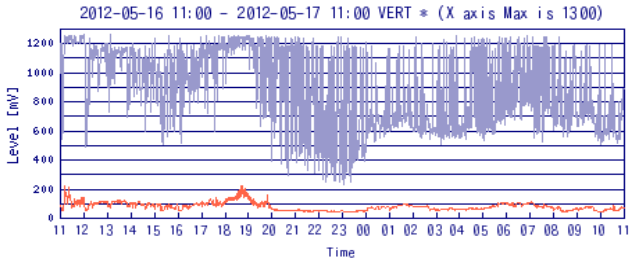
OSCAR Compiler gives us 211 times speedup with 128 cores

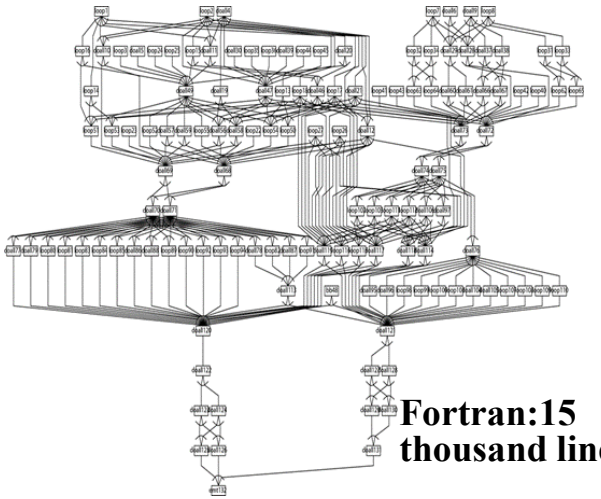
Commercial compiler gives us 0.9 times speedup with 128 cores (slow-downed against 1 core)

- Automatic parallelizing compiler available on the market gave us no speedup against execution time on 1 core on 64 cores
  - Execution time with 128 cores was slower than 1 core (0.9 times speedup)
- Advanced OSCAR parallelizing compiler gave us 211 times speedup with 128cores against execution time with 1 core using commercial compiler
  - OSCAR compiler gave us 2.1 times speedup on 1 core against commercial compiler by global cache optimization

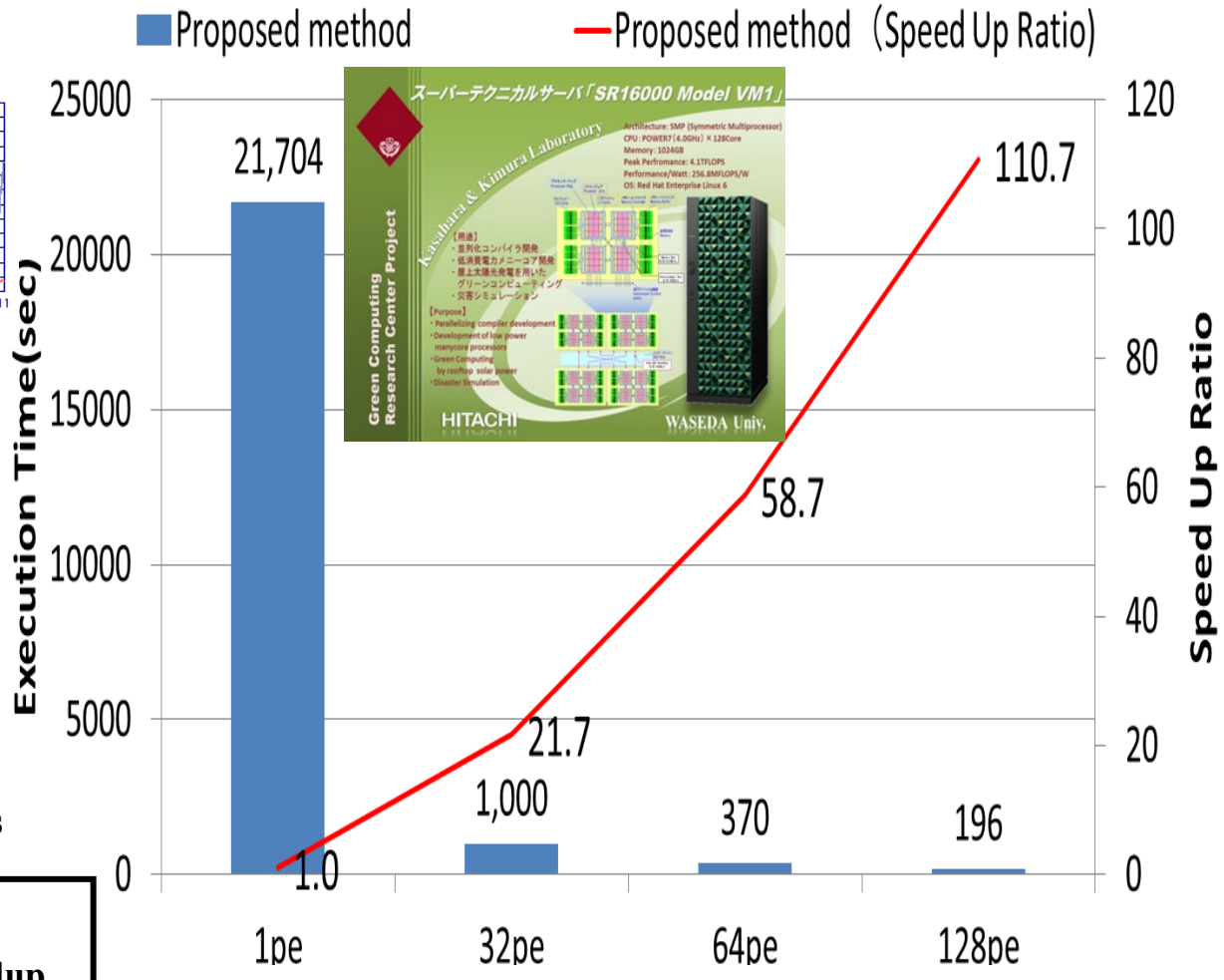
110 Times Speedup against the Sequential Processing for GMS Earthquake Wave Propagation Simulation on Hitachi SR16000

（Power7 Based 128 Core Linux SMP） (LCPC2015)





Fortran:15 thousand lines

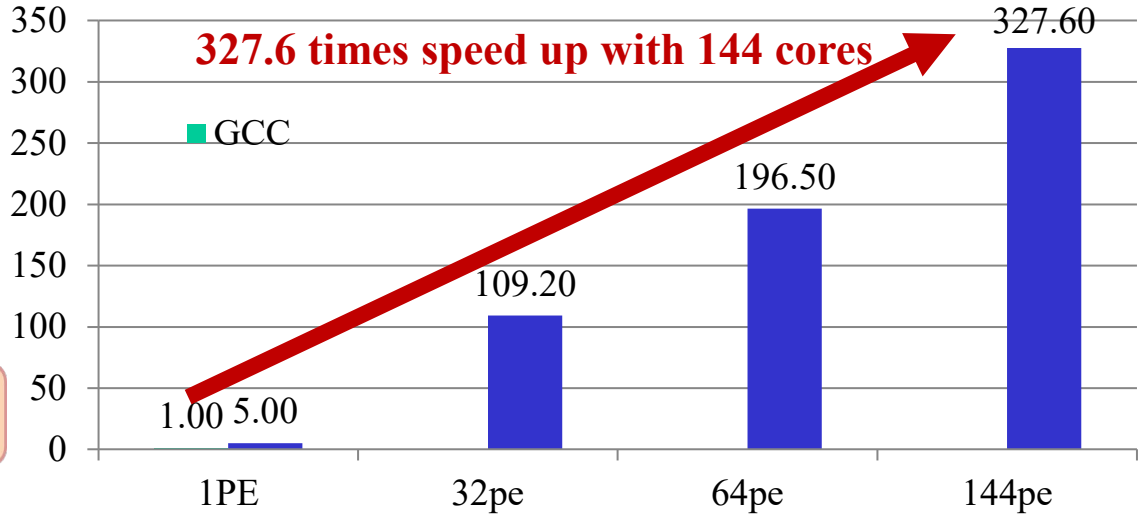


First touch for distributed shared memory and cache optimization over loops are important for scalable speedup

Performance on Multicore Server for Latest Cancer Treatment Using Heavy Particle (Proton, Carbon Ion)

327 times speedup on 144 cores

Hitachi 144cores SMP Blade Server BS500:
Xeon E7-8890 V3(2.5GHz　18core/chip) x8 chip

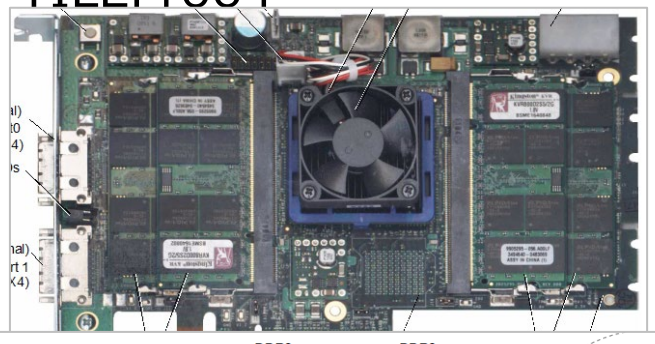


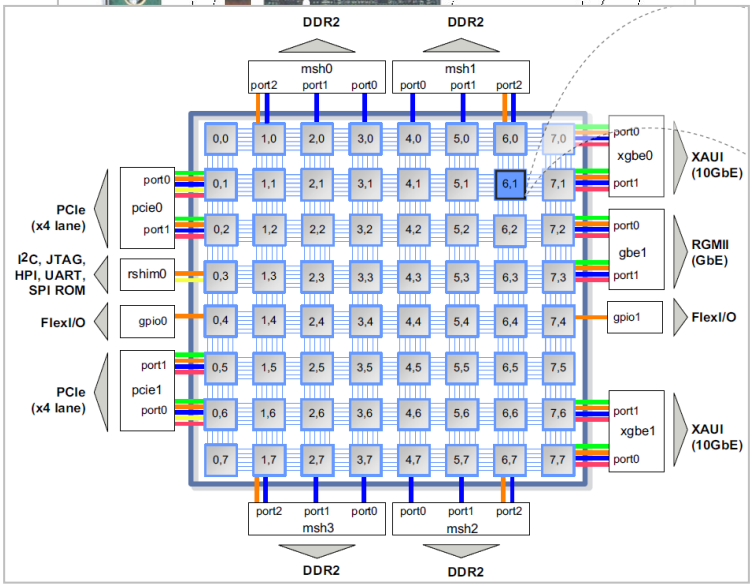
- Original sequential execution time 2948 sec（50 minutes）using GCC was reduced to 9 sec with 144 cores（327.6 times speedup）
  - Reduction of treatment cost and reservation waiting period is expected

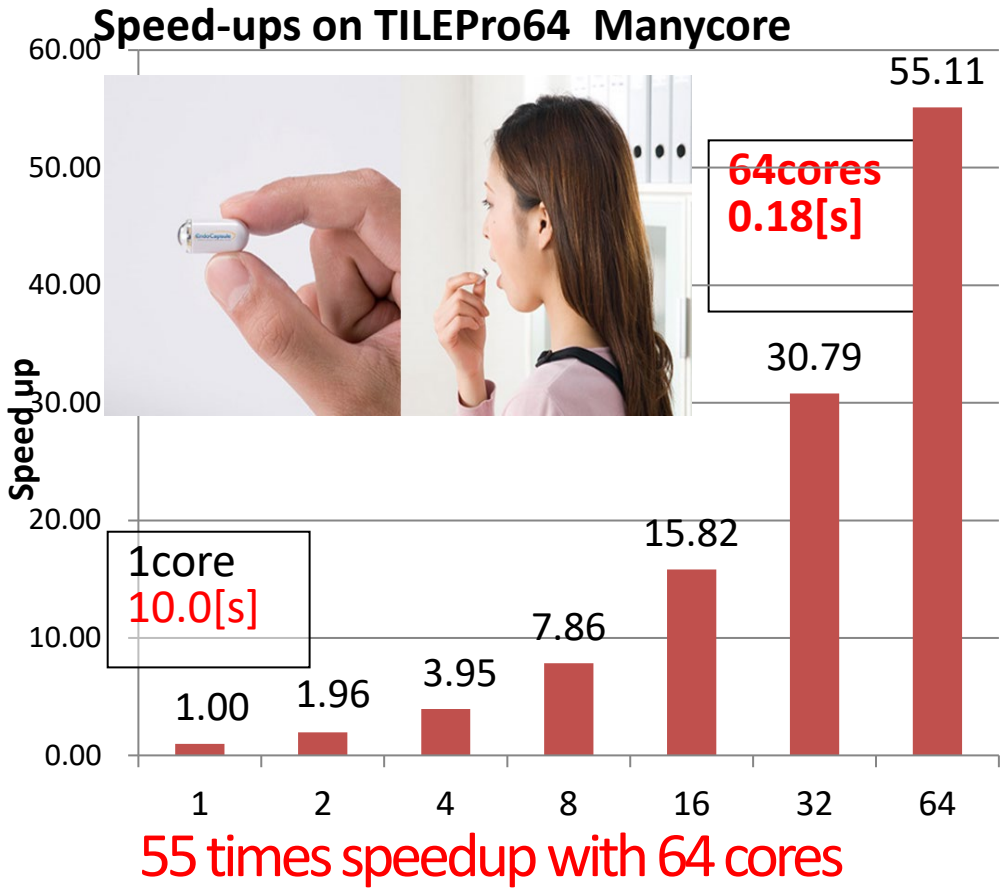
Automatic Parallelization of JPEG-XR for Drinkable Inner Camera (Endo Capsule)

10 times more speedup needed after parallelization for 128 cores of Power 7. Less than 35mW power consumption is required.

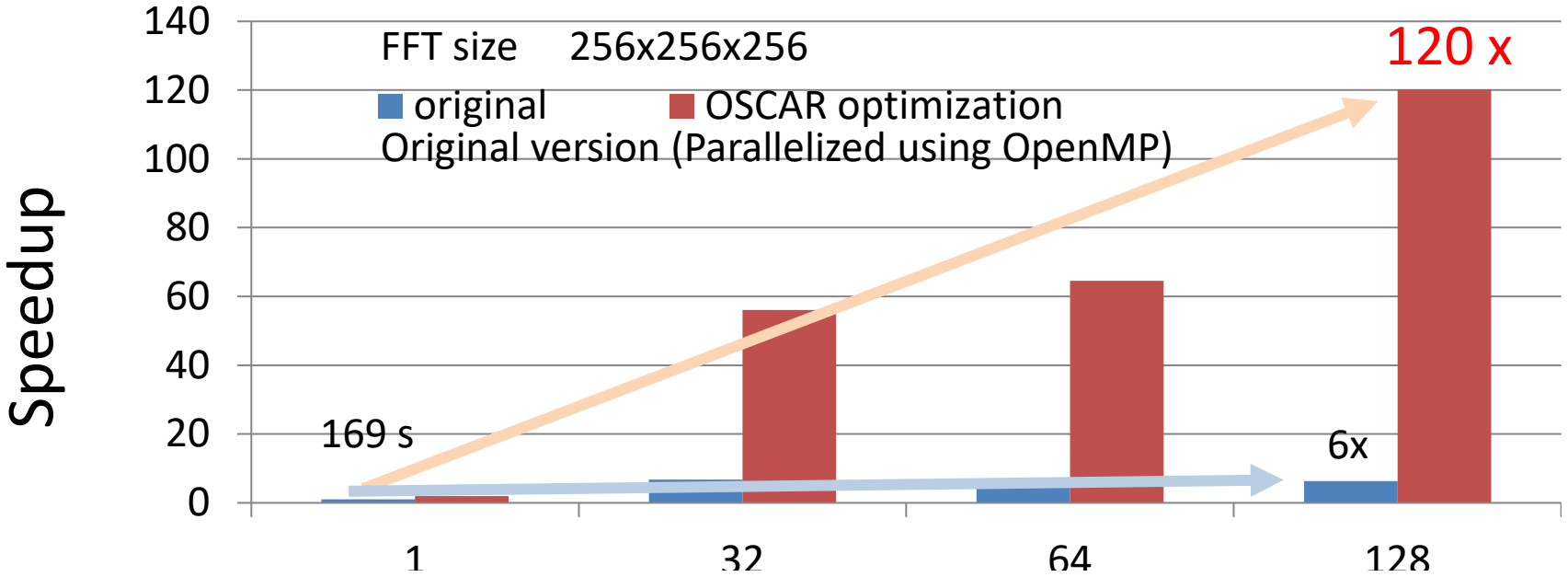
- TILEPro64







Parallelization of 3D-FFT for New Magnetic Material Computation on Hitachi SR16000 Power7 CC-Numa Server



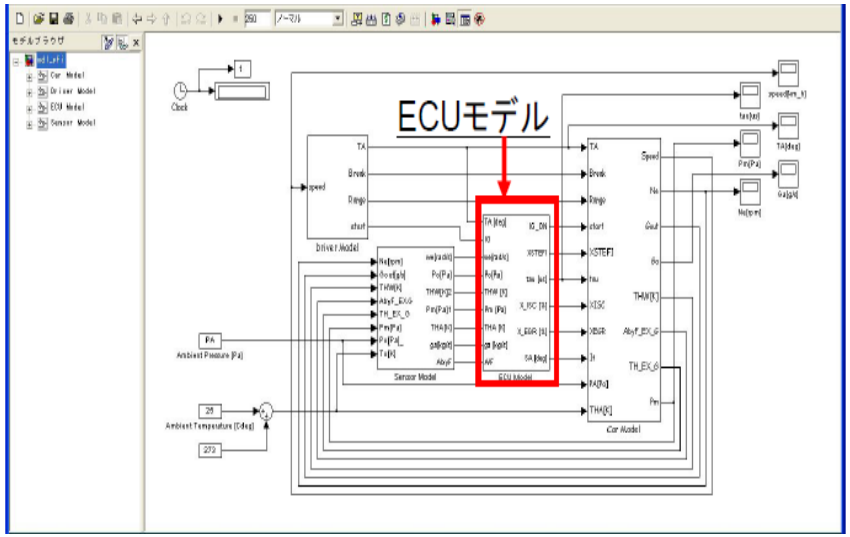
OSCAR optimization

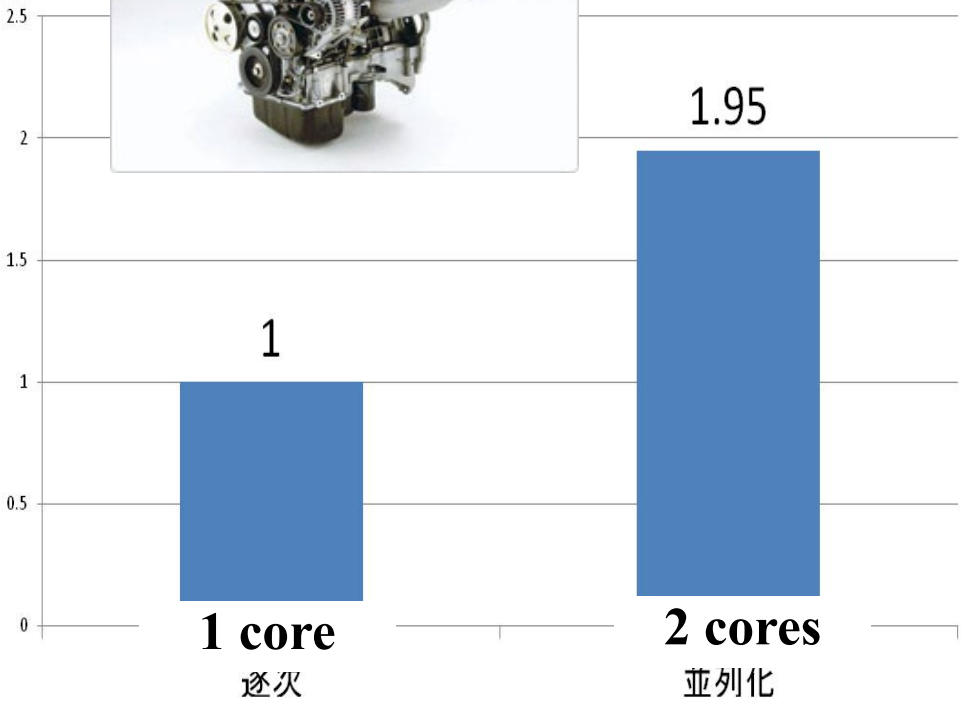
- reducing number of data transpose with interchange, code motion and loop fusion

Engine Control by multicore with Denso

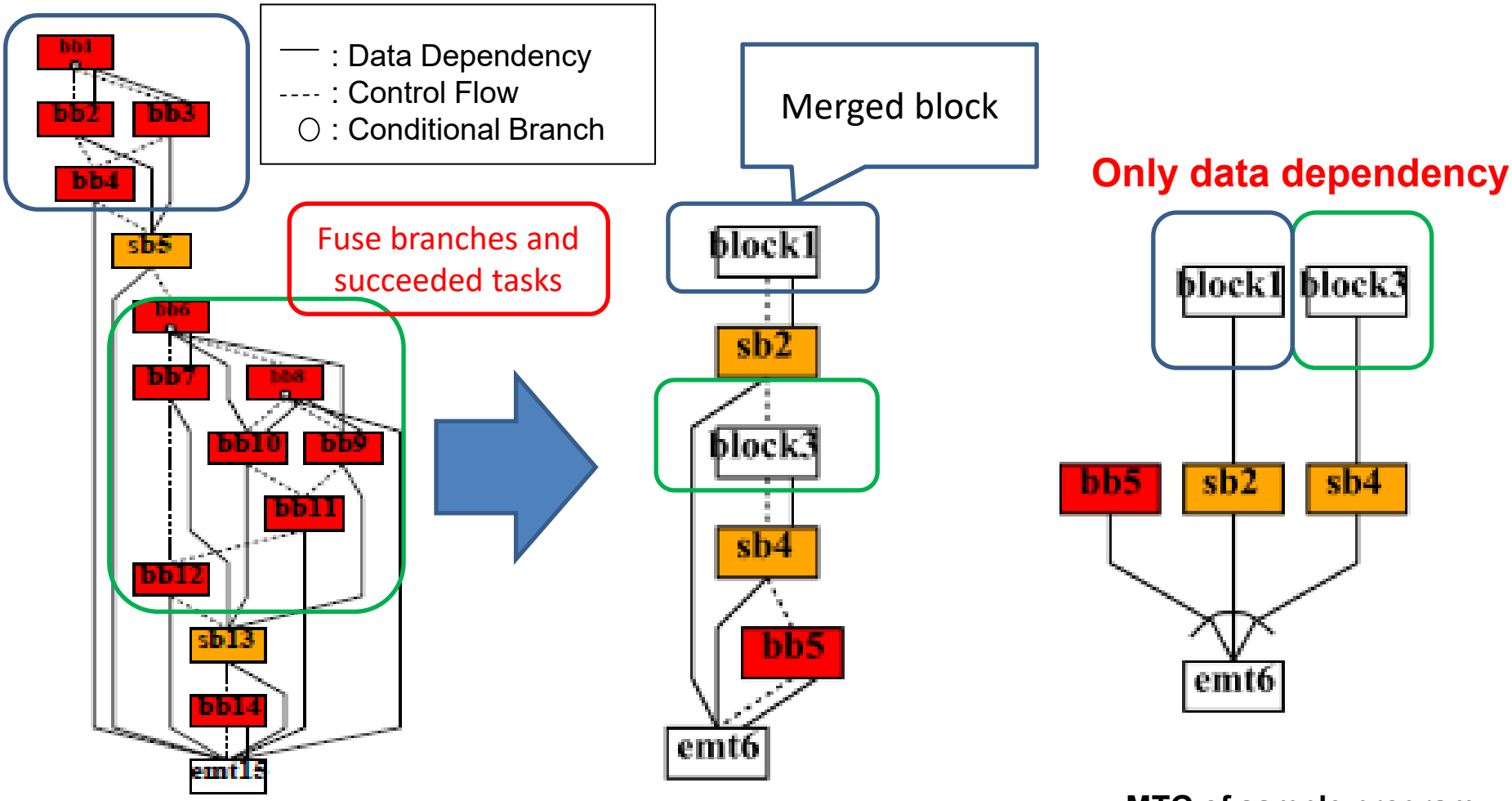
Though so far parallel processing of the engine control on multicore has been very difficult, Denso and Waseda succeeded 1.95 times speedup on 2core V850 multicore processor.

- Hard real-time automobile engine control by multicore using local memories
- Millions of lines C codes consisting conditional branches and basic blocks





Macro Task Fusion for Static Task Scheduling



MFG of sample program before maro task fusion

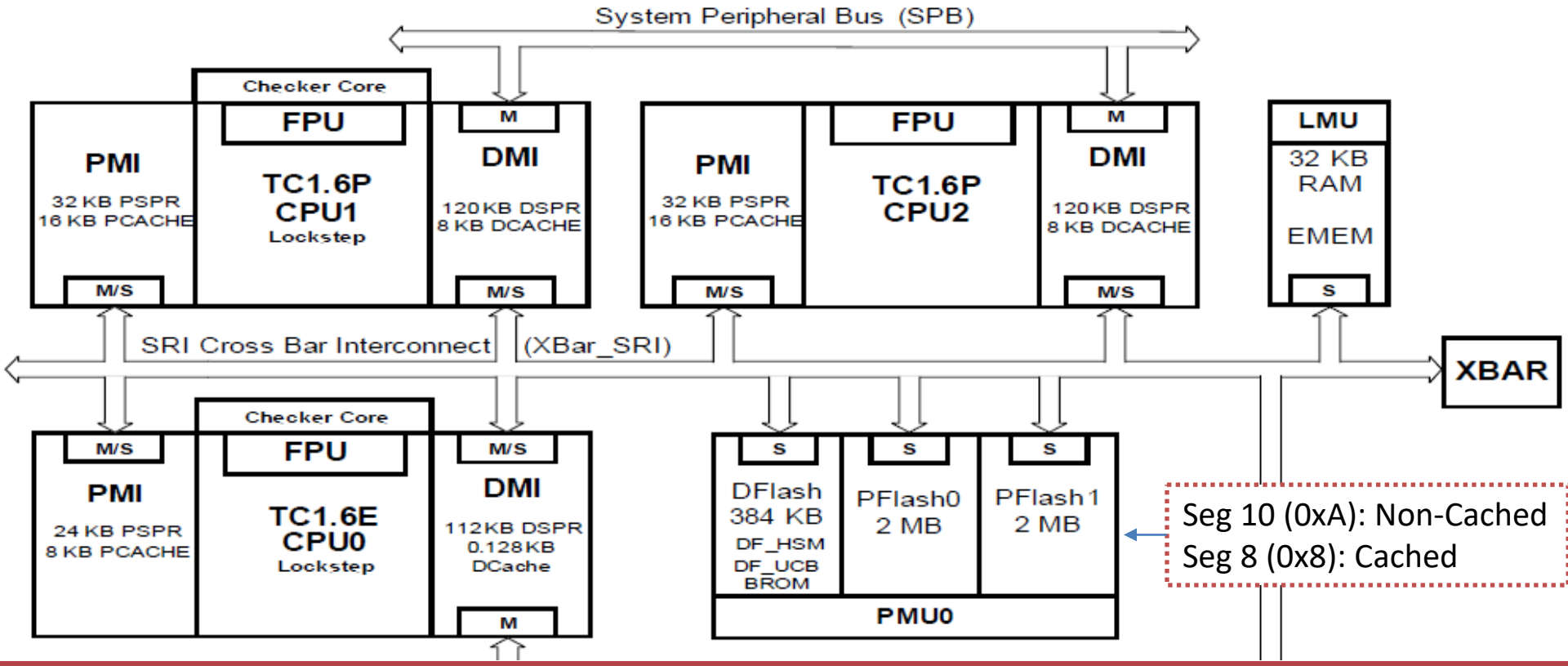
MFG of sample program after macro task fusion

MTG of sample program after macro task fusion

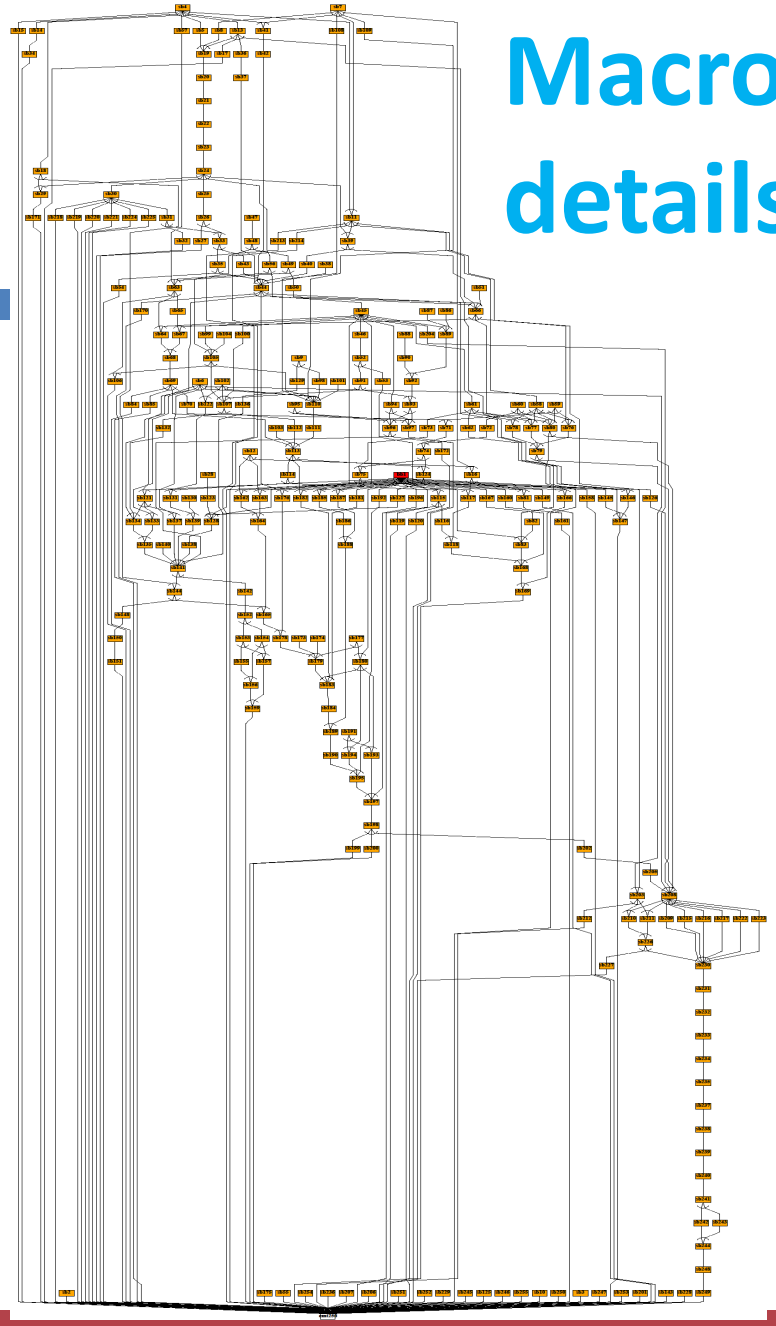
Infineon AURIX TC277

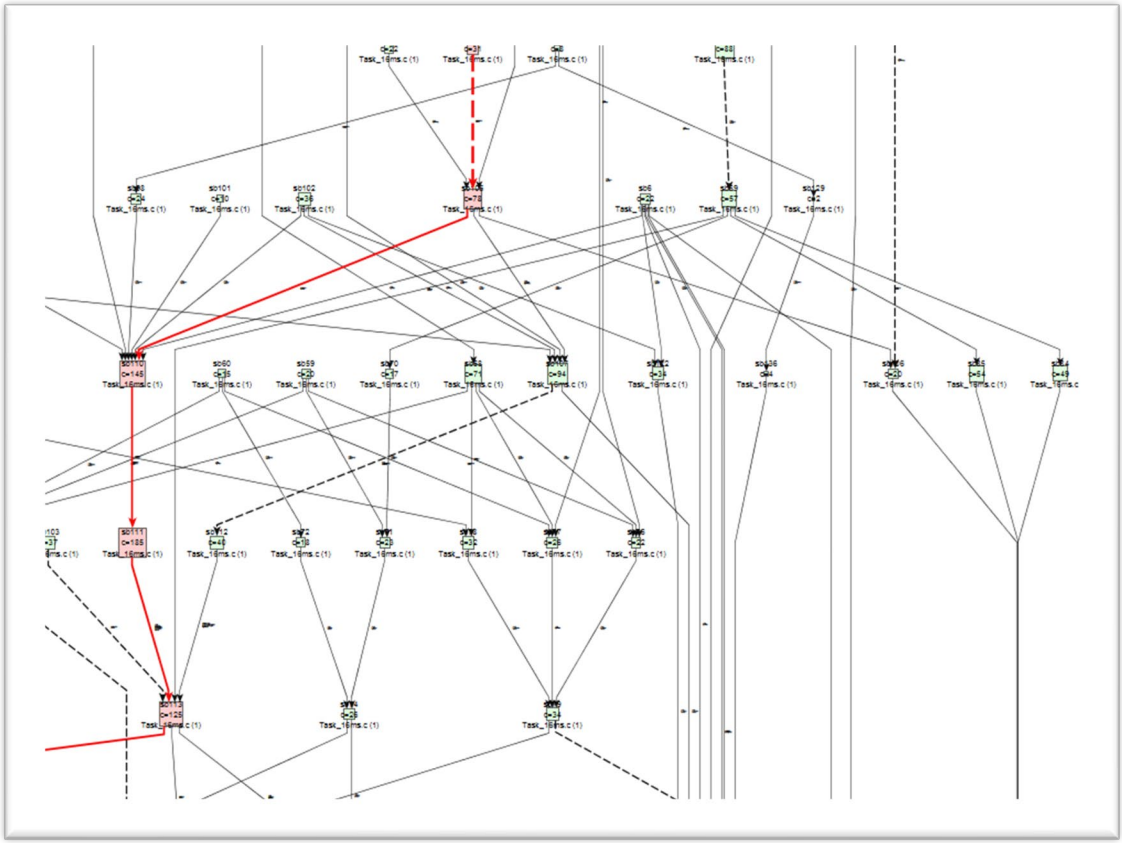
Abbreviations :

PCACHE:	Program Cache
DCACHE :	Data Cache
DSPR:	Data Scratch-Pad RAM
PSPR:	Program Scratch-Pad RAM
BROM:	Boot ROM
PFlash:	Program Flash
DFlash:	Data Flash (EEEPROM)
S :	SRI Slave Interface
M :	SRI Master Interface



Macrotask Graph, Dependence details and schedules

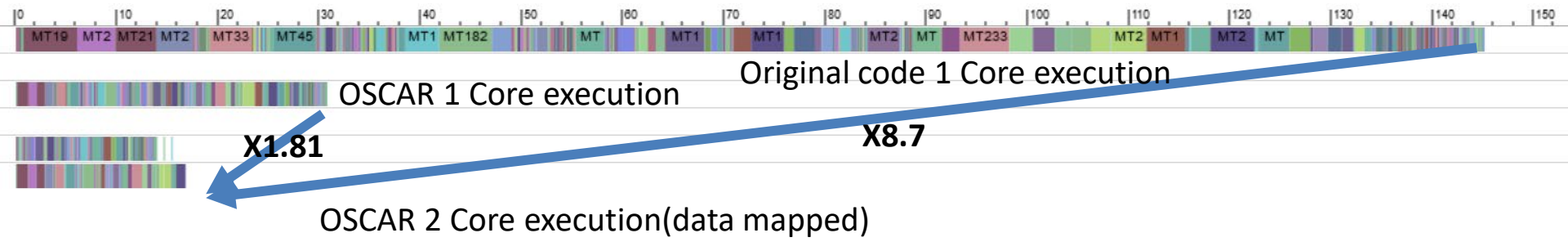




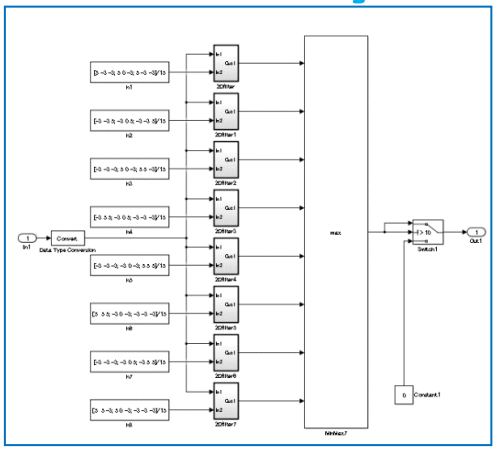
MTG – task_16ms

Automatic Parallelization of an Engine Control C Program with 400 thousands lines on AUTOSAR on 2 cores of Infineon AURIX TC277

- **Original sequential** execution time on 1 core: **145500** cycles
- **Sequential execution time by OSCAR** on 1 core:**29700** cycles
  - 4.9 times speedup on 1 core against original execution by OSCAR Compilers automatic data allocation for local scratch pad memory, flush memory modules
- **2 core execution by OSCAR** Compiler:**16400** cycles
  - **1.81 times speedup with 2 core against 1 core execution with OSCAR Compiler**
  - **8.7 times speedup against original sequential execution.**



Automatic Parallelization of MATLAB/Simulink Applications by OSCAR Compiler



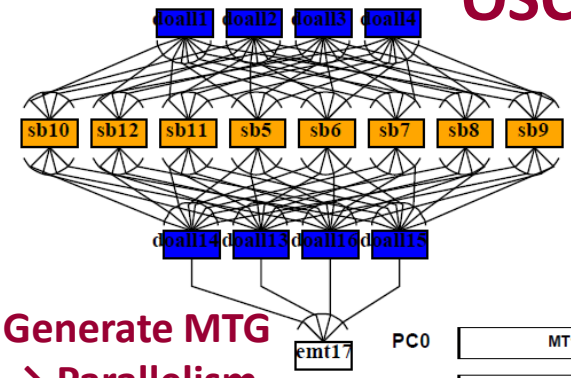
Simulink model

Generate C code using Embedded Coder

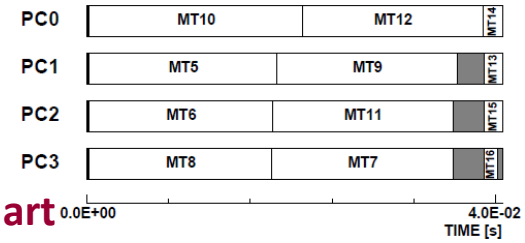
```
/* Model step function */
void VesselExtraction_step(void)
{
  int32_T i;
  real_T u0;

  /* DataTypeConversion: '<S1>/Data Type Conversion' incorporates:
   *  Inport: '<Root>/In1'
   */
  for (i = 0; i < 16384; i++) {
    VesselExtraction_B.DataTypeConversion[i] = VesselExtraction_U.In1[i];
  }

  /* End of DataTypeConversion: '<S1>/Data Type Conversion' */

  /* Outputs for Atomic SubSystem: '<S1>/2Dfilter' */

  /* Constant: '<S1>/h1' */
  VesselExtraction_Dfilter(VesselExtraction_B.DataTypeConversion,
    VesselExtraction_P.h1_Value, &VesselExtraction_B.Dfilter,
    (P_Dfilter_VesselExtraction_T *)&VesselExtraction_P.Dfilter);

  /* End of Outputs for SubSystem: '<S1>/2Dfilter' */

  /* Outputs for Atomic SubSystem: '<S1>/2Dfilter1' */

  /* Constant: '<S1>/h2' */
  VesselExtraction_Dfilter(VesselExtraction_B.DataTypeConversion,
    VesselExtraction_P.h2_Value, &VesselExtraction_B.Dfilter1,
    (P_Dfilter_VesselExtraction_T *)&VesselExtraction_P.Dfilter1);
```

C code

OSCAR Compiler



(1) Generate MTG
→ Parallelism



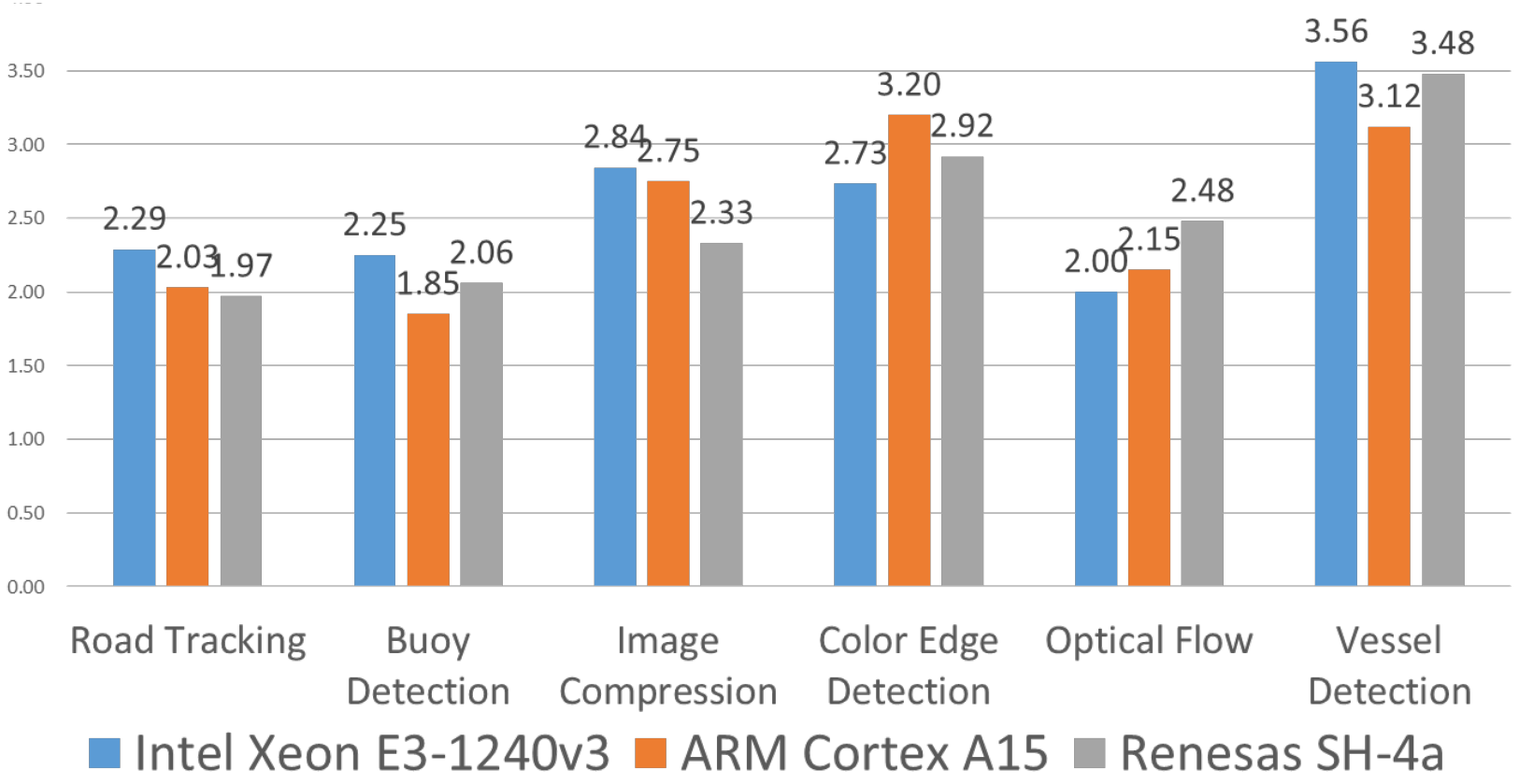
(2) Generate gantt chart
→ Scheduling in a multicore

```
void VesselExtraction_step ( )
{
    int thr1 ;
    int thr2 ;
    int thr3 ;
    {
        oscar_thread_create ( & thr1 ,
            thread_function_001 , (void*)1 ) ;
        oscar_thread_create ( & thr2 ,
            thread_function_002 , (void*)2 ) ;
        oscar_thread_create ( & thr3 ,
            thread_function_003 , (void*)3 ) ;

            VesselExtraction_step_PE0 ( ) ;

        oscar_thread_join ( thr1 ) ;
        oscar_thread_join ( thr2 ) ;
        oscar_thread_join ( thr3 ) ;
    }
}

void thread_function_001 ( void )
{
    VesselExtraction_step_PE1 ( ) ;
}
```

(3) Generate parallelized C code using the OSCAR API
→ Multiplatform execution (Intel, ARM and SH etc)

Speedups of MATLAB/Simulink Image Processing on Various 4core Multicores

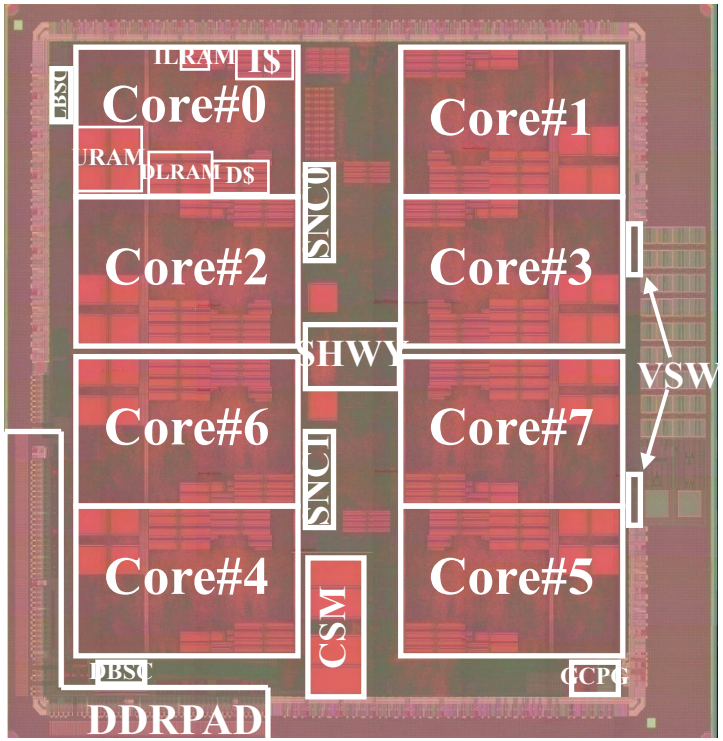
(Intel Xeon, ARM Cortex A15 and Renesas SH4A)



Road Tracking, Image Compression : http://www.mathworks.co.jp/jp/help/vision/examples
Buoy Detection : http://www.mathworks.co.jp/matlabcentral/fileexchange/44706-buoy-detection-using-simulink
Color Edge Detection : http://www.mathworks.co.jp/matlabcentral/fileexchange/28114-fast-edges-of-a-color-image--actual-color--not-converting-to-grayscale-/
Vessel Detection : http://www.mathworks.co.jp/matlabcentral/fileexchange/24990-retinal-blood-vessel-extraction/

Multicores for Performance and Low Power

Data Center: 100MW (Thermal Power Plant) → 1GW (Nuclear Power Plant)



IEEE ISSCC08: Paper No. 4.5, M.ITO, … and H. Kasahara, "An 8640 MIPS SoC with Independent Power-off Control of 8 CPUs and 8 RAMs by an Automatic Parallelizing Compiler"

Power ∝ Frequency * Voltage^2

(Voltage ∝ Frequency)

⇒ Power ∝ Frequency^3

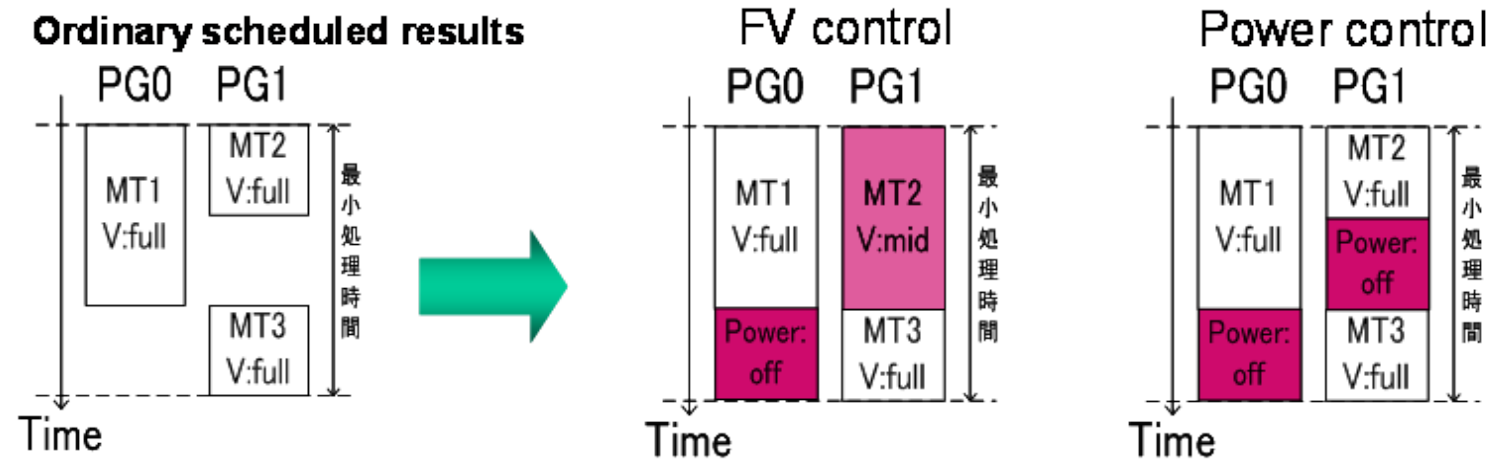
If Frequency is reduced to 1/4 (Ex. 4GHz→1GHz), Power is reduced to 1/64 and Performance falls down to 1/4 .

<Multicores>

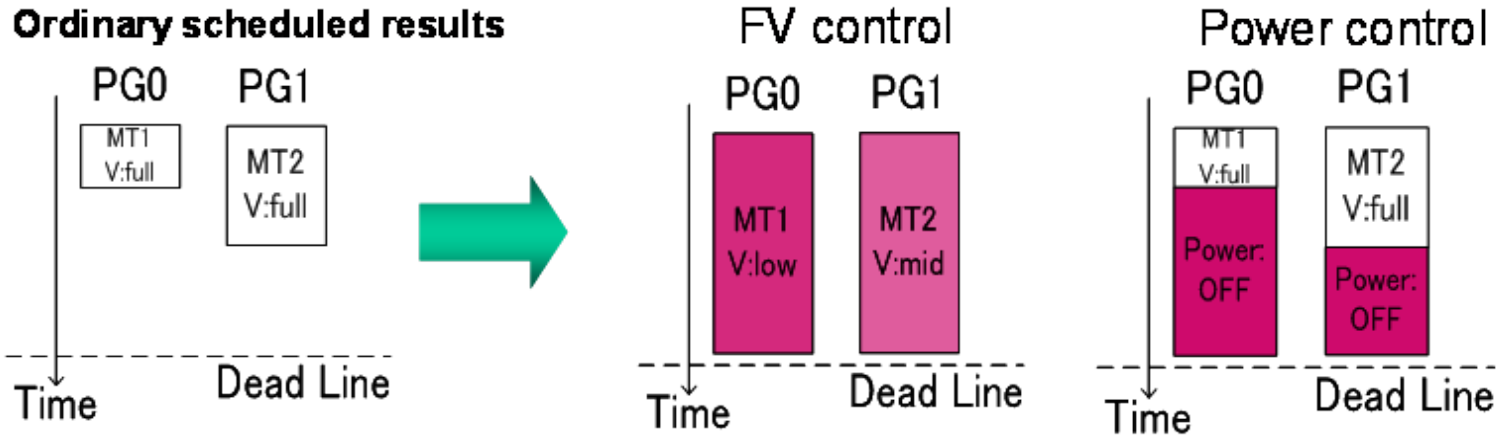
If 8cores are integrated on a chip, Power is still 1/8 and Performance becomes 2 times.

Power Reduction by Power Supply, Clock Frequency and Voltage Control by OSCAR Compiler

- Shortest execution time mode



- Realtime processing mode with dead line constraints



OSCAR API for Homogeneous/Heterogeneous Multicores with Multigrain Parallelization, Local Memory Management, Software Cache Coherent Control, DMA Data Transfer, Power Reduction

OSCAR API Ver. 2.0 for Homogeneous/Heterogeneous Multicores and Manycores

LCPC2009Homogeneous, 2010 Heterogeneous)

List of Directives (22 directives)

- Parallel Execution API
  - parallel sections (*)
  - flush (*)
  - critical (*)
  - execution
- Memoay Mapping API
  - threadprivate (*)
  - distributedshared
  - onchipshared
- Synchronization API
  - groupbarrier
- Data Transfer API
  - dma_transfer
  - dma_contiguous_parameter
  - dma_stride_parameter
  - dma_flag_check
  - dma_flag_send
- Power Control API
  - fvcontrol
  - get_fvstatus
- Timer API
  - get_current_time
- Accelerator
  - accelerator_task_entry
- Cache Control
  - cache_writeback
  - cache_selfinvalidate
  - complete_memop
  - noncacheable
  - aligncache

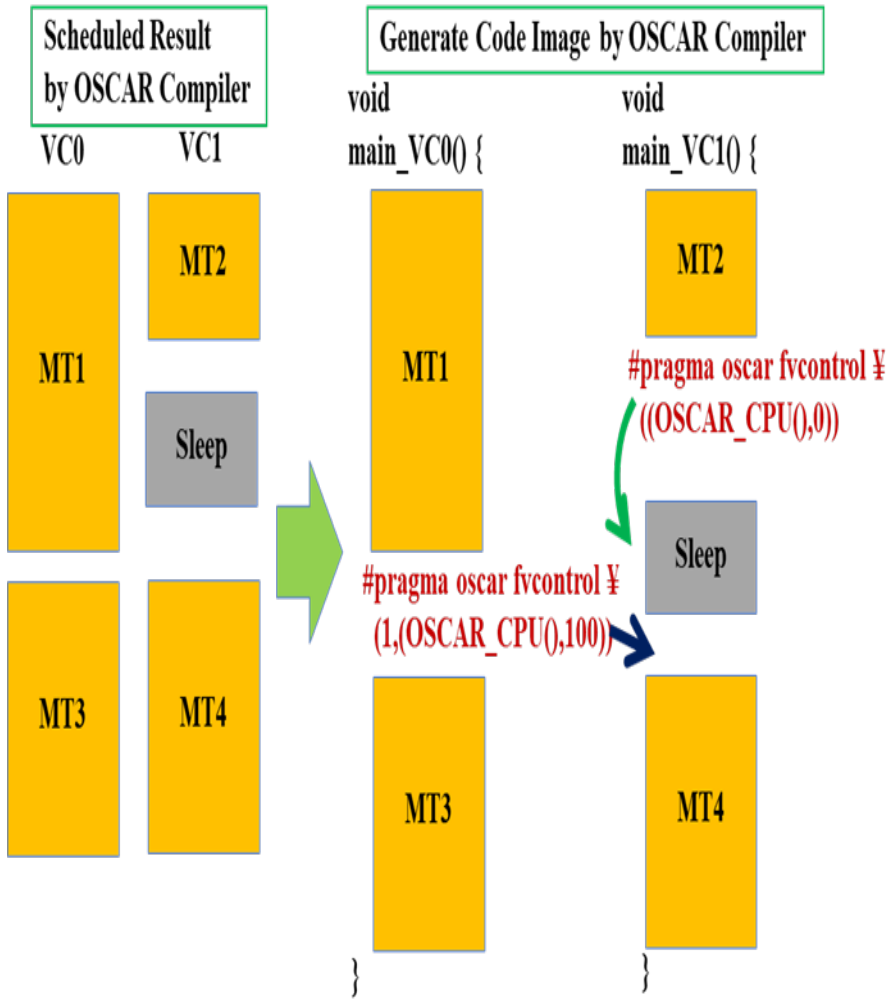
For systems having no cache coherent control hardware

2 hint directives for OSCAR compiler
- accelerator_task
- oscar_comment

from V2.0

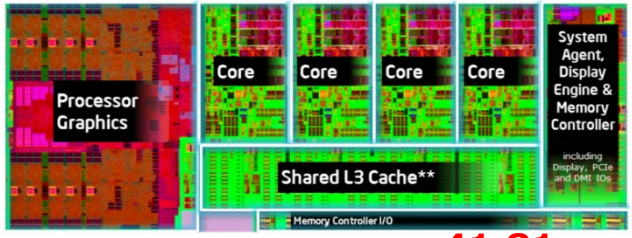
(* from OpenMP)

Low-Power Optimization with OSCAR API

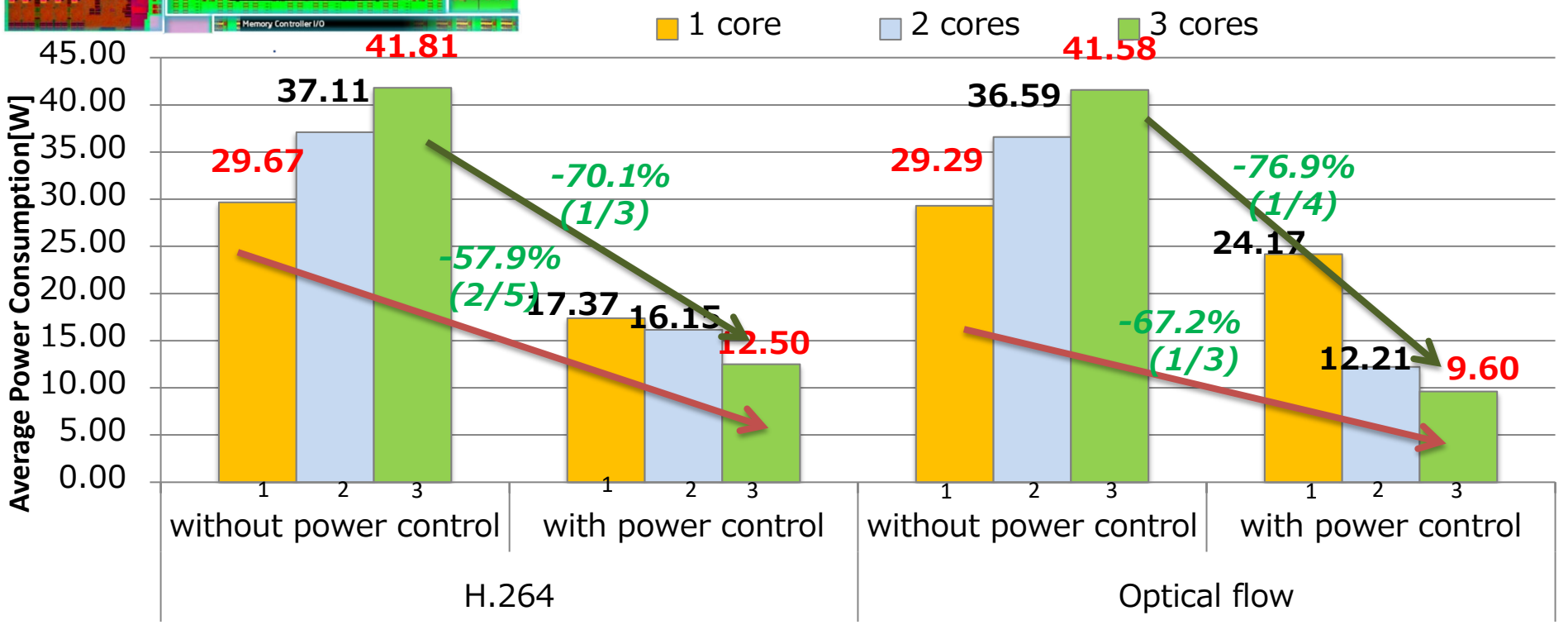


Automatic Power Reuction on Intel Haswell

H.264 decoder & Optical Flow (3cores)



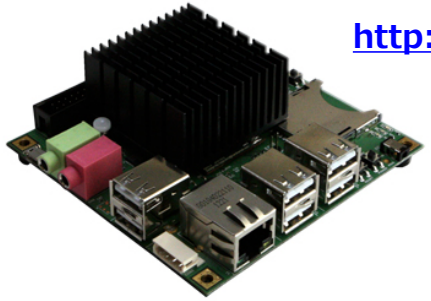
H81M-A, Intel Core i7 4770k
Quad core, 3.5GHz～0.8GHz



Power for 3cores was reduced to 1/3～1/4 against without software power control
Power for 3cores was reduced to 2/5～1/3 against ordinary 1core execution

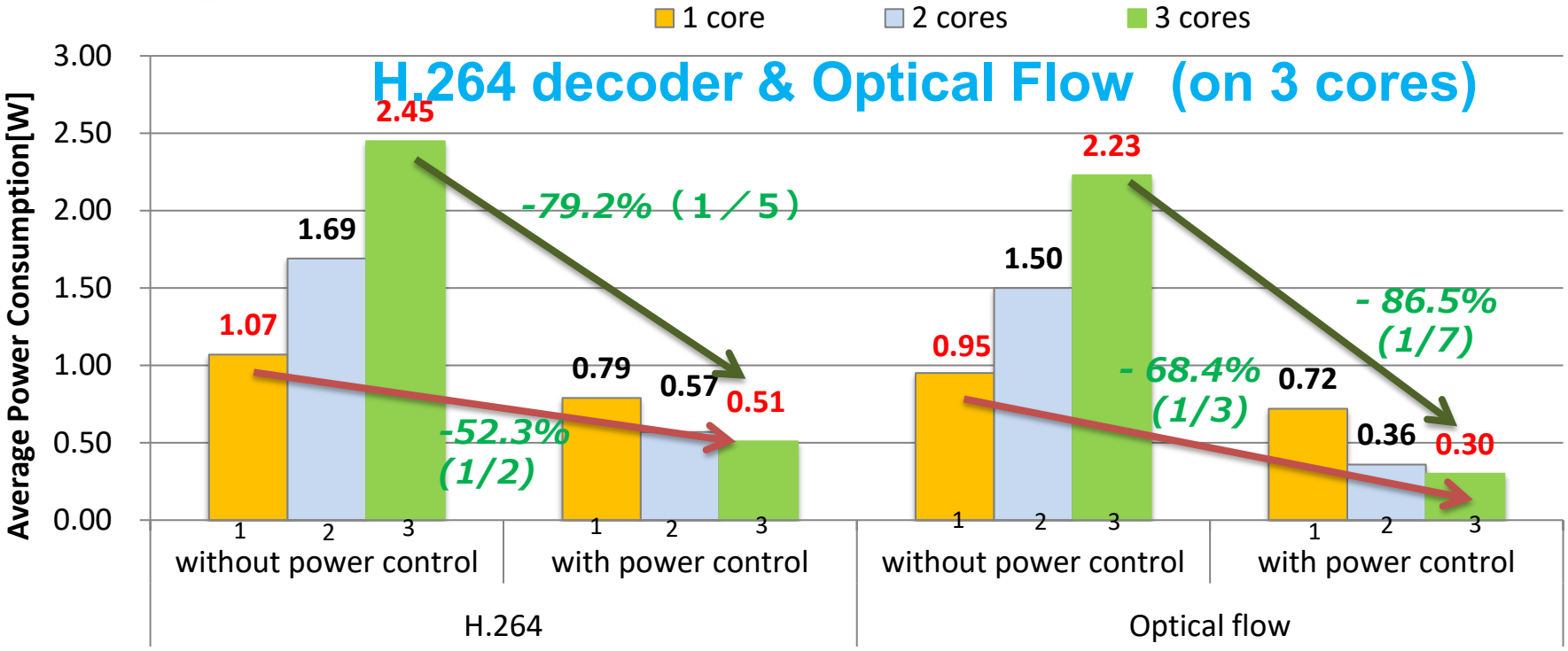
Automatic Power Reduction on ARM CortexA9 with Android

http://www.youtube.com/channel/UCS43lNYEIkC8i_KIgFZYQBQ



ODROID X2

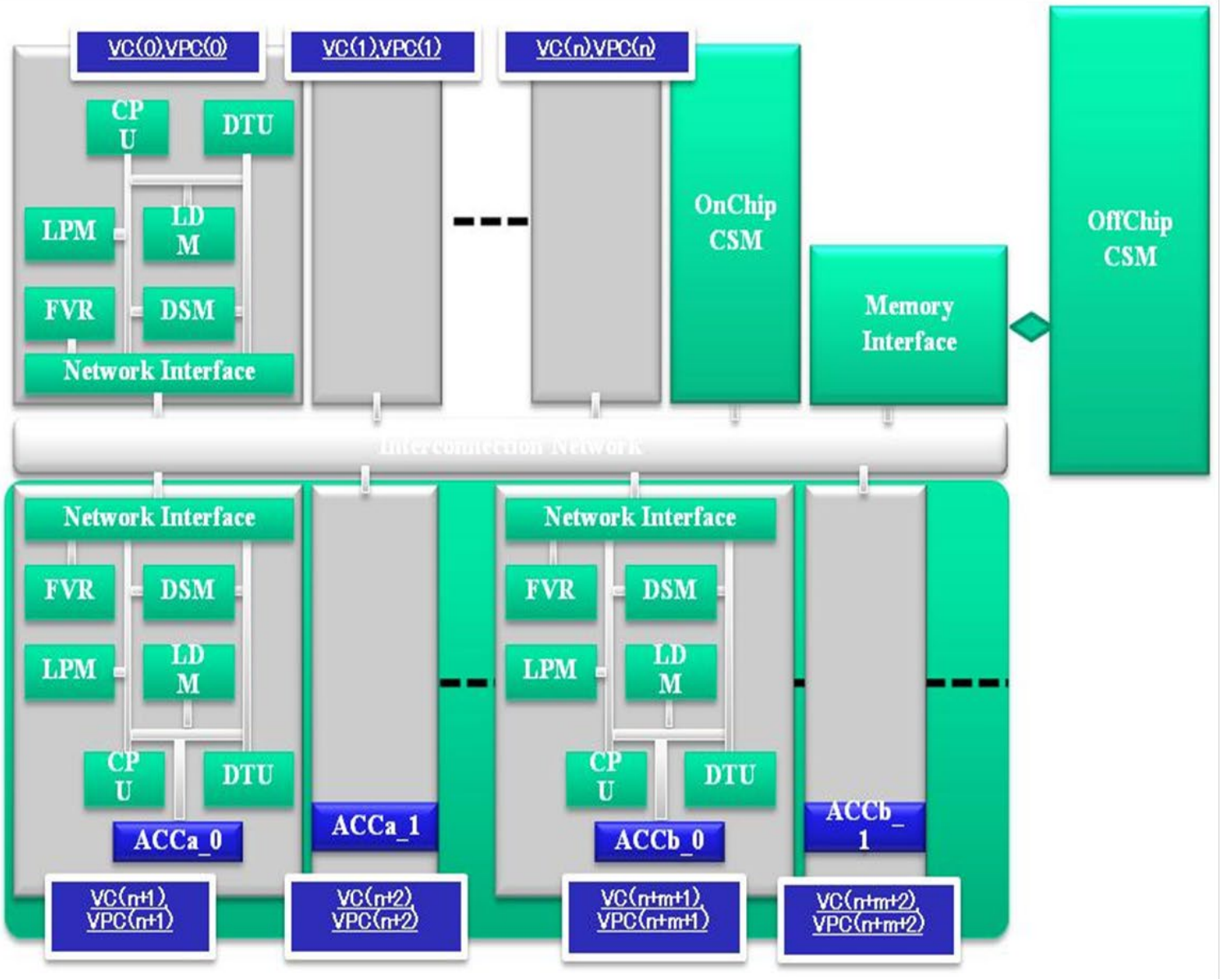
Samsung Exynos4412 Prime, ARM Cortex-A9 Quad core 1.7GHz～0.2GHz, used by Samsung's Galaxy S3



Power for 3cores was reduced to 1/5～1/7 against without software power control

Power for 3cores was reduced to 1/2～1/3 against ordinary 1core execution

OSCAR Heterogeneous Multicore



DTU
- Data Transfer Unit

LPM
- Local Program Memory

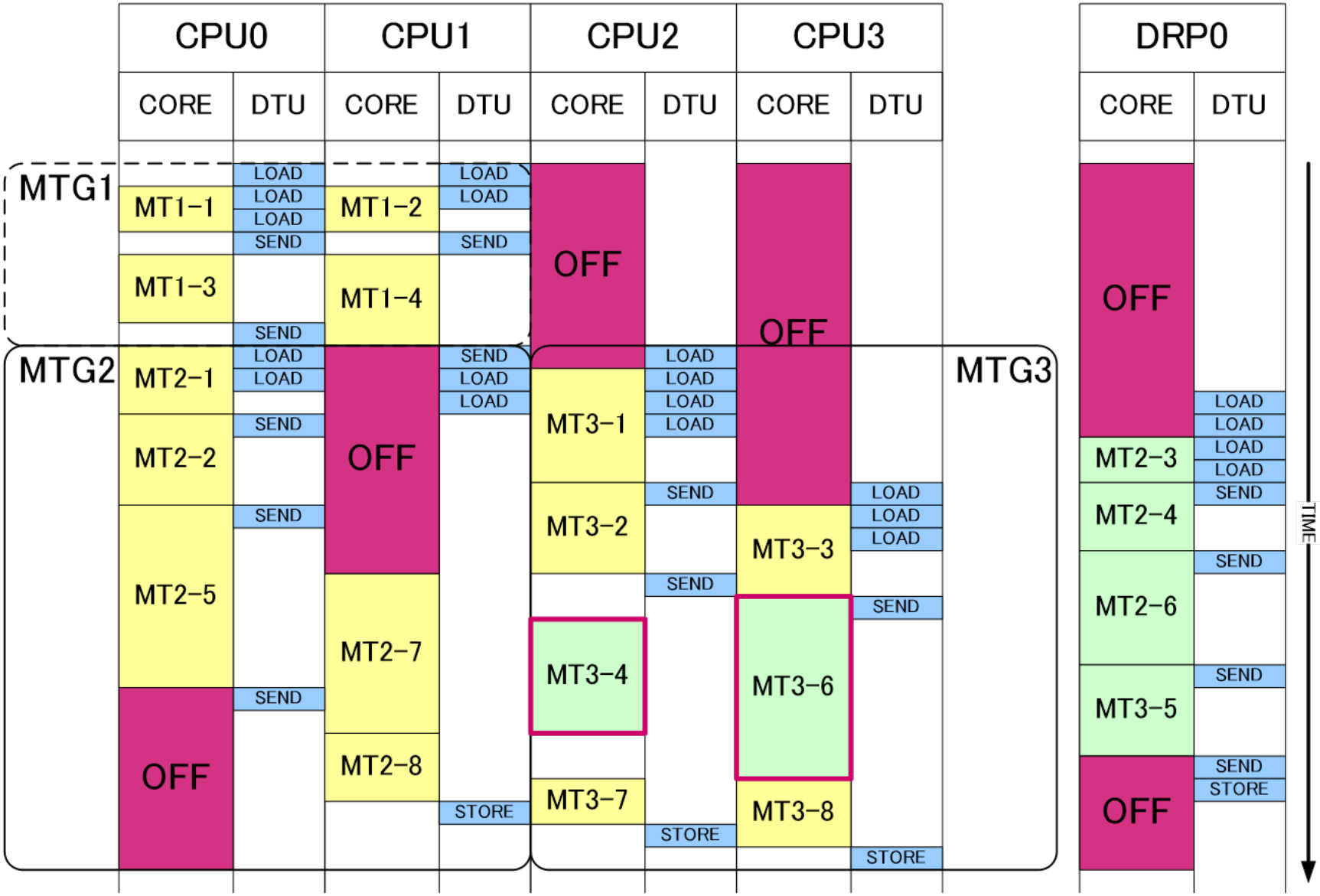
LDM
- Local Data Memory

DSM
- Distributed Shared Memory

CSM
- Centralized Shared Memory

FVR
- Frequency/Voltage Control Register

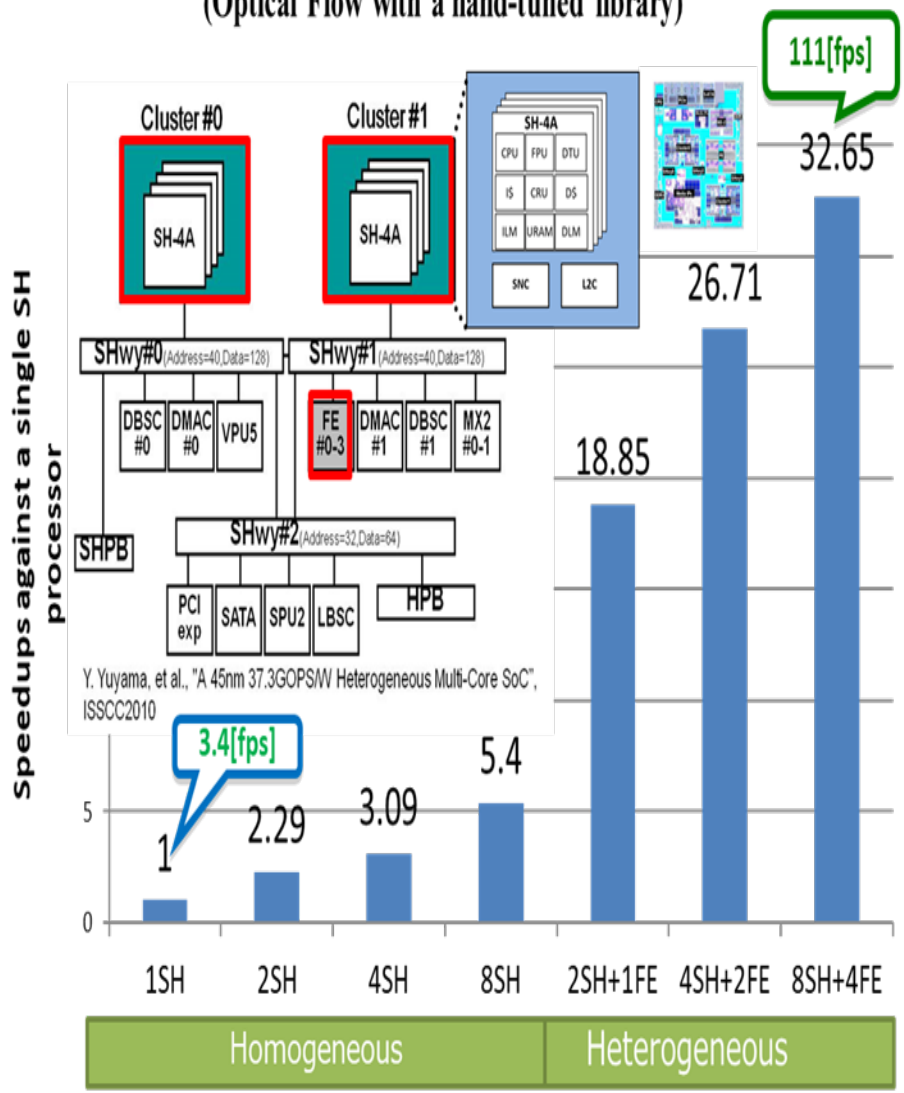
An Image of Static Schedule for Heterogeneous Multi-core with Data Transfer Overlapping and Power Control



Speedups and Power Reduction on RP-X Heterogeneous Multicore with 8 CPUs and 4 DRPs

33 Times Speedup Using OSCAR Compiler and API on Renesas RP-X with 8 CPUs & 4 DRP Accelerators

(Optical Flow with a hand-tuned library)



Power Reduction in a real-time execution controlled by OSCAR Compiler and OSCAR API on RP-X (Optical Flow with a hand-tuned library)

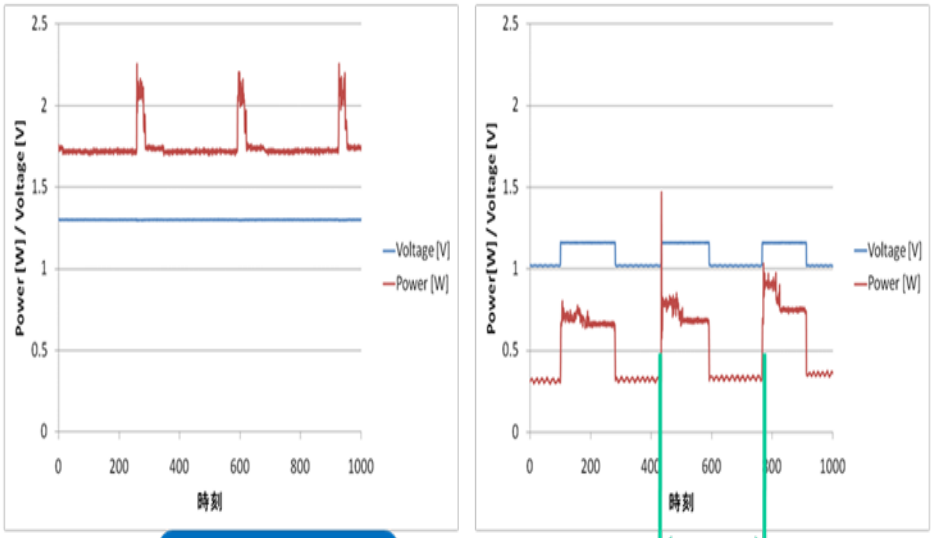
Without Power Reduction

With Power Reduction by OSCAR Compiler

70% of power reduction

Average:1.76[W]

Average:0.54[W]

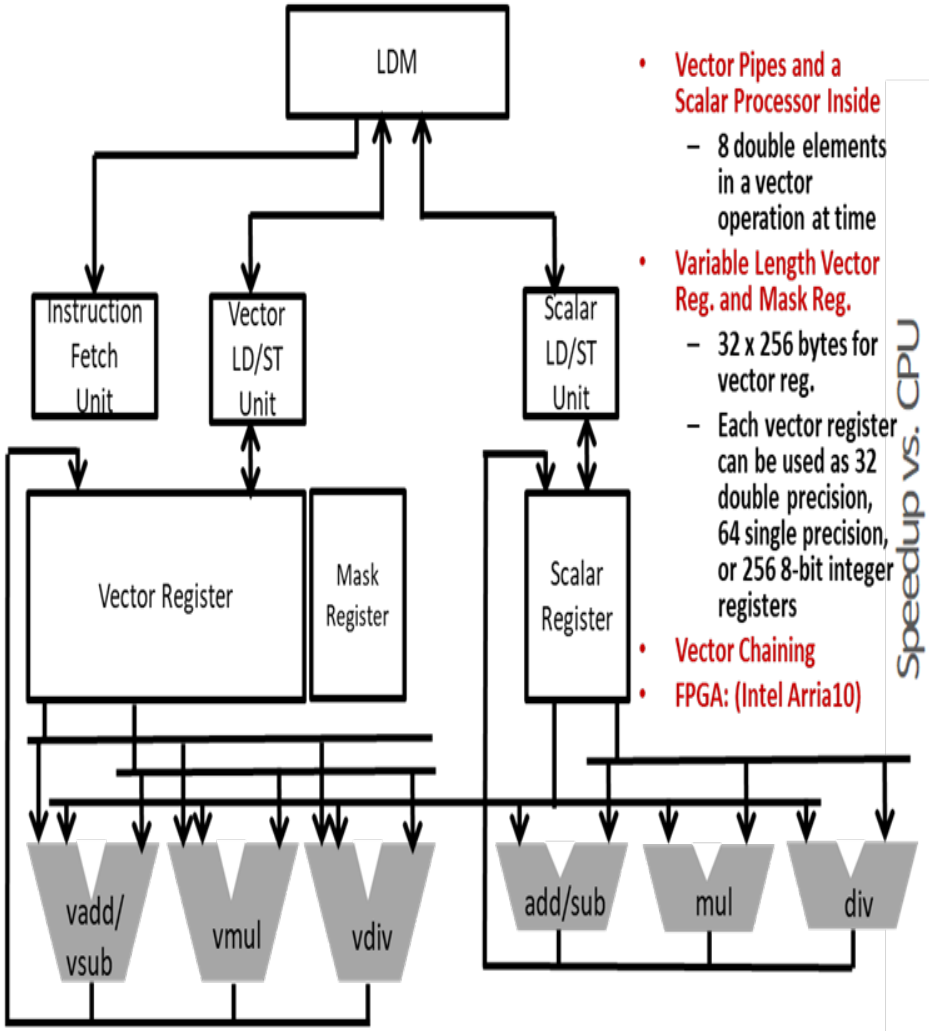


1cycle : 33[ms]
→30[fps]

Performance for Multimedia and Scientific Applications on OSCAR Vector Accelerator

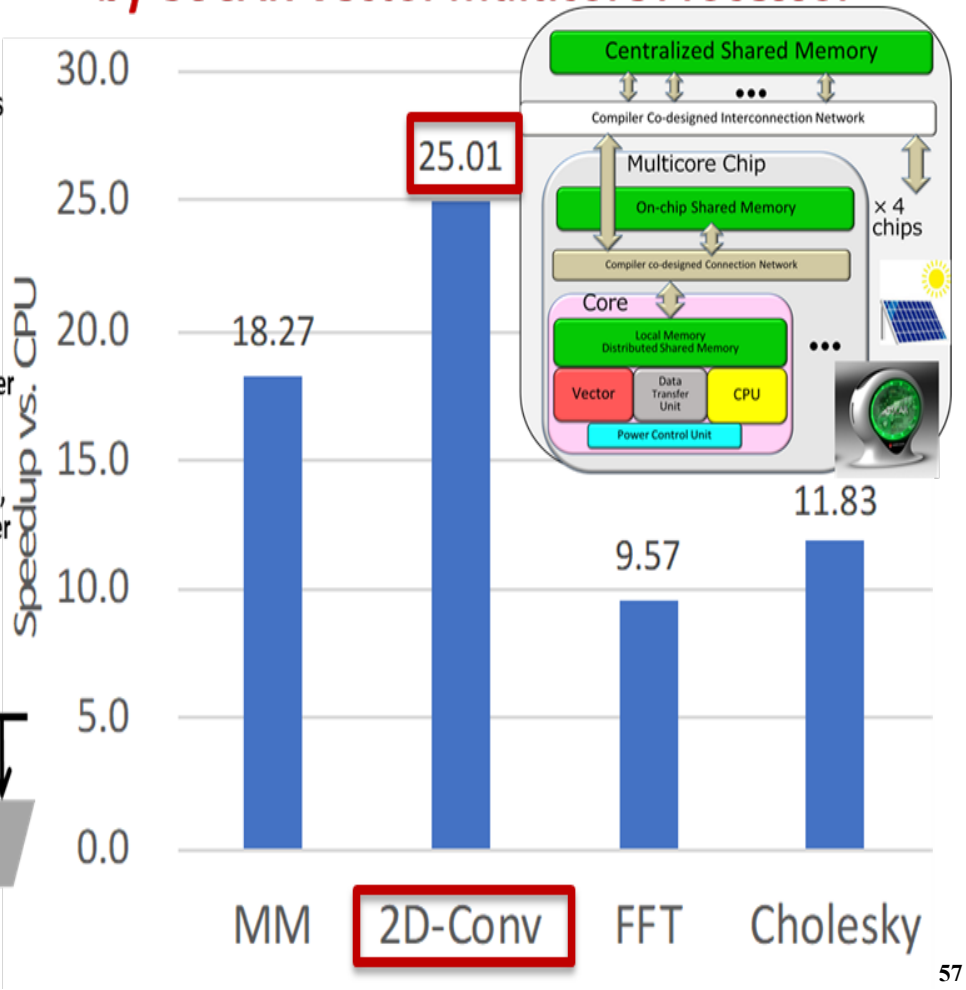
(A Vector Processor with Local Memory or Distributed Shared Memory and DMA Controller Managed by the OSCAR Compiler Redesigned Improving Japanese Supercomputer Technology in 1980-2000)

OSCAR Vector Accelerator on FPGA

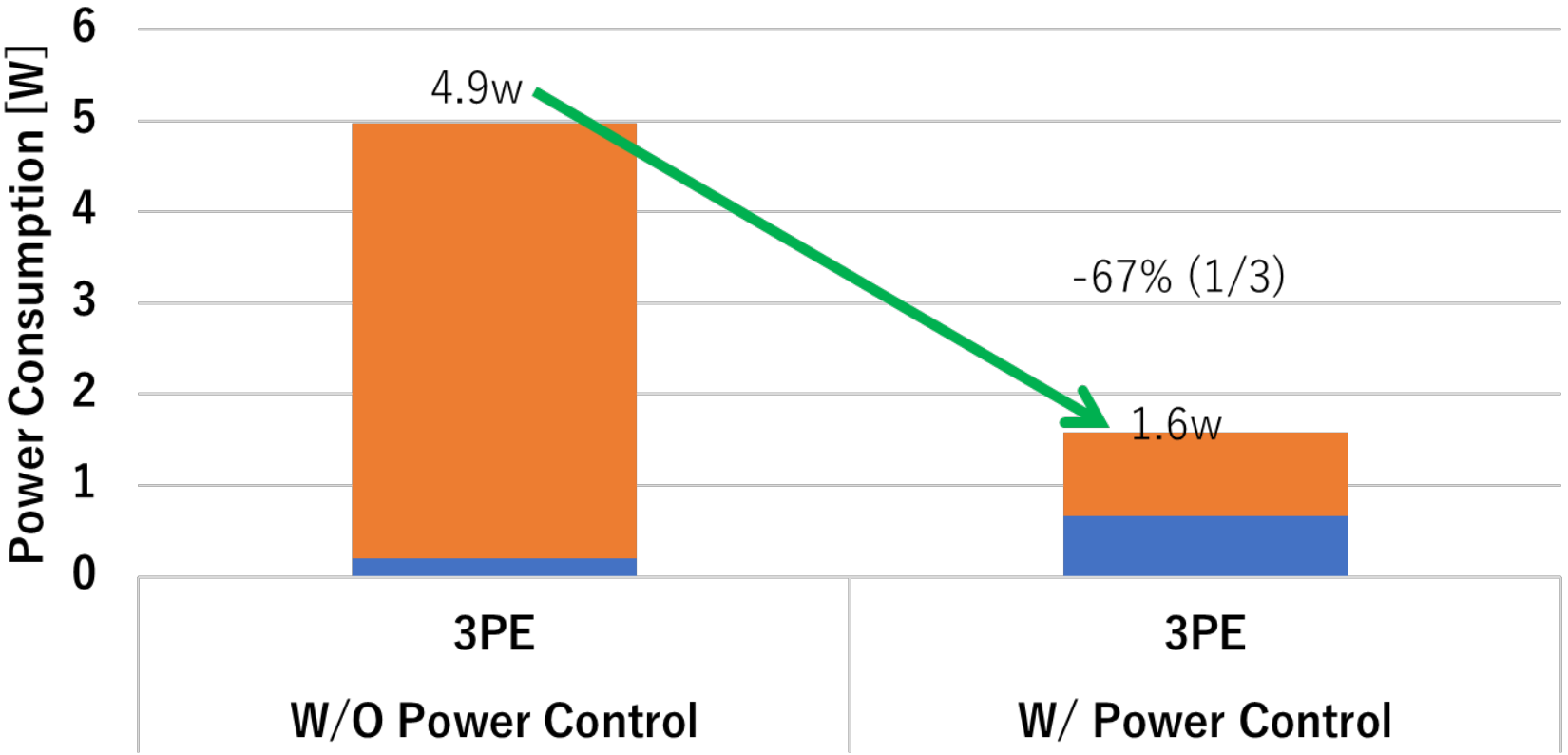


- Vector Pipes and a Scalar Processor Inside
  - 8 double elements in a vector operation at time
- Variable Length Vector Reg. and Mask Reg.
  - 32 x 256 bytes for vector reg.
  - Each vector register can be used as 32 double precision, 64 single precision, or 256 8-bit integer registers
- Vector Chaining
- FPGA: (Intel Arria10)

Speedups against General Purpose Processor by OSCAR Vector Multicore Processor



Automatic Power Reduction of OpenCV Face Detection on big.LITTLE ARM Processor

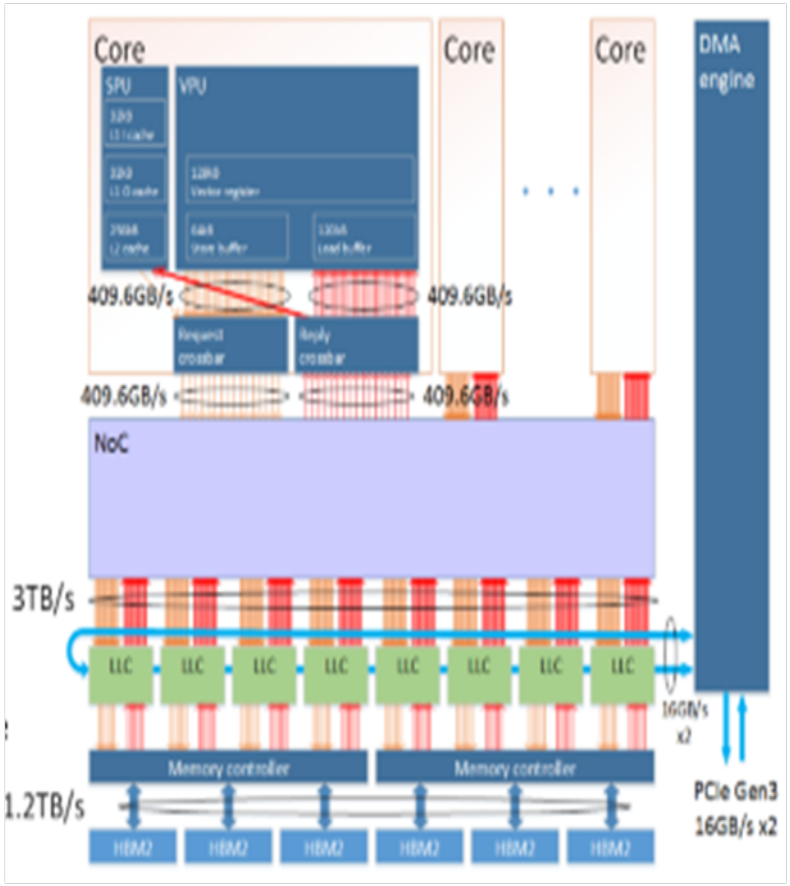


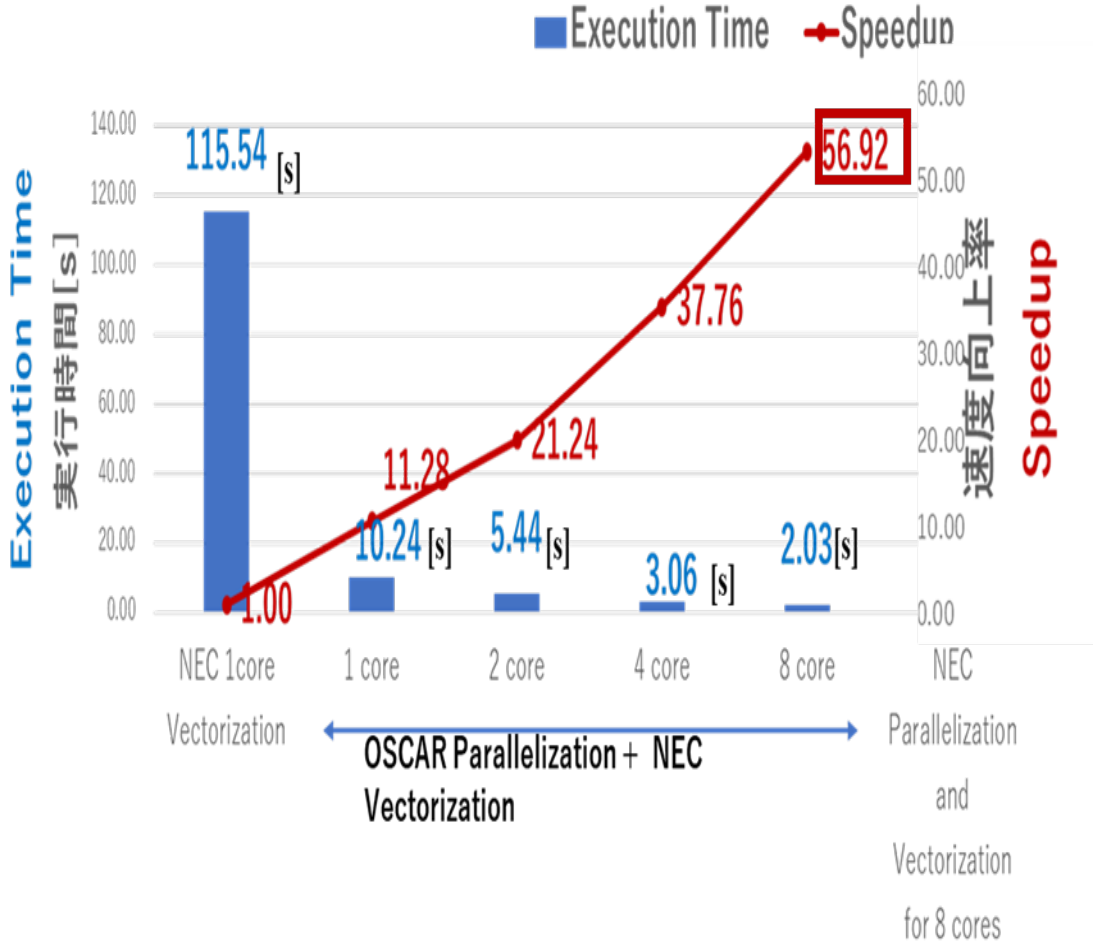
- ODROID-XU3
  - Samsung Exynos 5422 Processor
    - 4x Cortex-A15 2.0GHz, 4x Cortex-A7 1.4GHz big.LITTLE Architecture
    - 2GB LPDDR3 RAM

Frequency can be changed by each cluster unit

Speedups of NPB/CG Scientific Code by OSCAR Compiler on NEC SX-Aurora TSUBASA A100-1 8 cores 10C VE

57 times speedup for 8 vector cores by OSCAR Parallelization & NEC Vectorization against NEC 1 core Vectorization



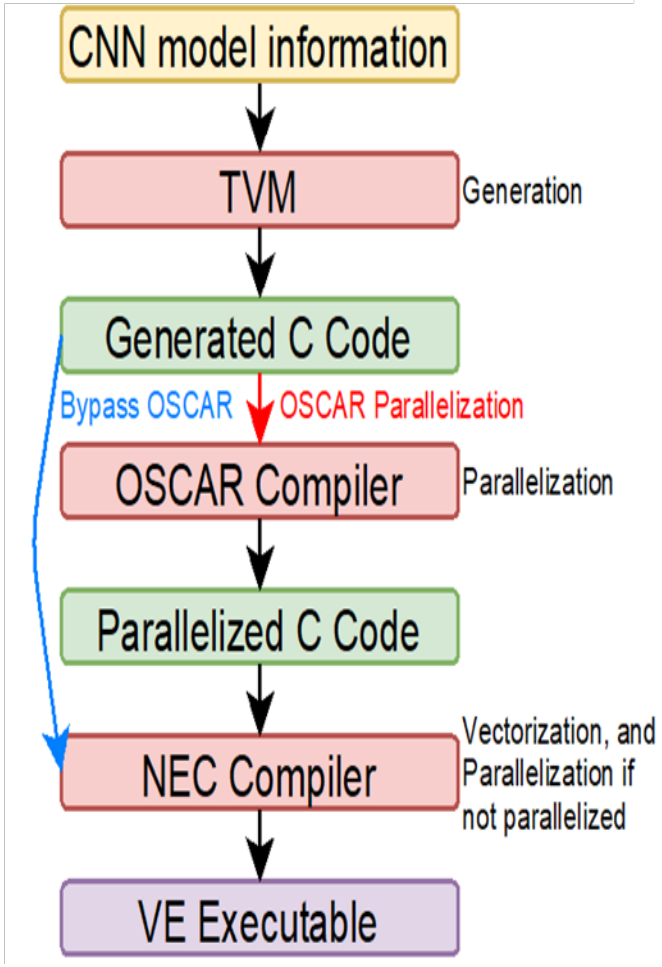


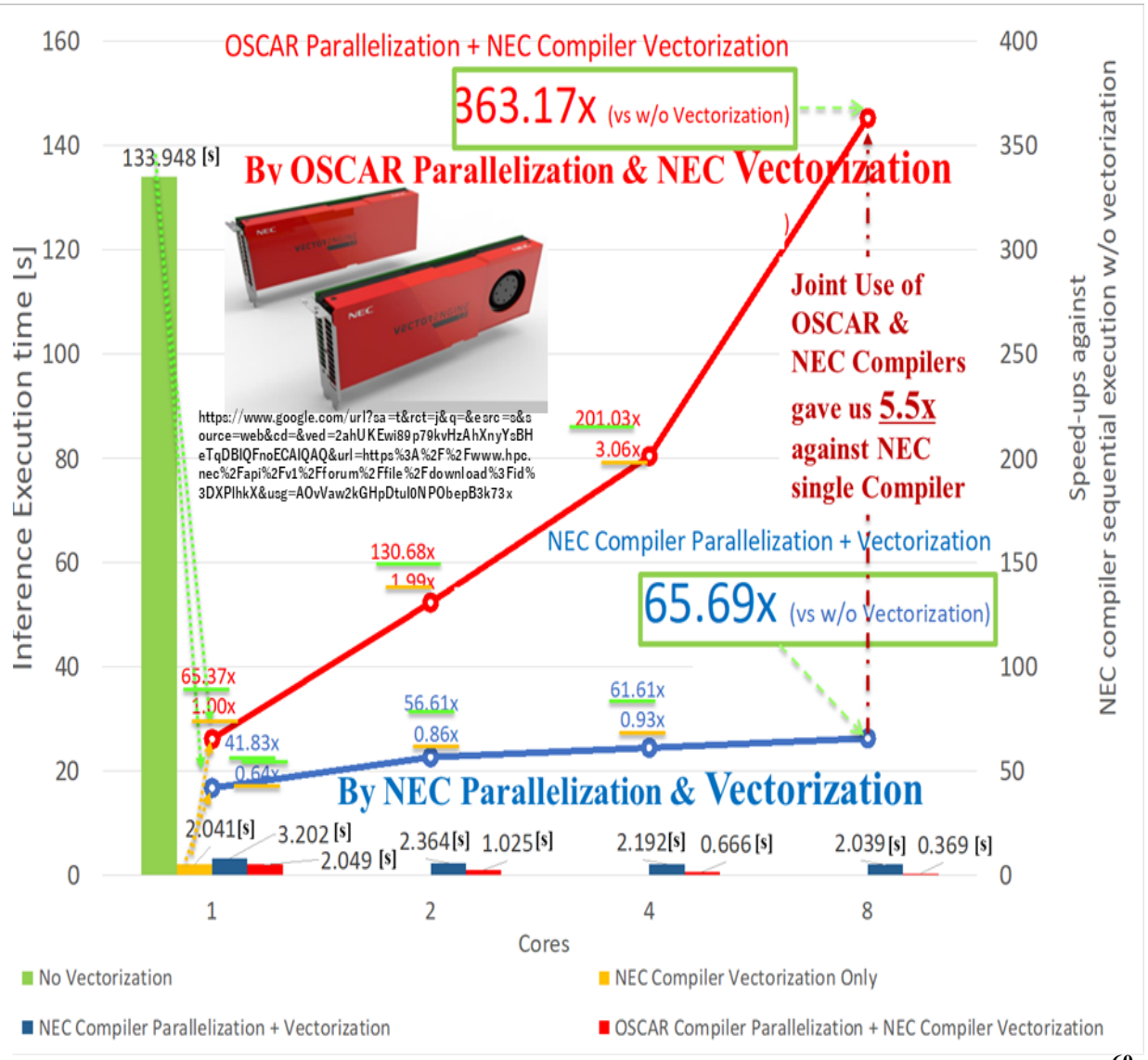
NEC Corporation. (2021) SX-Aurora TSUBASA Performance Tuning Guide. [Online]. Available: https://www.hpc.nec/documents/guide/pdfs/AuroraVETuningGuide.pdf

Speedups of Deep Learning Winograd 2D-Convolution generated by TVM on NEC Personal Vector Supercomputer SX-Aurora TSUBASA 8 Core Type 10C

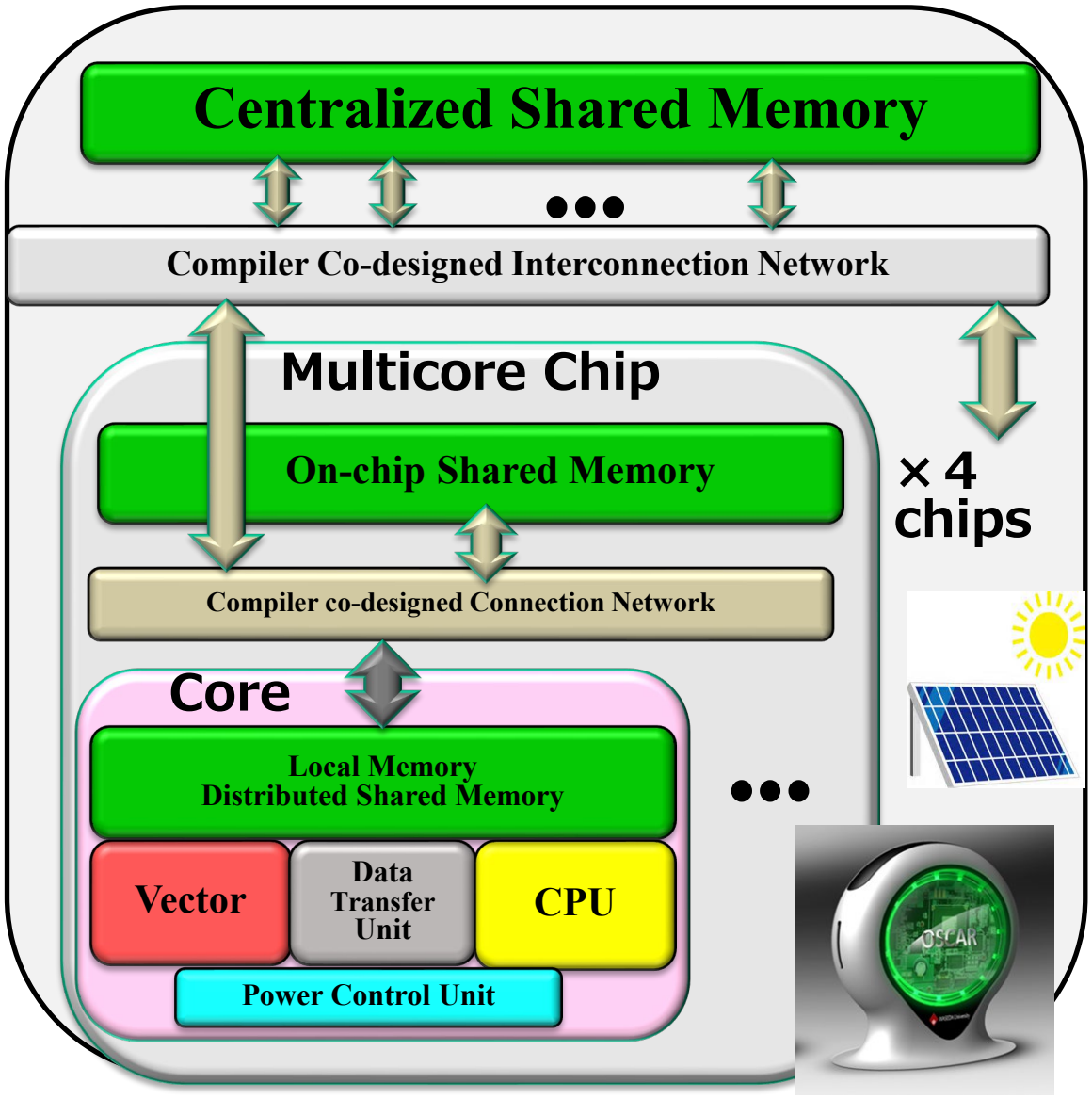
OSCAR Parallelization and NEC Vectorization gave us 363x Speedup against a Scalar Core

Parallelization of Deep Leaning C Code generated by TVM





OSCAR Vector Multicore and Compiler for Embedded to Severs with OSCAR Technology



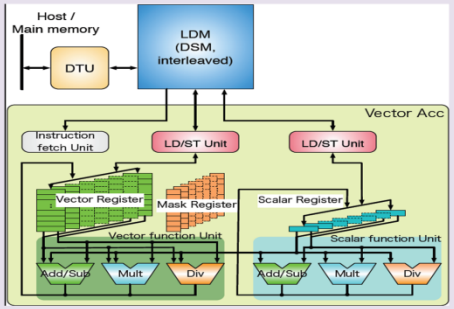
Target:

- **Solar Powered**
- **Compiler power reduction.**
- **Fully automatic parallelization and vectorization including local memory management and data transfer.**

Vector Accelerator

Features

- Attachable for any CPUs (Intel, ARM, IBM)
- Data driven initiation by sync flags



Function Units [tentative]

- Vector Function Unit
  - 8 double precision ops/clock
  - 64 characters ops/clock
  - Variable vector register length
  - Chaining LD/ST & Vector pipes
- Scalar Function Unit

Registers[tentative]

- Vector Register 256Bytes/entry, 32entry
- Scalar Register 8Bytes/entry
- Floating Point Register 8Bytes/entry
- Mask Register 32Bytes/entry

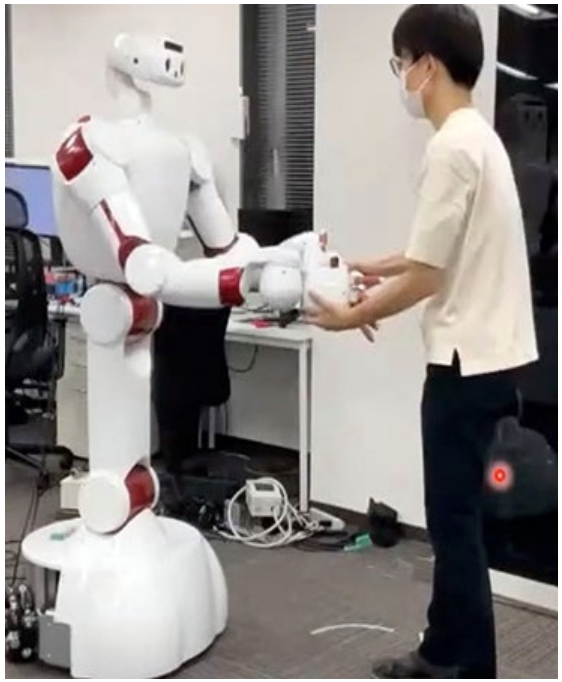
https://www8.cao.go.jp/cstp/english/moonshot/sub3_en.html

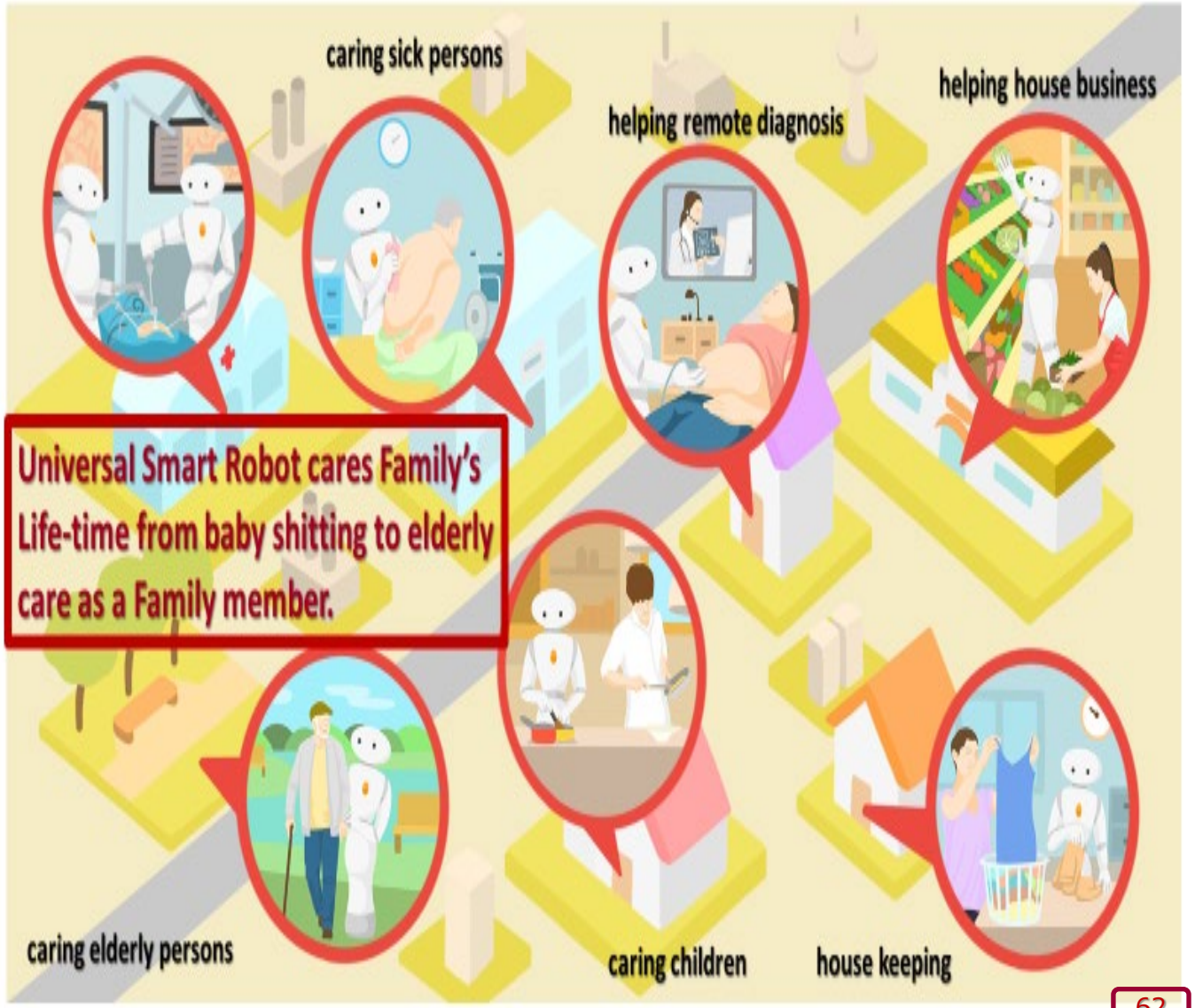
Cabinet Office

MOONSHOT
RESEARCH & DEVELOPMENT PROGRAM

AIREC (AI-driven Robot for Embrace and Care) Led by Prof. Sugano

Supported by Japanese Government "Moonshot" Project from 2020





Future Multicore Products with Automatic Parallelizing Compiler

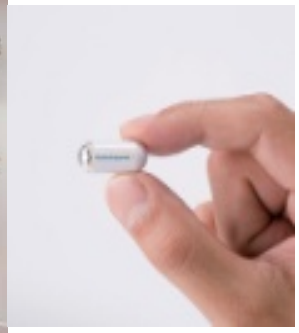
Next Generation Automobiles

- Safer, more comfortable, energy efficient, environment friendly
- Cameras, radar, car2car communication, internet information integrated brake, steering, engine, moter control

Smart phones

-From everyday recharging to less than once a week
- Solar powered operation in emergency condition
- Keep health

Advanced medical systems



Cancer treatment,
Drinkable inner camera

- Emergency solar powered
- No cooling fun, No dust , clean usable inside OP room

Personal / Regional Supercomputers

Solar powered with more than 100 times power efficient : FLOPS/W

- Regional Disaster Simulators saving lives from tornadoes, localized heavy rain, fires with earth quakes