OSCAR Automatic Parallelizing and Power Reducing Multicore Compiler for Realtime Embedded to High Performance Computing

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

IEEE Computer Society President 2018

Professor, Dept. of Computer Science & Engineering

Director, Advanced Multicore Processor Research Institute

Waseda University, Tokyo, Japan

URL: http://www.kasahara.cs.waseda.ac.jp/

1980 BS, 82 MS, 85 Ph.D. , Dept. EE, Waseda Univ.
1985 Visiting Scholar: U. of California, Berkeley
1986 Assistant Prof., 1988 Associate Prof., 1997, Waseda Univ., Now Dept. of Computer Sci. & Eng.
1989-90 Research Scholar: U. of Illinois, Urbana-Champaign, Center for Supercomputing R&D
2004 Director, Advanced Multicore Research Institute, 2017 member: the Engineering Academy of Japan and the Science Council of Japan

2005 STARC Academia-Industry Research Award
2008 LSI of the Year Second Prize
2008 Intel AsiaAcademic Forum Best Research Award
2010 IEEE CS Golden Core Member Award
2014 Minister of Edu., Sci. & Tech. Research Prize
2015 IPSJ Fellow
2017 IEEE Fellow, IEEE Eta Kappa Nu

Reviewed Papers: 216, Invited Talks: 155, Published Unexamined Patent Application:59 (Japan, US, GB, China Granted Patents: 30), Articles in News Papers, Web News, Medias incl. TV etc.: 578

Committees in Societies and Government 255
IEEE Computer Society President 2018, BoG(2009-14), Multicore STC Chair (2012-), Japan Chair (2005-07), IPSJ Chair: HG for Mag. & 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, etc.

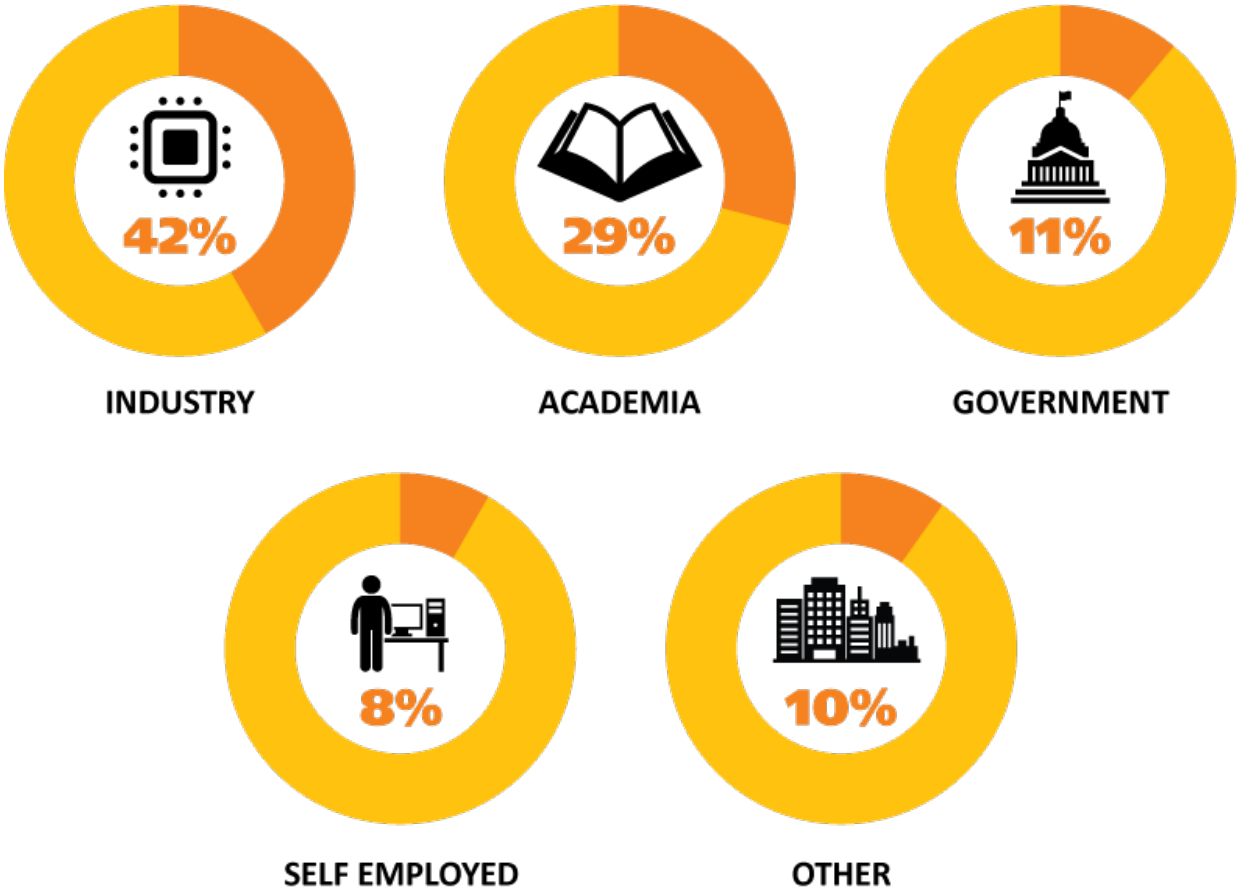
IEEE Computer Society

60,000+ members, volunteer-led organization,
200 technical conferences, 17 scholarly journals
and **13 magazines, awards program,**
Digital Library with more than 550,000 articles and papers ,
400 local and regional chapters, 40 technical committees,

- Region 1 (Northeastern US)
- Region 2 (Eastern US)
- Region 3 (Southern US)
- Region 4 (Central US)
- Region 5 (Southwestern US)
- Region 6 (Western US)
- Region 7 (Canada)
- Region 8 (Africa, Europe, Middle East)
- Region 9 (Latin America)
- Region 10 (Asia and Pacific)

IEEE-USA (Regions 1-6)

IEEE Computer Society Members



IEEE Computer Society BoG (Board of Governors) Feb.1, 2018

https://www.computer.org/web/cshistory/officers-2018

Past IEEE Computer Society Presidents

Chairs of the IRE Professional Group on Electronic Computers

1951-53 Morton M. Astrahan
1953-54 John H. Howard
1954-55 Harry Larson
1955-56 Jean H. Felker
1956-57 Jerre D. Noe
1957-58 Werner Buchholz
1958-59 Willis H. Ware
1959-60 Richard O. Endres
1960-62 Arnold A. Cohen
1962-64 Walter L. Anderson

Chairs of the AIEE Committee on Large-Scale Computing Devices

1946-49 Charles Concordia
1949-51 John Grist Brainerd
1951-53 Walter H. MacWilliams
1953-55 Frank J. Maginniss
1955-57 Edwin L. Harder
1957-59 Morris Rubinoff
1959-61 Ruben A. Imm
1961-63 Claude A. Kagan
1963-64 Gerhard L. Hollander

Chairs & Presidents of the IEEE Computer Society

1964-65 Keith Uncapher
1965-66 Richard I. Tanaka
1966-67 Samuel Levine
1968-69 Charles L. Hobbs
1970-71 Edward J. McCluskey
1972-73 Albert S. Hoagland
1974-75 Stephen S. Yau
1976 Dick B. Simmons
1977-78 Merlin G. Smith
1979-80 Tse-Yun Feng
1981 Richard E. Merwin
1982-83 Oscar N. Garcia
1984-85 Martha Sloan
1986-87 Roy L. Russo
1988 Edward A. Parrish
1989 Kenneth A. Anderson
1990 Helen M. Wood
1991 Duncan H. Lawrie
1992 Bruce D. Shriver
1993 James H. Aylor
1994 Laurel V. Kaleda
1995 Ronald G. Hoelzeman
1996 Mario R. Barbacci
1997 Barry W. Johnson
1998 Doris L. Carver
1999 Leonard L. Tripp
2000 Guylaine M. Pollock
2001 Benjamin W. Wah
2002 Willis K. King
2003 Stephen Diamond
2004 Carl K. Chang
2005 Gerald L. Engel
2006 Deborah M. Cooper
2007 Michael R. Williams
2008 Rangachar Kasturi
2009 Susan K. (Kathy) Land,
2010 James D. Isaak
2011 Sorel Reisman
2012 John W. Walz
2013 David Alan Grier
2014 Dejan S. Milojicic
2015 Thomas M. Conte
2016 Roger U. Fujii
2017 Jean-Luc Gaudiot
2018 Hironori Kasahara

IPSJ/IEEE-CS Young Computer Researcher Award

For members of the IPSJ and the IEEE-CS

The First Award Ceremony:

COMPSAC2018, July 23-27, NII, Tokyo

https://ieeecompsac.computer.org/2018/

Bjarne Stroustrup
2018 Computer Society Computer Pioneer Award
Columbia University

Masaru Kitsuregawa
Director General of NII, Past President of IPSJ

Margaret Martonosi
2018 Computer Society Technical Achievement
Princeton University

Dejan Milojicic
CS President 2014
HP Labs
CS 2022 Report

IEEE

IEEE Computer Society Membership

Join Your Peers

To share knowledge, solve problems and network

Be A Leader

Participate on a national community, committee, or special interest groups

Access Trusted, Reliable Information

Computer Magazine & ComputingEdge Magazine Included. Member only access to Computer Society Digital Library.

Save Money

Member receive discounted pricing on IEEE Computer Society Conferences and Events and additional member subscriptions.

Develop Skills

Online education, certification preparation, webinars, face to face meetings

Choose Your Content Bundle

	Preferred Plus More Info	Training & Development More Info	Research More Info	Basic More Info	Student More Info
IEEE Member: Add Computer Society to your existing IEEE membership. IEEE Membership is an additional charge. **Affiliate Member:** Join Computer Society only as a Computer Society affiliate. IEEE benefits are not included. **Student Member:** Students must join IEEE when joining Computer Society. IEEE Student members can add Computer Society.	**IEEE Member: $30** **Affiliate Member: $78** JOIN NOW	**IEEE Member: $28** **Affiliate Member: $71** JOIN NOW	**IEEE Member: $28** **Affiliate Member: $71** JOIN NOW	**IEEE Member: $20** **Affiliate Member: $63** JOIN NOW	**IEEE Student Member: $4** **New Student Member: $20** JOIN NOW
Computer magazine (12 digital issues)*	✓	✓	✓	✓	✓
ComputingEdge (12 issues)	✓	✓	✓	✓	✓
Members-only discounts on conferences and events	✓	✓	✓	✓	✓
Members-only webinars	✓	✓	✓	✓	✓
Unlimited access to Computing Now, computer.org, and the new mobile-ready myCS	✓	✓	✓	✓	✓
Local chapter membership	✓	✓	✓	✓	✓
Skillsoft Skillchoice™ Complete with 67,000+ books, videos, courses, practice exams and mentorship resources	✓	✓			✓
Books24x7 on-demand access to 15,000 technical and business resources	✓	✓			✓
Two complimentary Computer Society magazines subscriptions	✓		✓		
myComputer mobile app	30 tokens		30 tokens		30 tokens
Computer Society Digital Library	12 FREE downloads	Member pricing	12 FREE downloads	Member pricing	Included
Training webinars	3 FREE webinars	3 Free webinars	Member pricing	Member pricing	Member pricing
Priority registration to Computer Society events	✓				
Right to vote and hold office	✓	✓	✓	✓	
One-time 20% Computer Society online store discount	✓				

** Print publications are available for an additional fee. See IEEE catalog for details.*

https://www.computer.org/web/education/multicore-video-series#

IEEE computer society

Home // Professional Education // Certification Credentials // Certificates of Achievement // Multicore Video Series

Multicore Video Series

- Automatic Parallelization: David Padua
- Autoparallelization for GPUs: Wen-mei Hwu
- Dependences and Dependence Analysis: Utpal Banerjee
- Dynamic Parallelization: Rudolf Eigenmann
- Instruction Level Parallelization: Alexandru Nicolau
- Multigrain Parallelization and Power Reduction: Hironori Kasahara
- The Polyhedral Model: Paul Feautrier
- Vector Computation: David Kuck (Computer Pioneer)
- Vectorization: P. Sadayappan
- Vectorization/Parallelization in the IBM Compiler: Yaoqing Gao
- Vectorization/Parallelization in the Intel Compiler: Peng Tu
- Roundtable Discussion by all presenters

Self-Paced Learning:

Approximate time = 12 hours

- PDH: 12.0
- CEU: 1.2

Full Series Price:

- IEEE CS Member: $195
- Nonmember: $1,000

Individual Videos:

- IEEE CS Member: $30
- Nonmember: $125

See individual videos below.

For questions, please contact certification@computer.org.

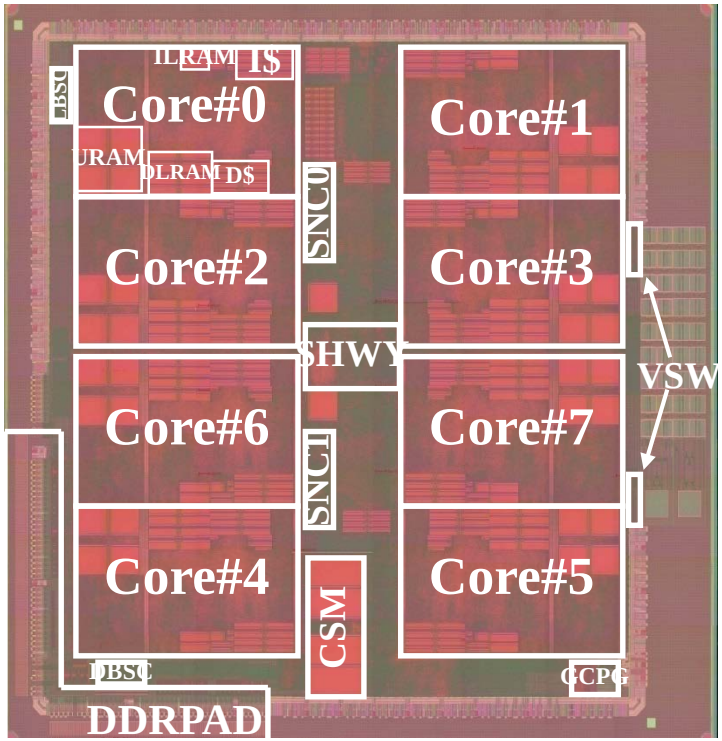
IEEE

Toward 2018

1. Refining content and services to further improve the satisfaction of CS members;
2. Considering an incentive for volunteers to further accelerate CS activities and promptly provide technical benefits for people around the globe;
   To express appreciation to volunteers:
   CS Point (Mileage) System: Annual & Life Time Honor,
   Premier Seating, Premier Registration, Distinguished Reviewer, etc
3. Offering more attractive services for practitioners in industry;
4. Providing the world's best educational content and historical treasures for future generations, which only the CS can create with our pioneering researchers (for example, the Multicore Compiler Video Series found at www.computer.org/web/education/multicore-video-series);
5. Thinking about sustainable membership fees while considering the diversity of economic situations within the 10 regions;
6. Cooperating with other IEEE societies and sister societies in a timely and efficient manner;
7. Intelligibly introducing the latest computer-related technologies to younger generations, including children, so that they can realize their technological dreams.

Multicores for Performance and Low Power

Power consumption is one of the biggest problems for performance scaling from smartphones to cloud servers and supercomputers (“K” more than 10MW) .



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 .

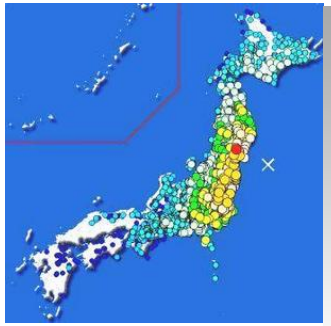
<Multicores>

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

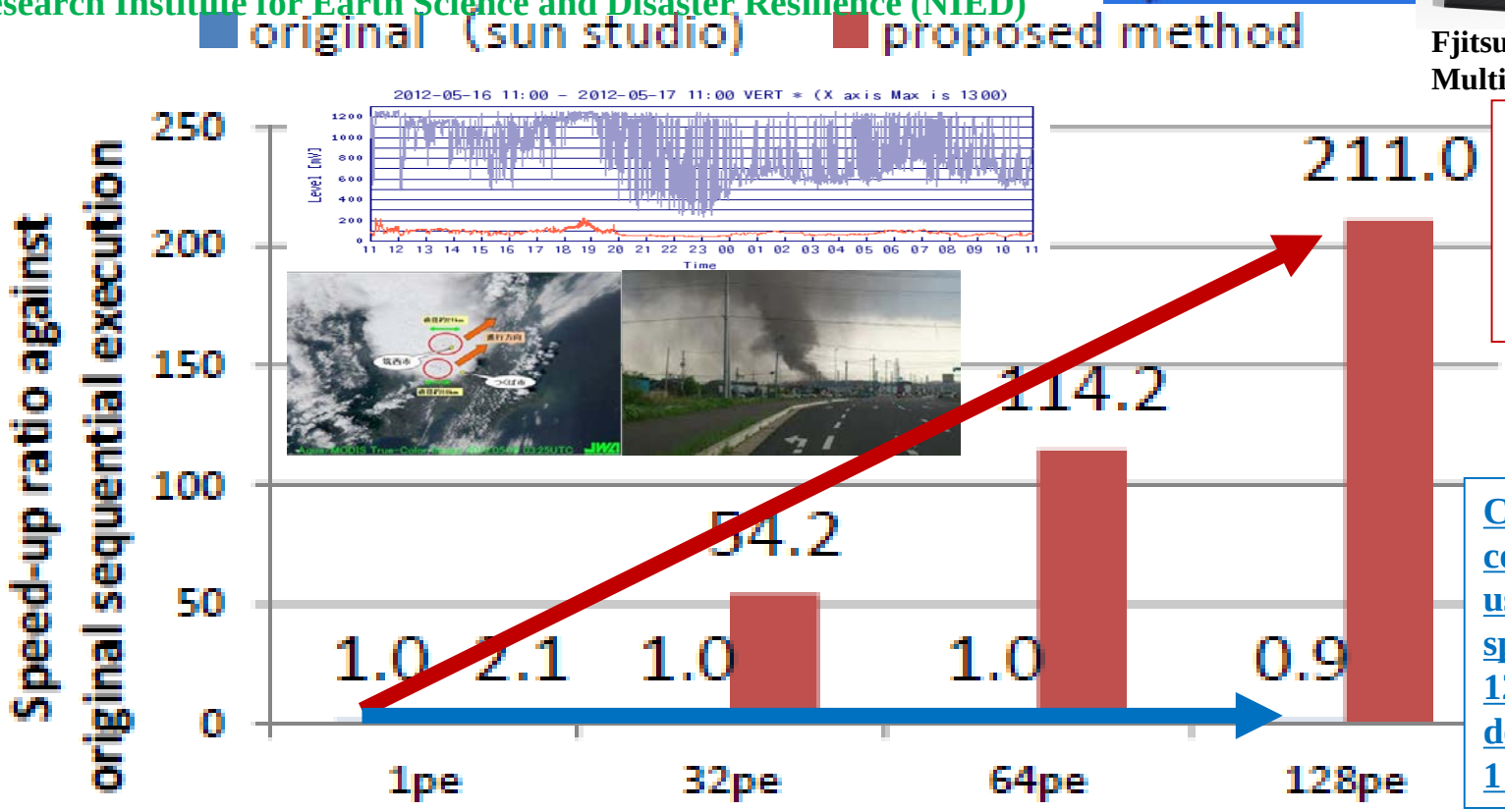
Parallel Soft is important for scalable performance of multicore (LCPC2015)

- 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

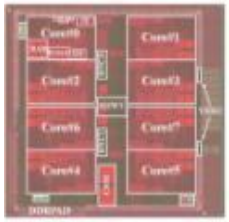


OSCAR Compiler gives us 211 times speedup with 128 cores

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

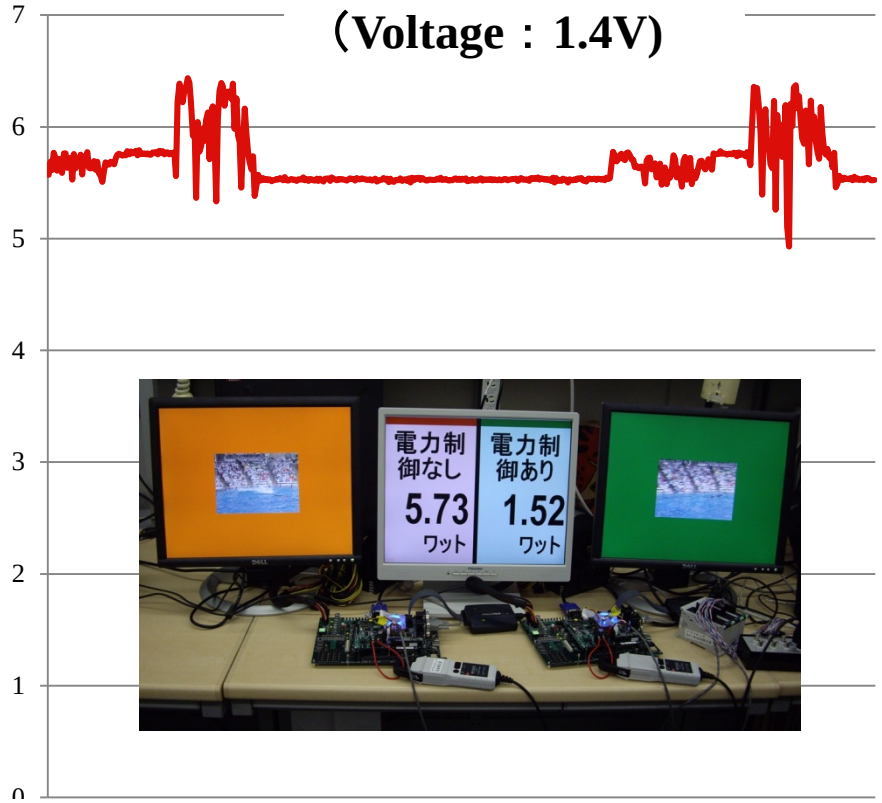
- Automatic parallelizing compiler available on the market gave us no speedup against execution time on 1 core on 64 cores
  - Execution time with 128 cores was slower than 1 core (0.9 times speedup)
- Advanced OSCAR parallelizing compiler gave us 211 times speedup with 128cores against execution time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

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

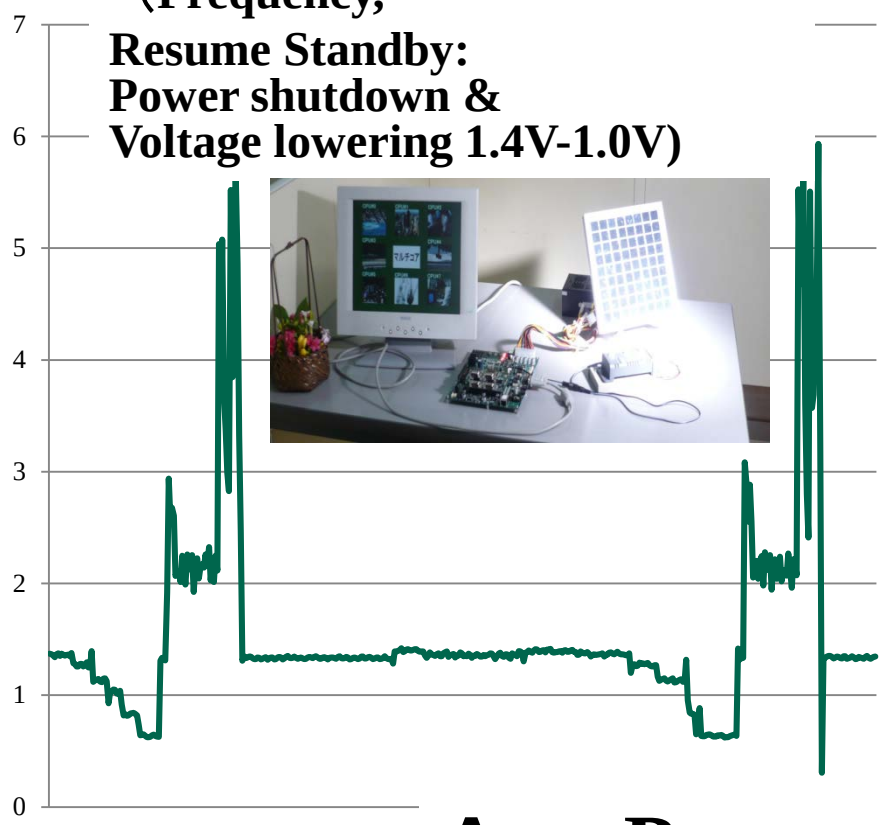


MPEG2 Decoding with 8 CPU cores

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]

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)

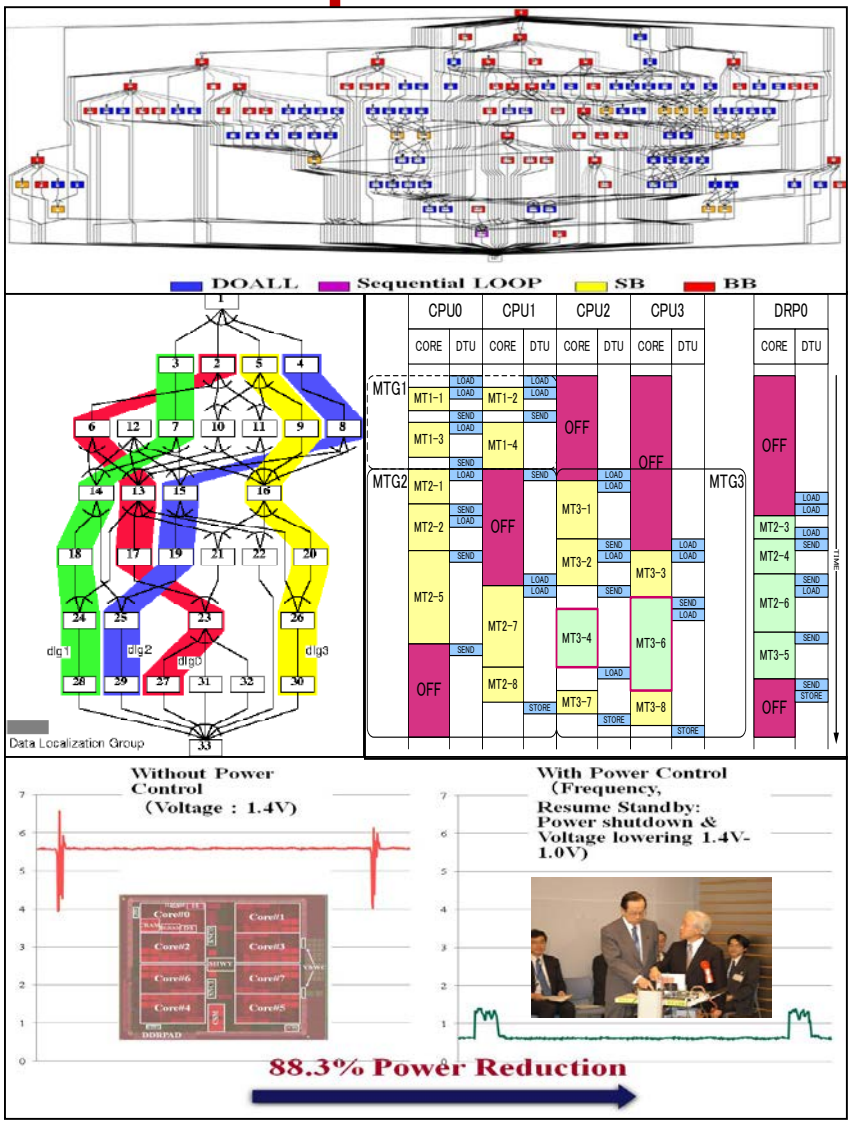
Data Transfer Overlapping (2016 partially)

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.



Multicore Program Development Using 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
- Coarse grain task parallelization
- Data Localization
- DMAC data transfer
- Power reduction using DVFS, Clock/ Power gating

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

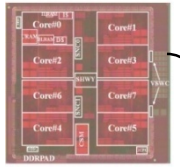
Parallelized API F or C program
- Proc0: Code with directives, Thread 0
- Proc1: Code with directives, Thread 1
- Accelerator 1: Code
- Accelerator 2: Code

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

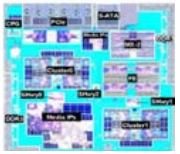
Low Power Heterogeneous Multicore Code Generation
API Analyzer (Available from Waseda) | Existing sequential compiler

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

Shred memory servers

Executable on various multicores

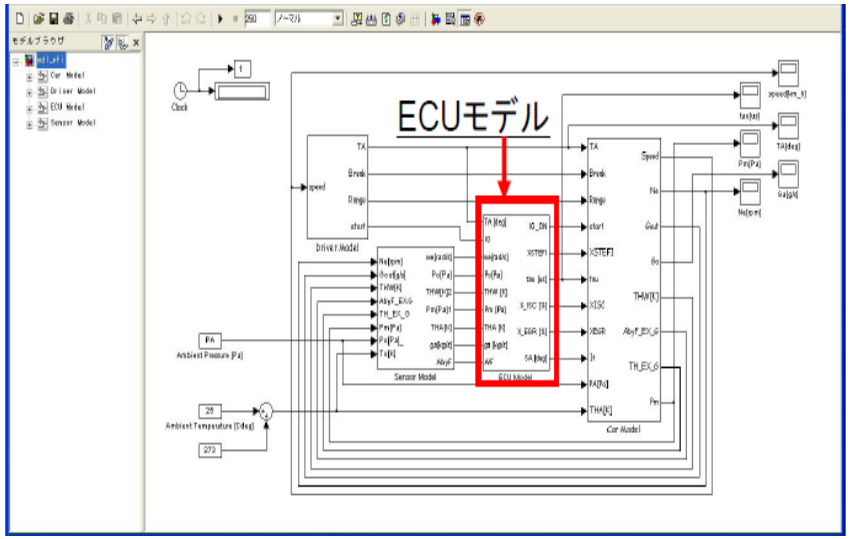
Hitachi, Renesas, NEC, Fujitsu, Toshiba, Denso, Olympus, Mitsubishi, Esol, Cats, Gaio, 3 univ.

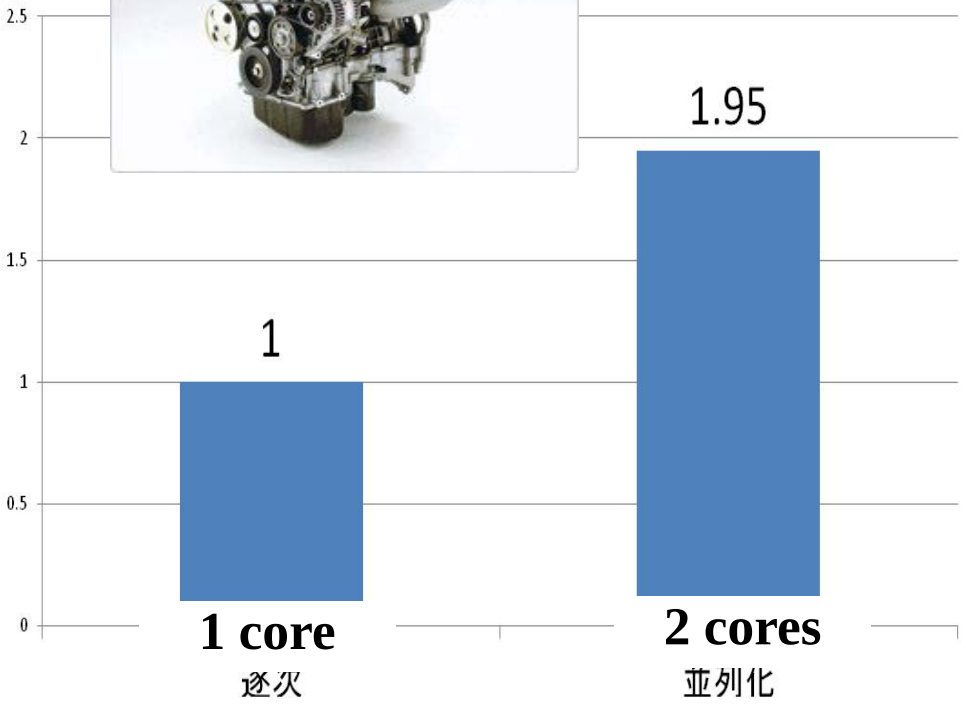
OSCAR: Optimally Scheduled Advanced Multiprocessor
API : Application Program Interface

Engine Control by multicore with Denso

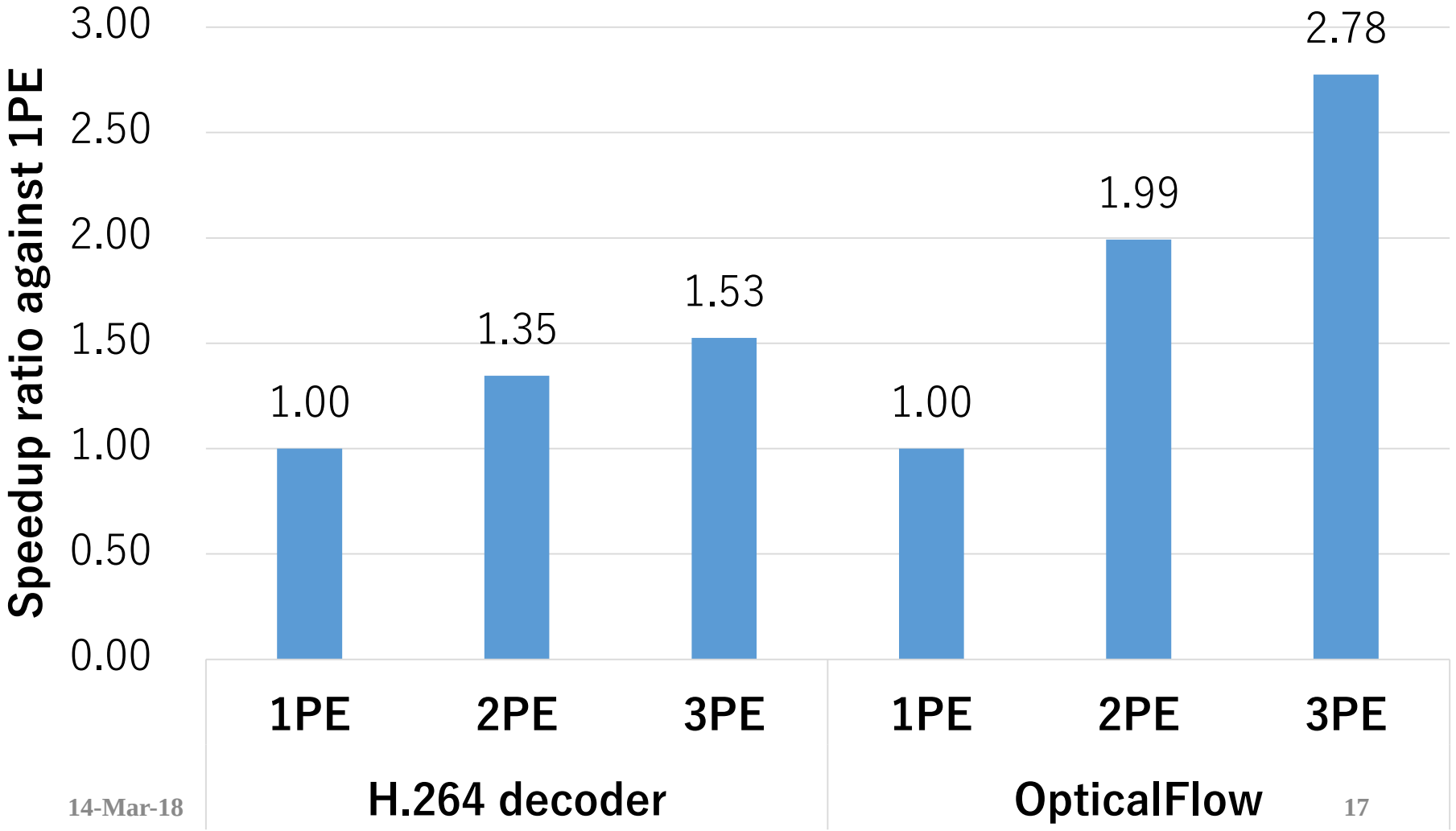
Though so far parallel processing of the engine control on multicore has been very difficult, Denso and Waseda succeeded 1.95 times speedup on 2core V850 multicore processor.

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

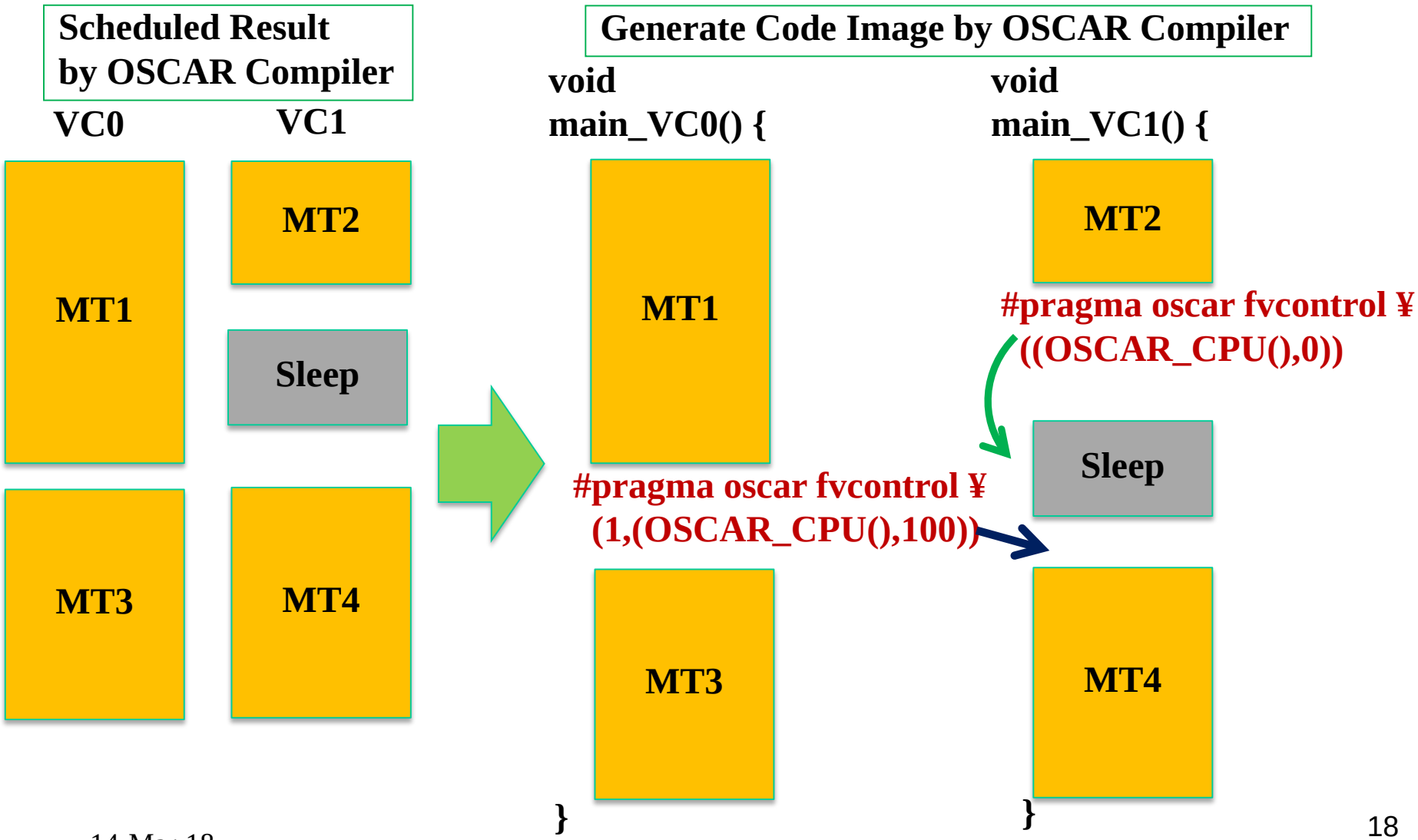




Speedup ratio for H.264 and Optical Flow on ARM Cortex-A9 Android 3 cores by OSCAR Automatic Parallelization

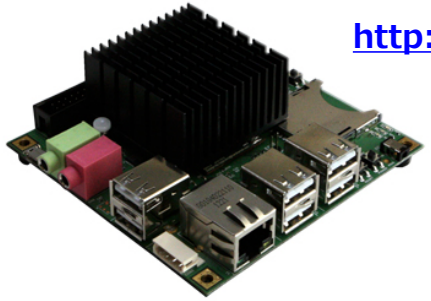


Low-Power Optimization with OSCAR API



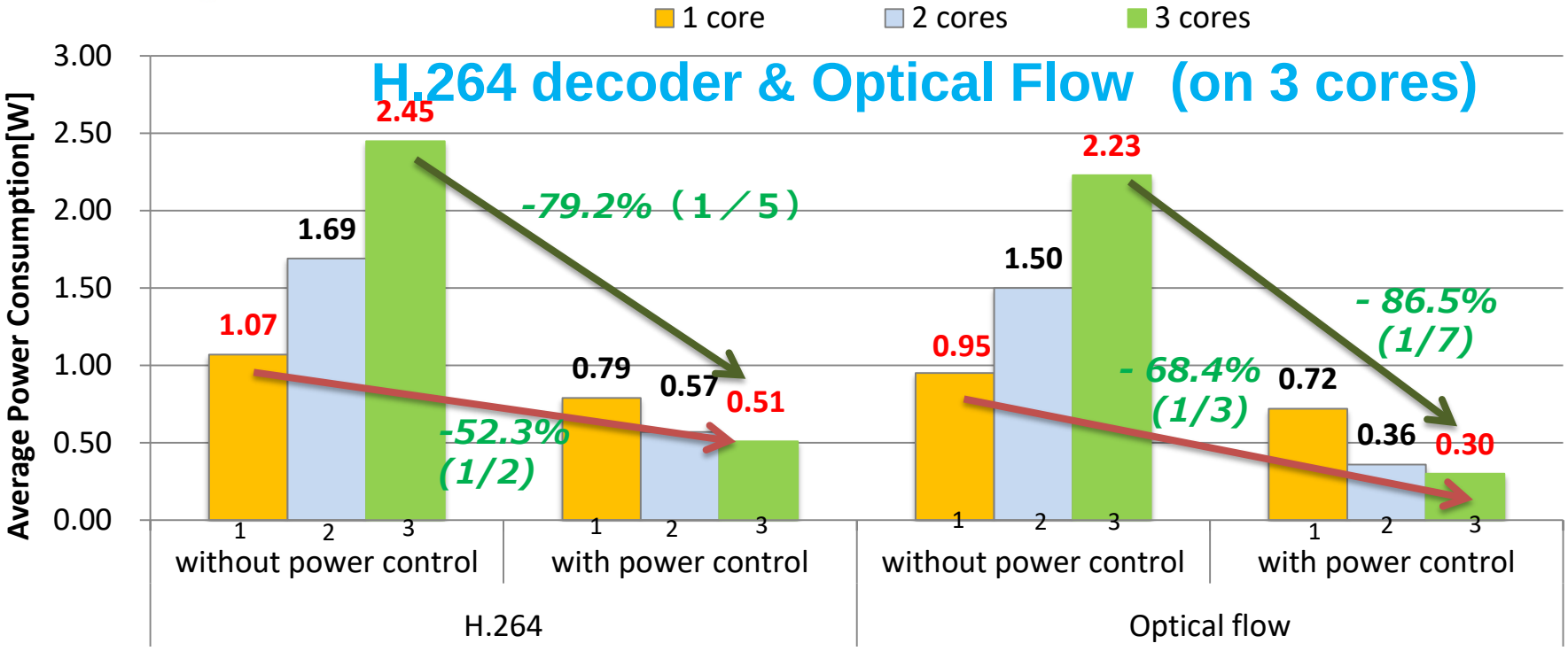
Automatic Power Reduction on ARM CortexA9 with Android

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

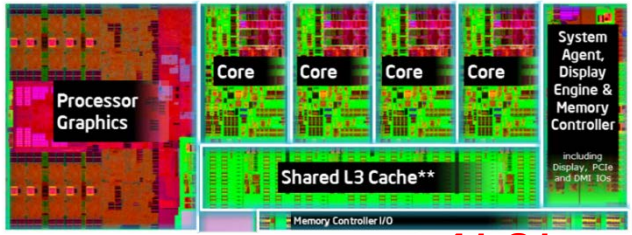


Power for 3cores was reduced to 1/5～1/7 against without software power control

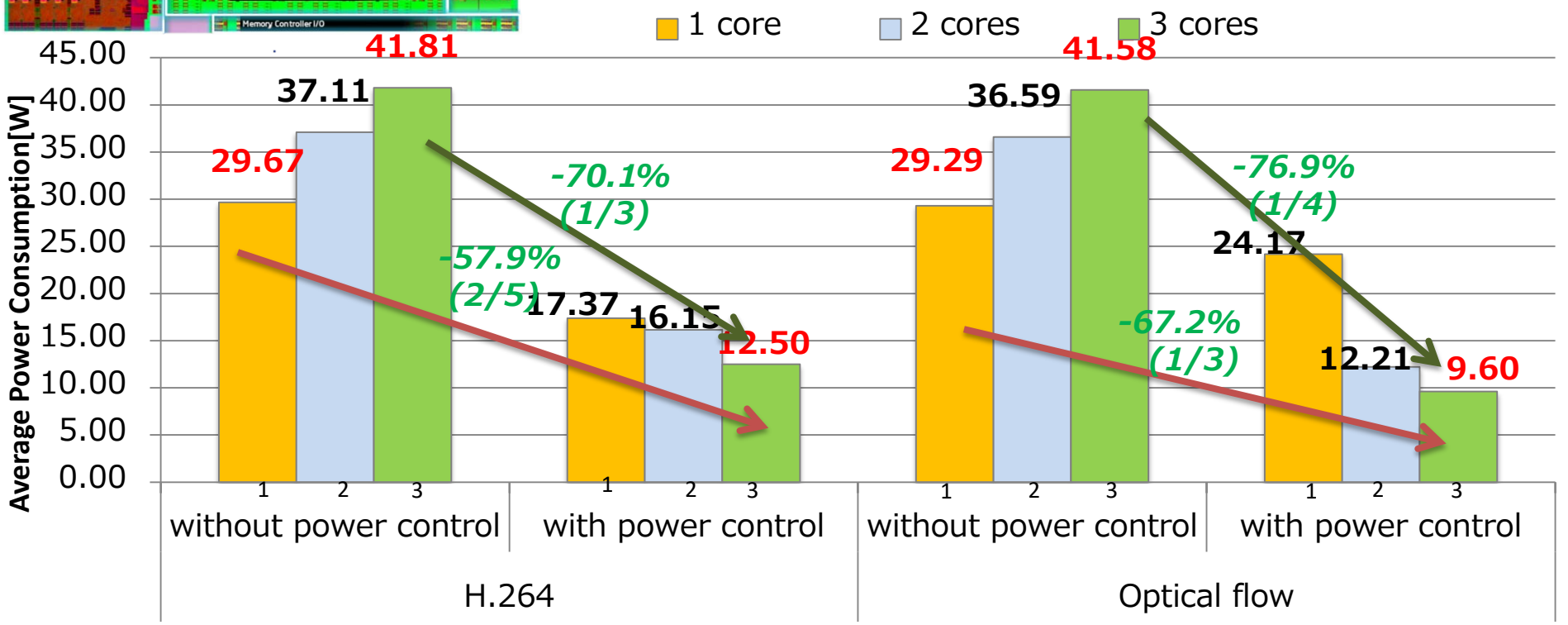
Power for 3cores was reduced to 1/2～1/3 against ordinary 1core execution

Automatic Power Reuction on Intel Haswell

H.264 decoder & Optical Flow (3cores)

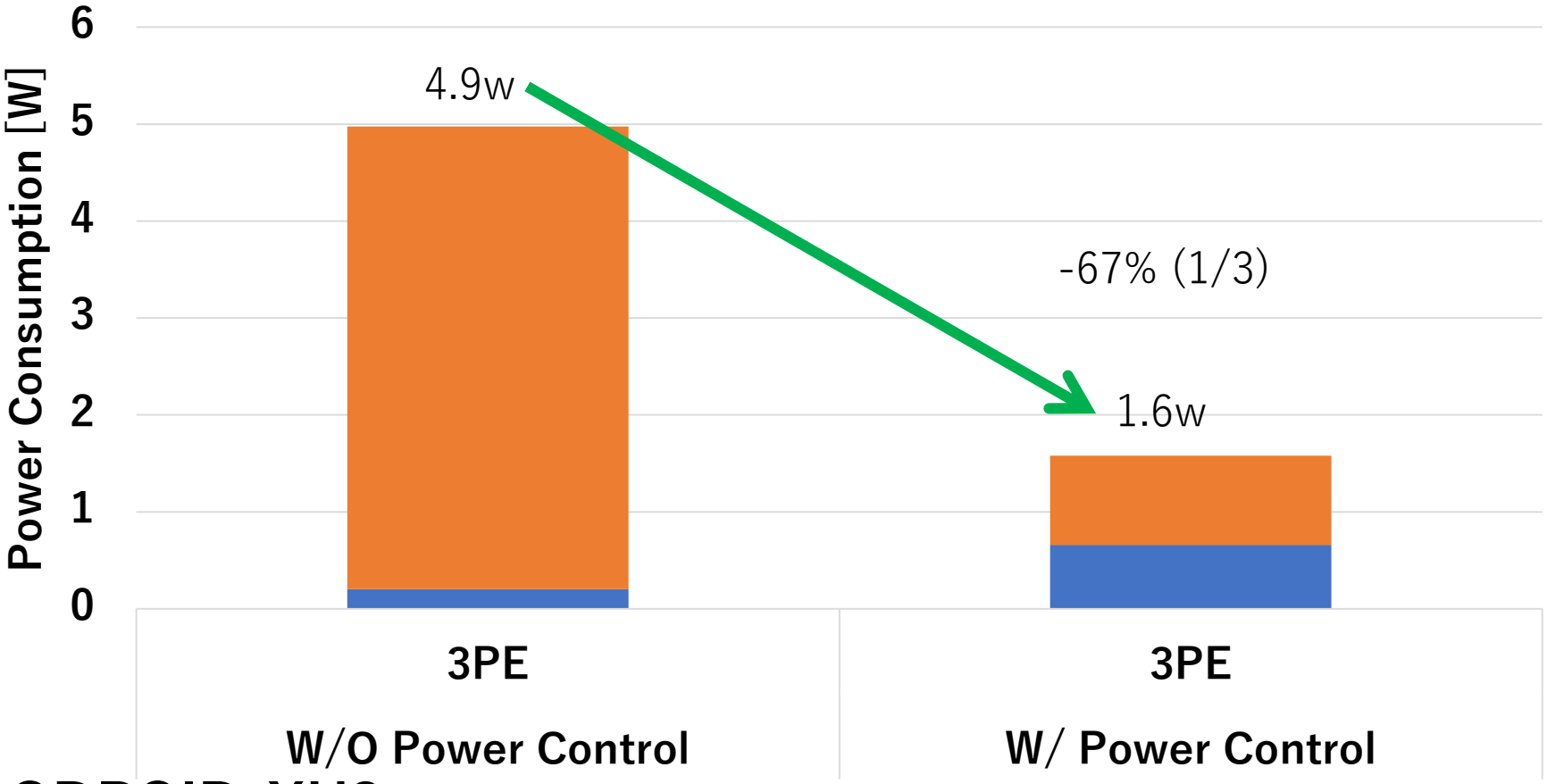


H81M-A, Intel Core i7 4770k
Quad core, 3.5GHz～0.8GHz



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 of OpenCV Face Detection on big.LITTLE ARM Processor

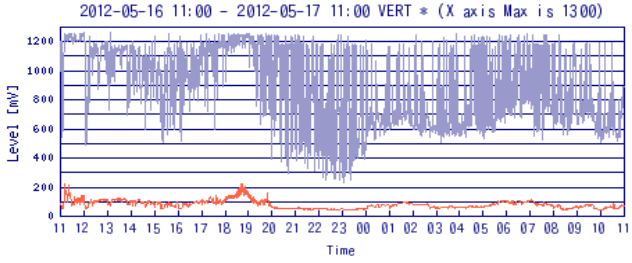


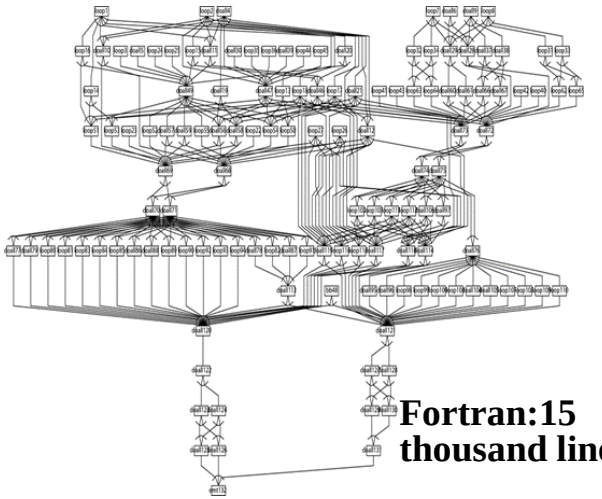
- ODROID-XU3
  - Samsung Exynos 5422 Processor
    - 4x Cortex-A15 2.0GHz, 4x Cortex-A7 1.4GHz big.LITTLE Architecture
    - 2GB LPDDR3 RAM

Frequency can be changed by each cluster unit

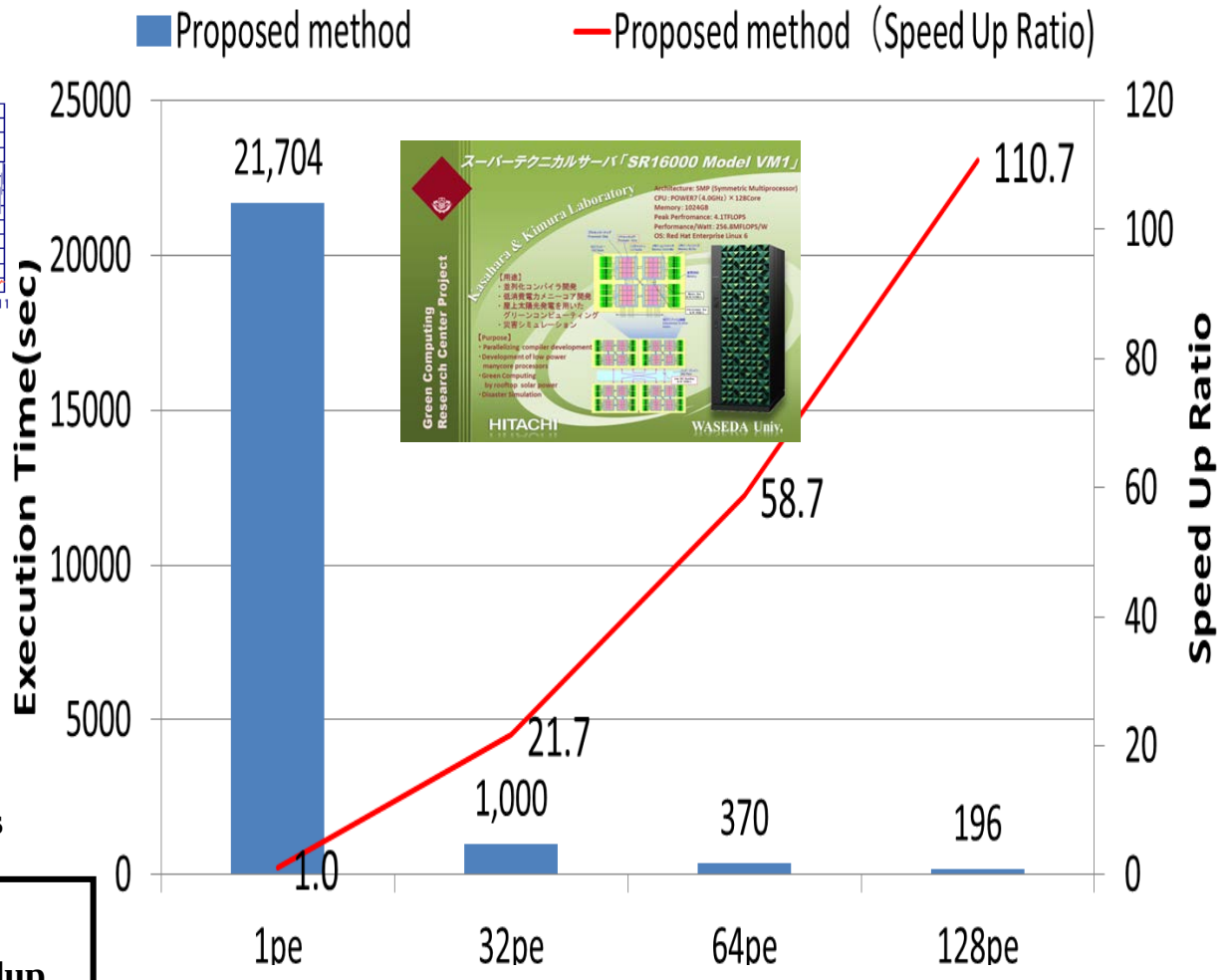
110 Times Speedup against the Sequential Processing for GMS Earthquake Wave Propagation Simulation on Hitachi SR16000

（Power7 Based 128 Core Linux SMP） (LCPC2015)





Fortran:15 thousand lines

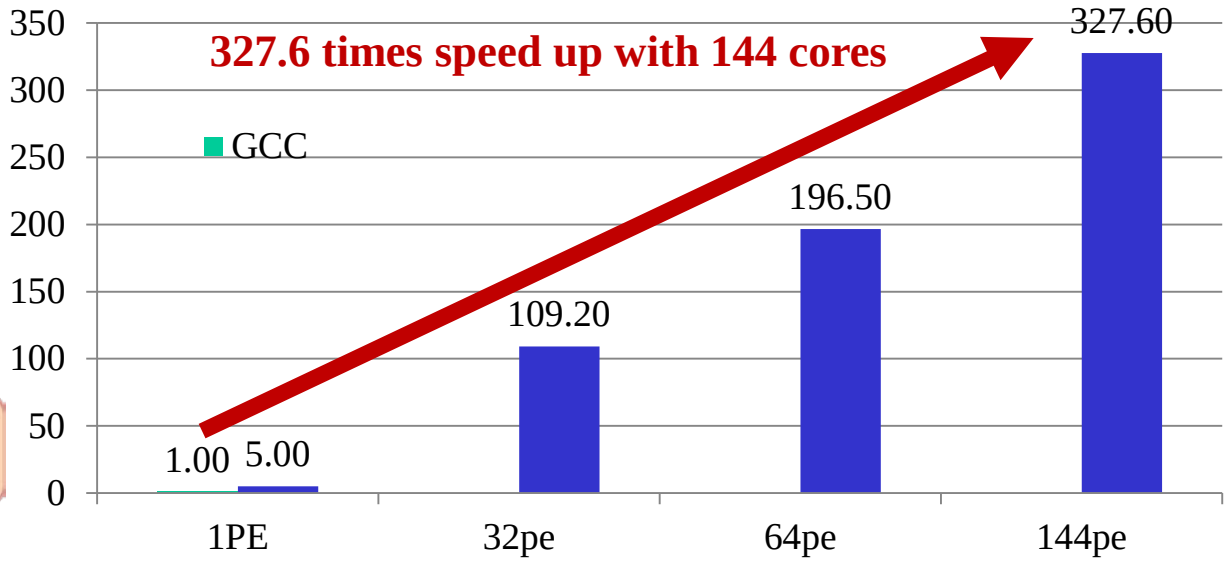


First touch for distributed shared memory and cache optimization over loops are important for scalable speedup

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

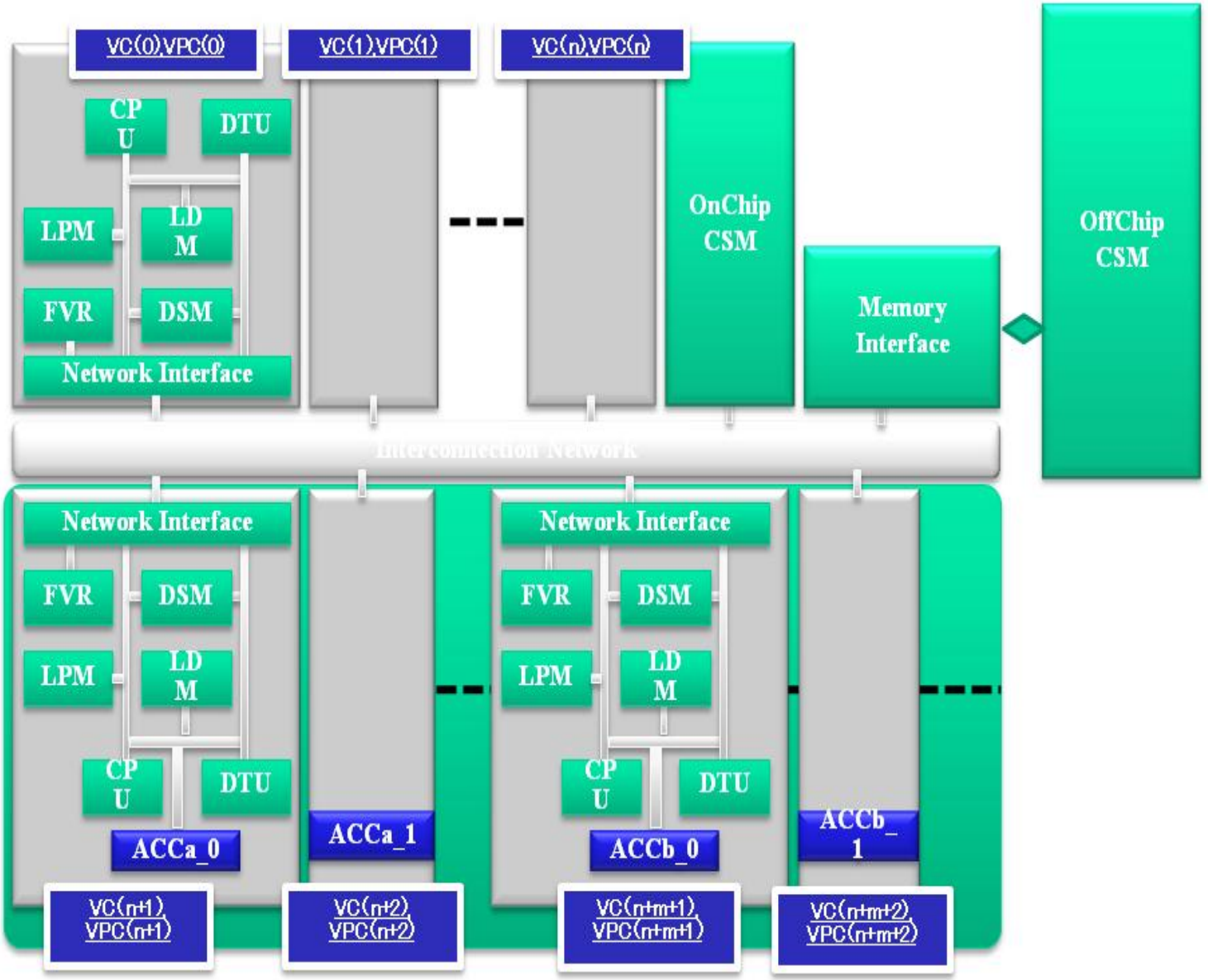
327 times speedup on 144 cores

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**

OSCAR Heterogeneous Multicore



DTU
- Data Transfer Unit

LPM
- Local Program Memory

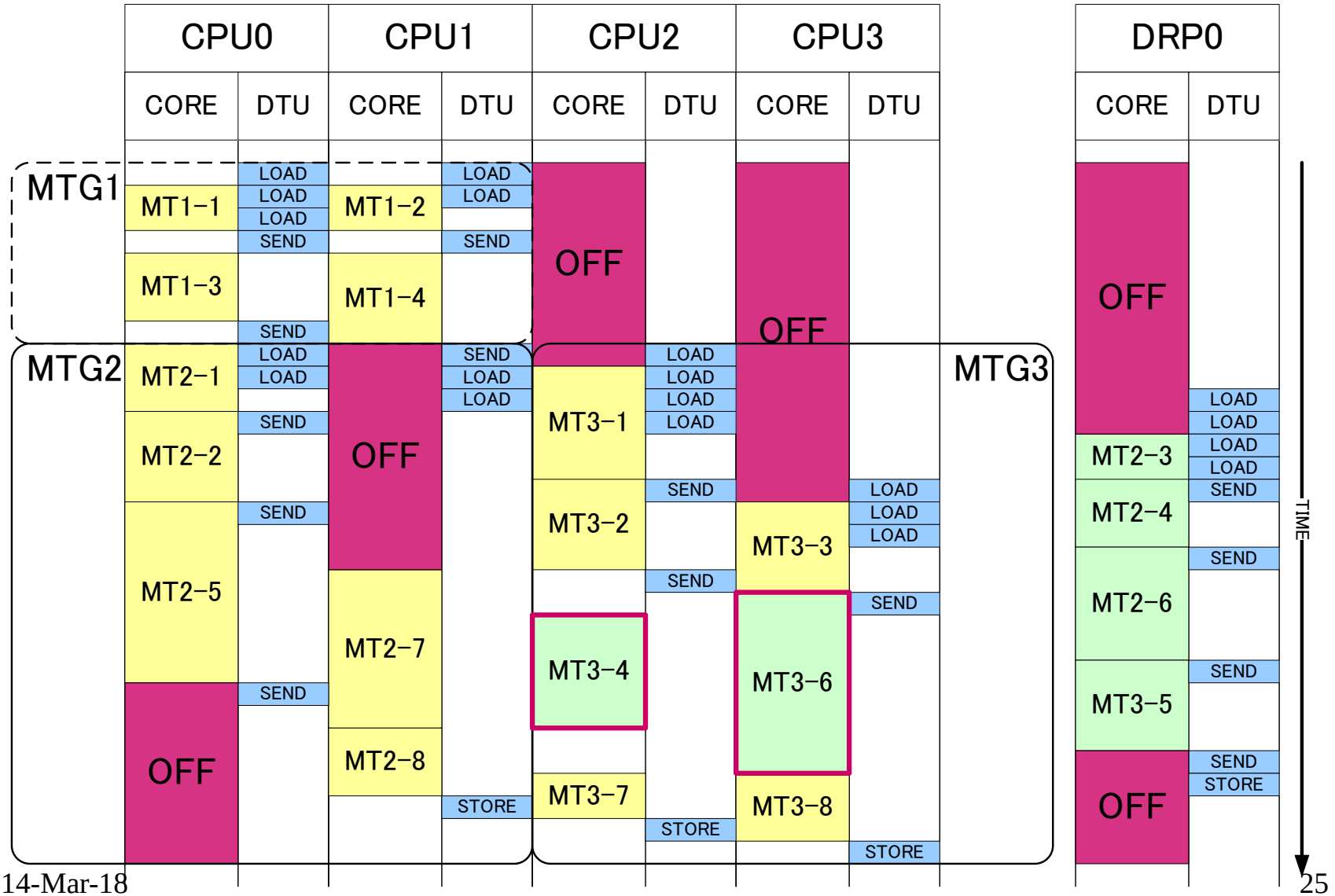
LDM
- Local Data Memory

DSM
- Distributed Shared Memory

CSM
- Centralized Shared Memory

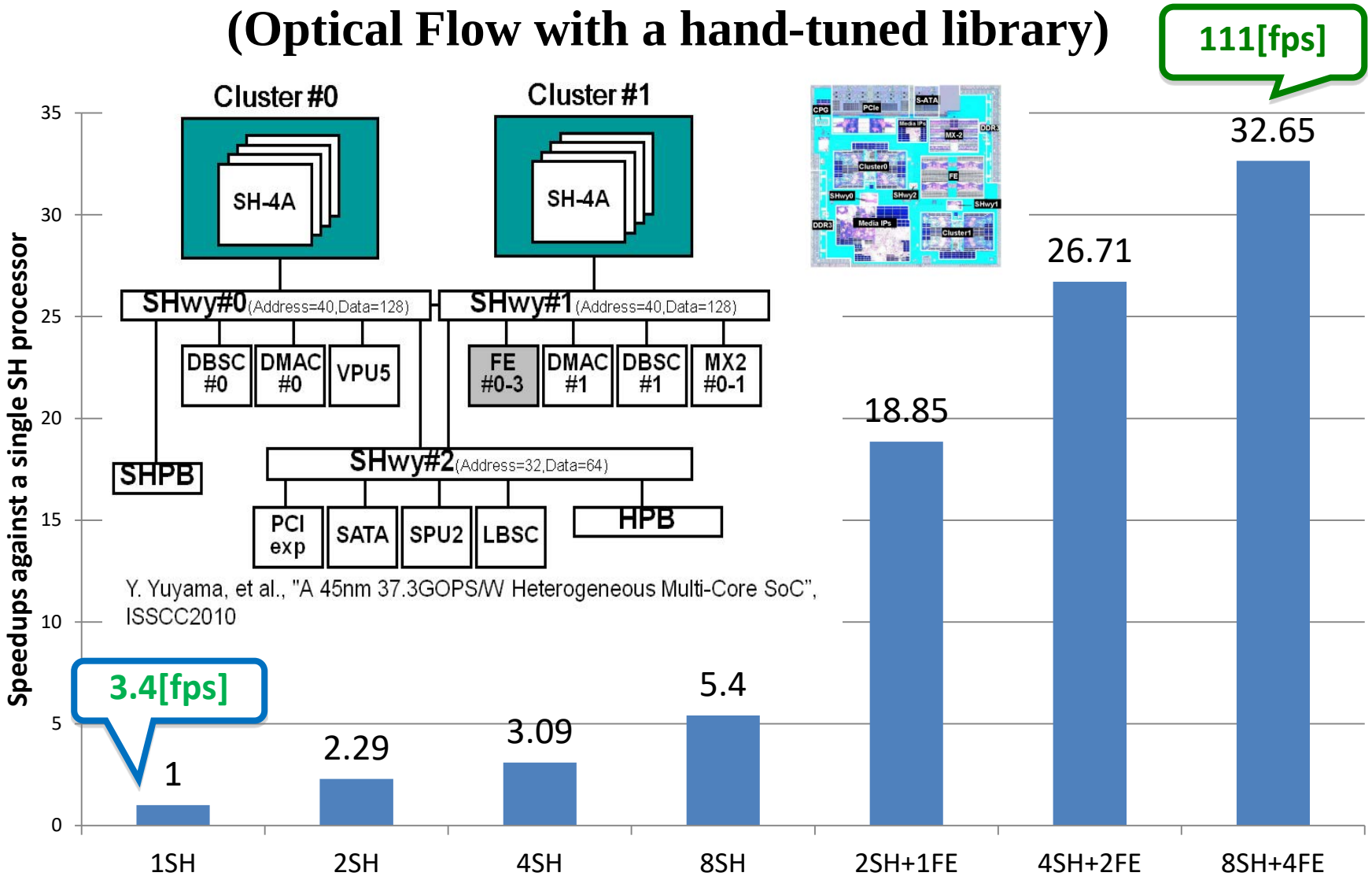
FVR
- Frequency/Voltage Control Register

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



33 Times Speedup Using OSCAR Compiler and OSCAR API on RP-X

(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)

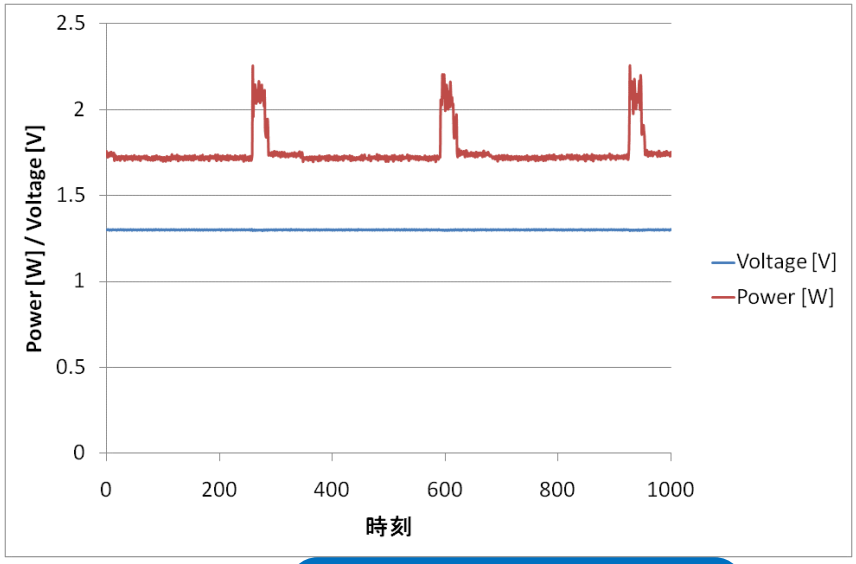
Without Power Reduction

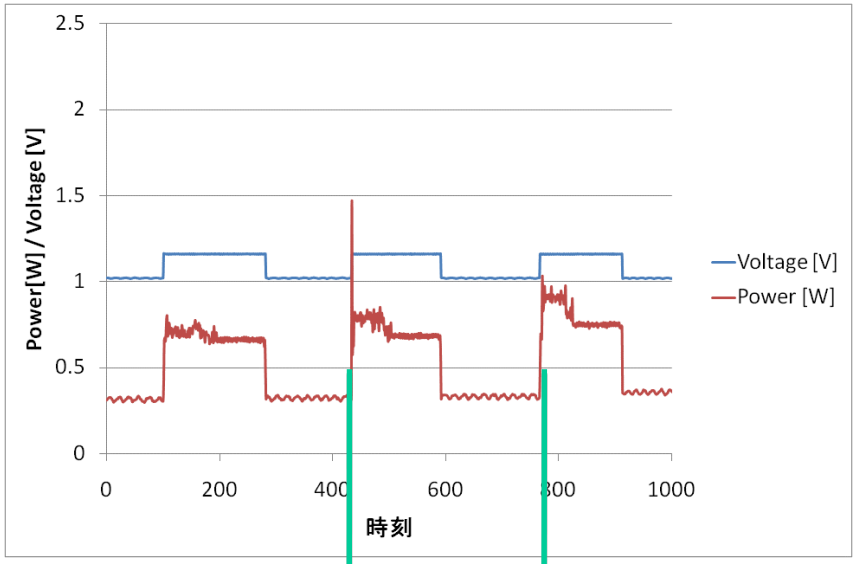
With Power Reduction by OSCAR Compiler

70% of power reduction

Average:1.76[W]

Average:0.54[W]

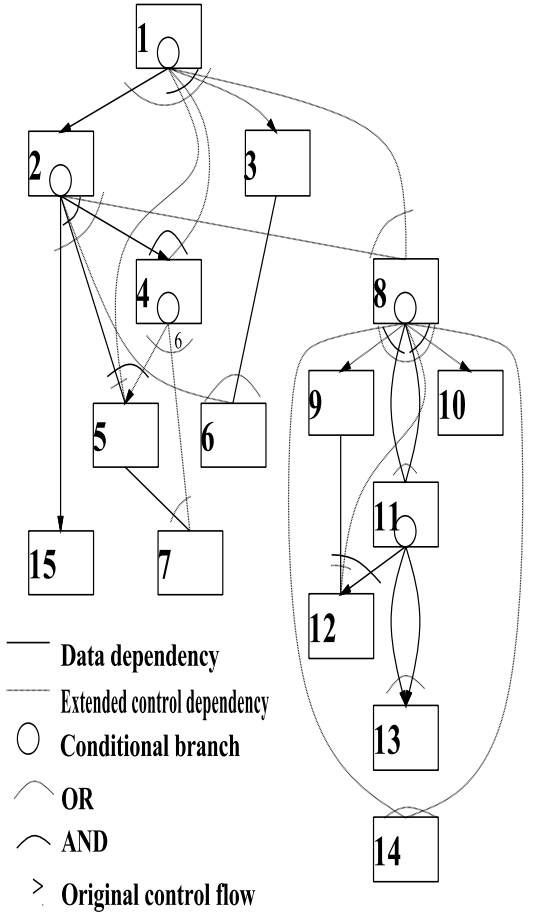




1cycle : 33[ms]
→30[fps]

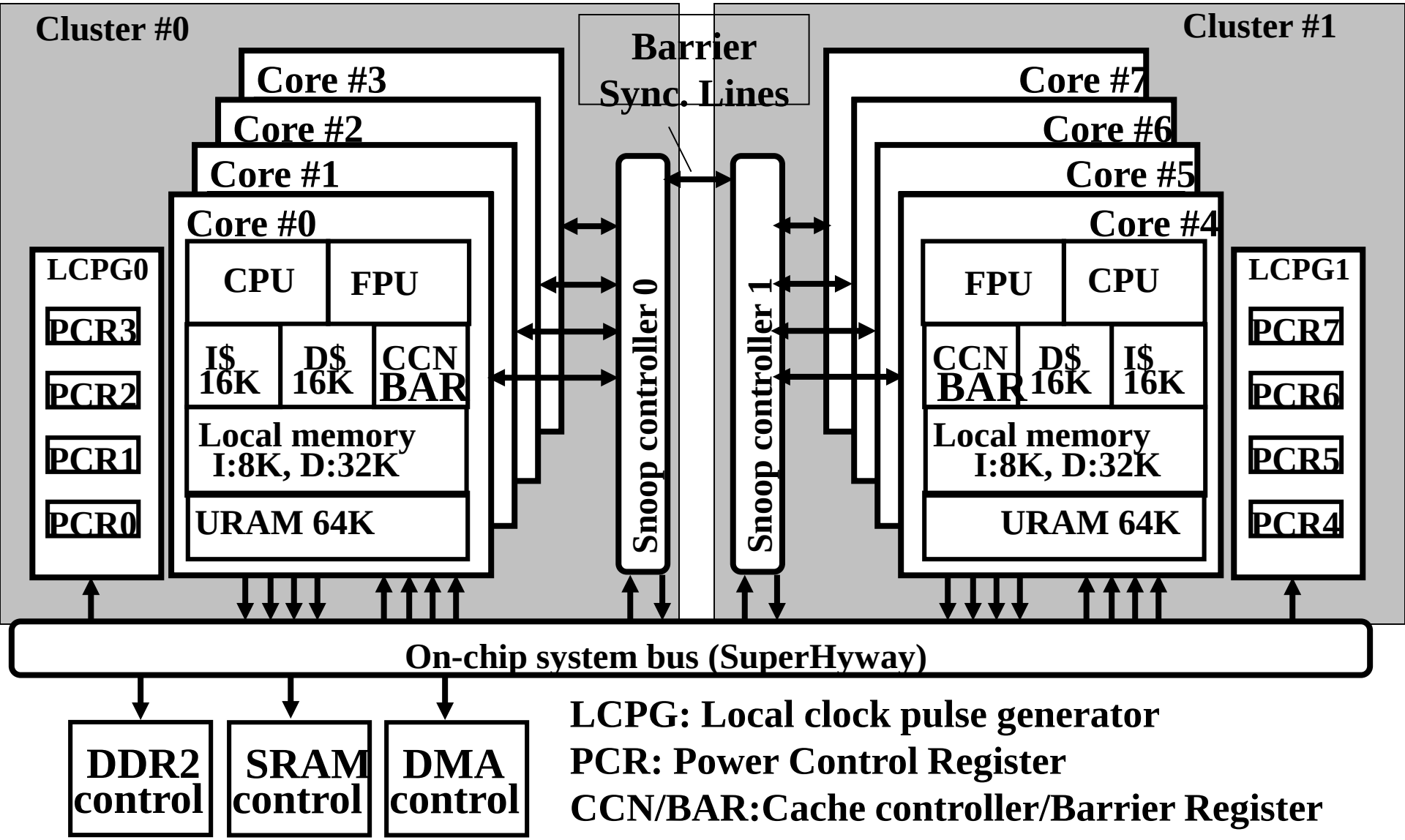
Software Coherence Control Method on OSCAR Parallelizing Compiler

- Coarse grain task parallelization with earliest condition analysis (control and data dependency analysis to detect parallelism among coarse grain tasks).
- OSCAR compiler automatically controls coherence using following simple program restructuring methods:
  - To cope with stale data problems:
    - **Data synchronization by compilers**
  - To cope with false sharing problem:
    - **Data Alignment**
    - **Array Padding**
    - **Non-cacheable Buffer**



MTG generated by earliest executable condition analysis

8 Core RP2 Chip Block Diagram



LCPG: Local clock pulse generator

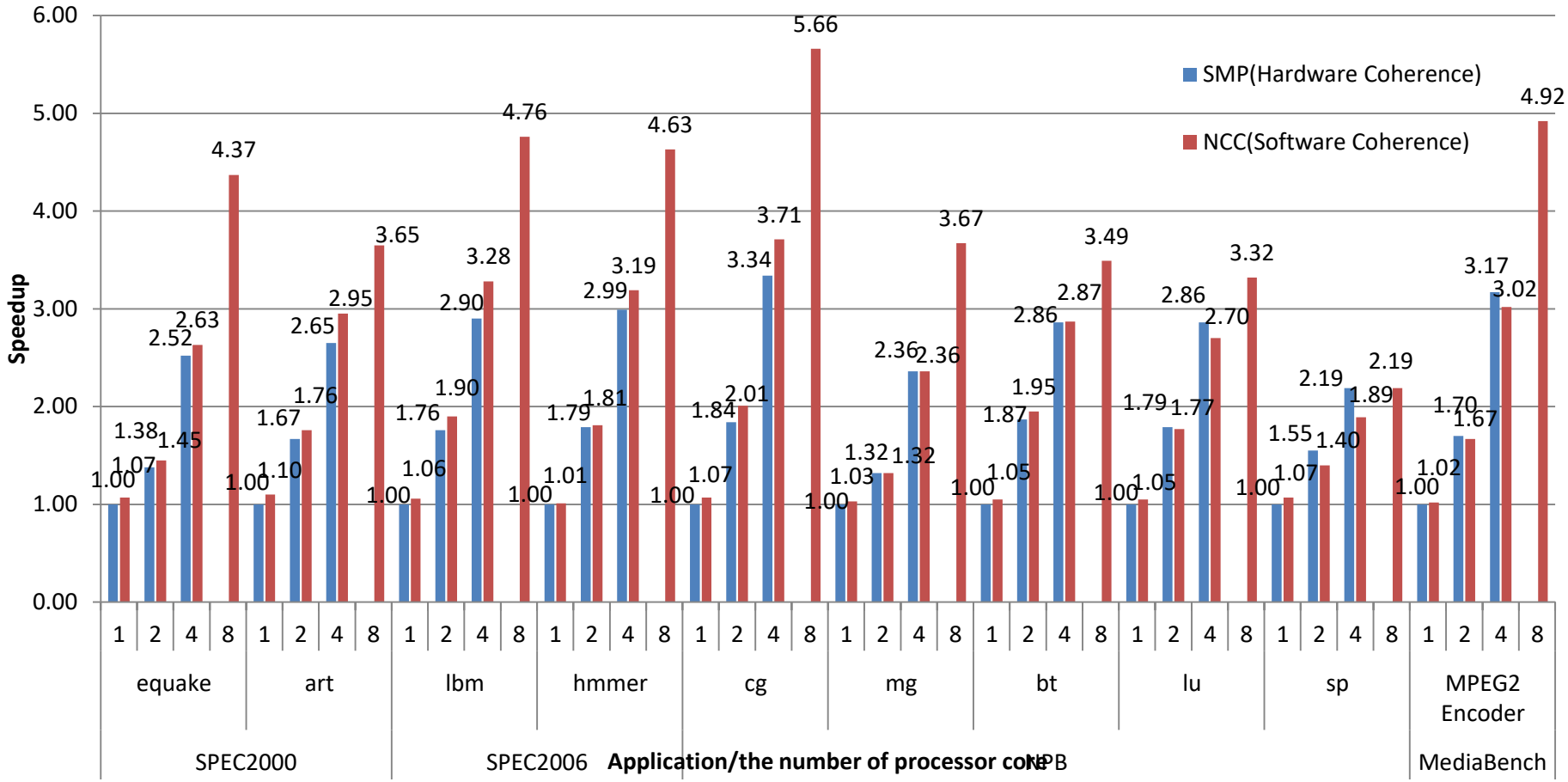
PCR: Power Control Register

CCN/BAR:Cache controller/Barrier Register

URAM: User RAM (Distributed Shared Memory)

Automatic Software Coherent Control for Manycores

Performance of Software Coherence Control by OSCAR Compiler on 8-core RP2

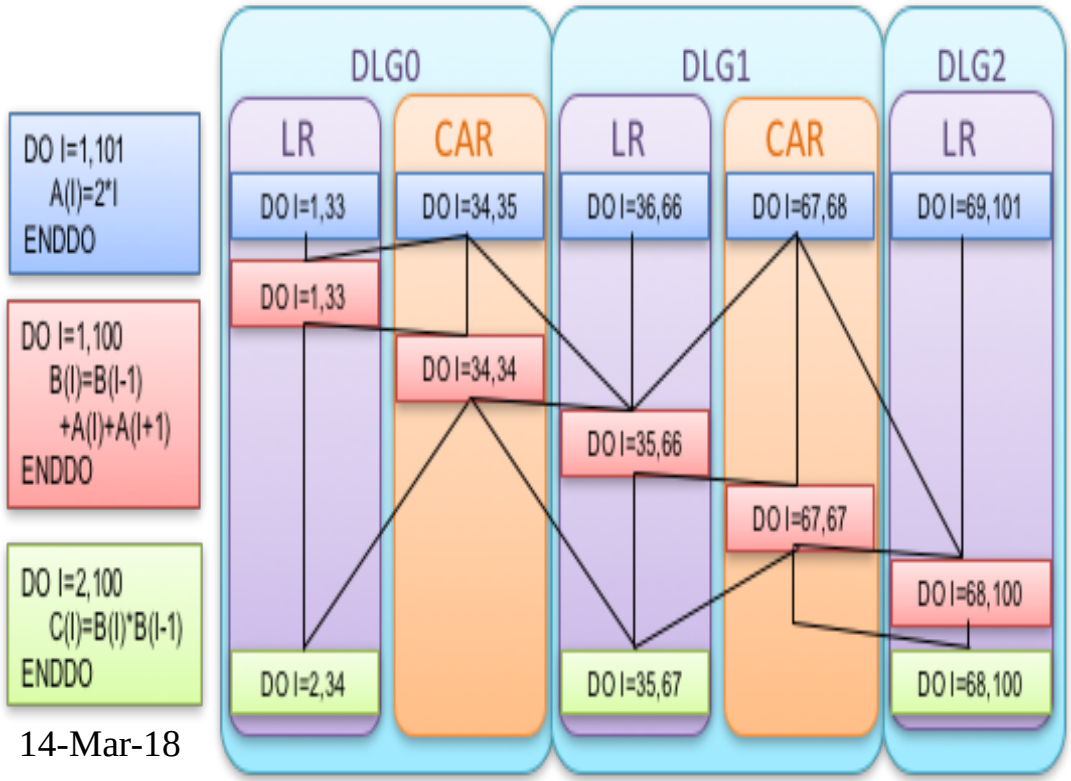


Automatic Local Memory Management

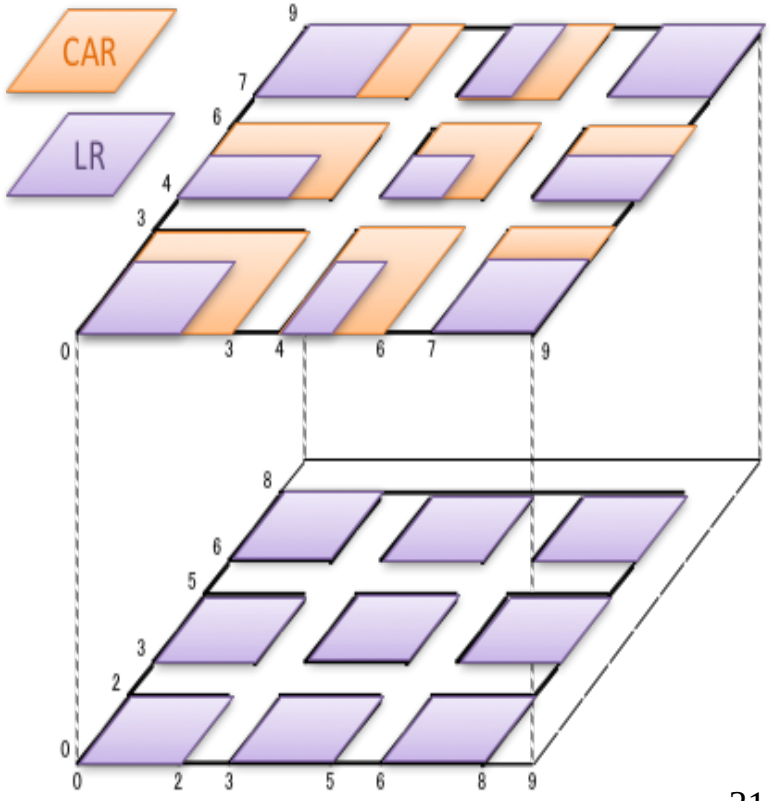
Data Localization: Loop Aligned Decomposition

- **Decomposed loop into LRs and CARs**
  - **LR (Localizable Region): Data can be passed through LDM**
  - **CAR (Commonly Accessed Region): Data transfers are required among processors**

Single dimension Decomposition

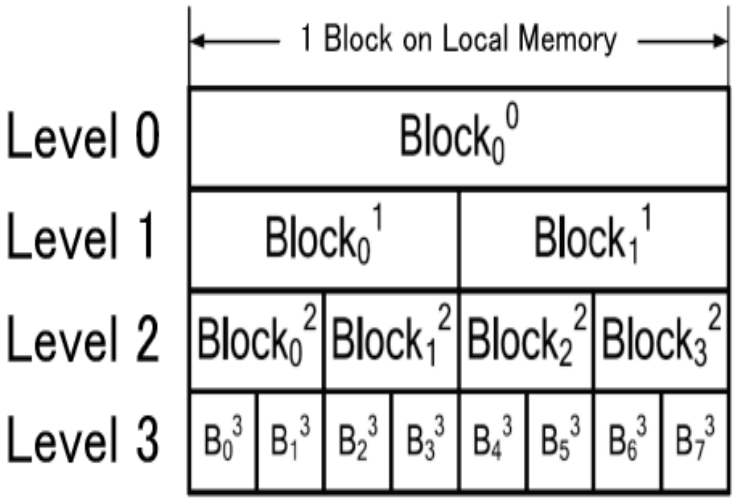


Multi-dimension Decomposition



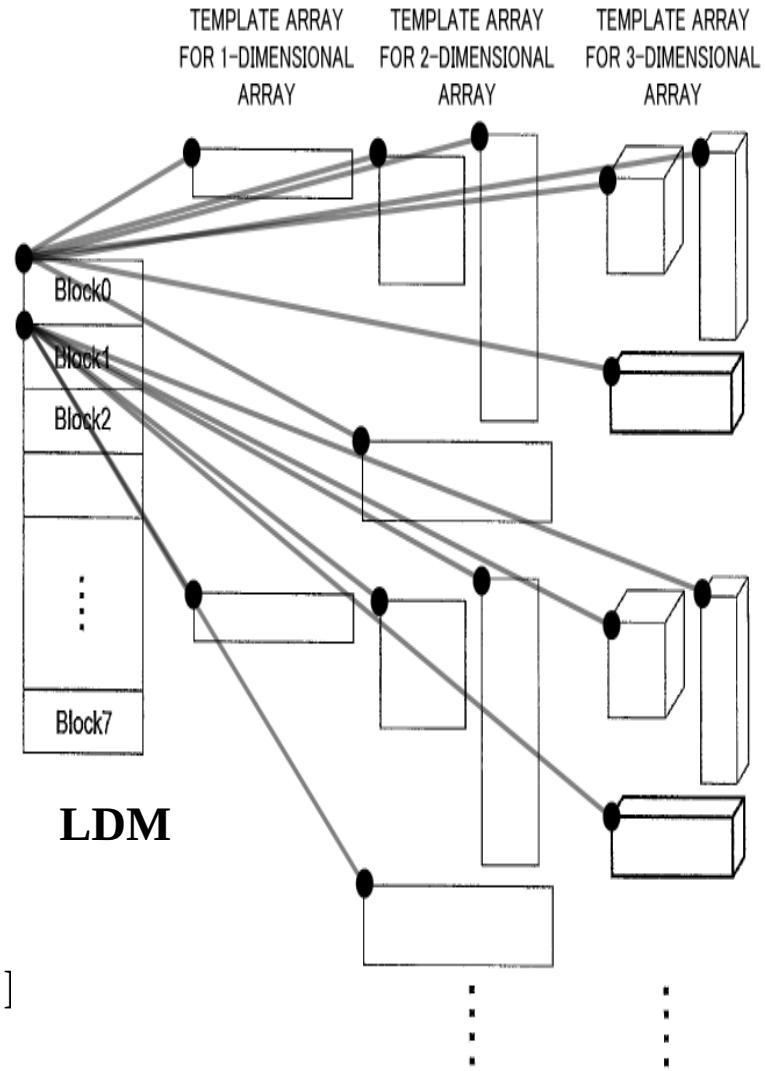
Adjustable Blocks

- Handling a suitable block size for each application
  - different from a fixed block size in cache
  - each block can be divided into smaller blocks with intege $Block_{Number}^{Level}$ small arrays and scalar



Multi-dimensional Template Arrays for Improving Readability

- a mapping technique for arrays with varying dimensions
  - each block on LDM corresponds to multiple empty arrays with varying dimensions
  - these arrays have an additional dimension to store the corresponding block number
    - TA[Block#][] for single dimension
    - TA[Block#][][] for double dimension
    - TA[Block#][][][] for triple dimension
    - ...
- LDM are represented as a one dimensional array
  - without Template Arrays, multi-dimensional arrays have complex index calculations
    - A[i][j][k] -> TA[offset + i' * L + j' * M + k']
  - Template Arrays provide readability
    - A[i][j][k] -> TA[Block#][i'][j'][k']

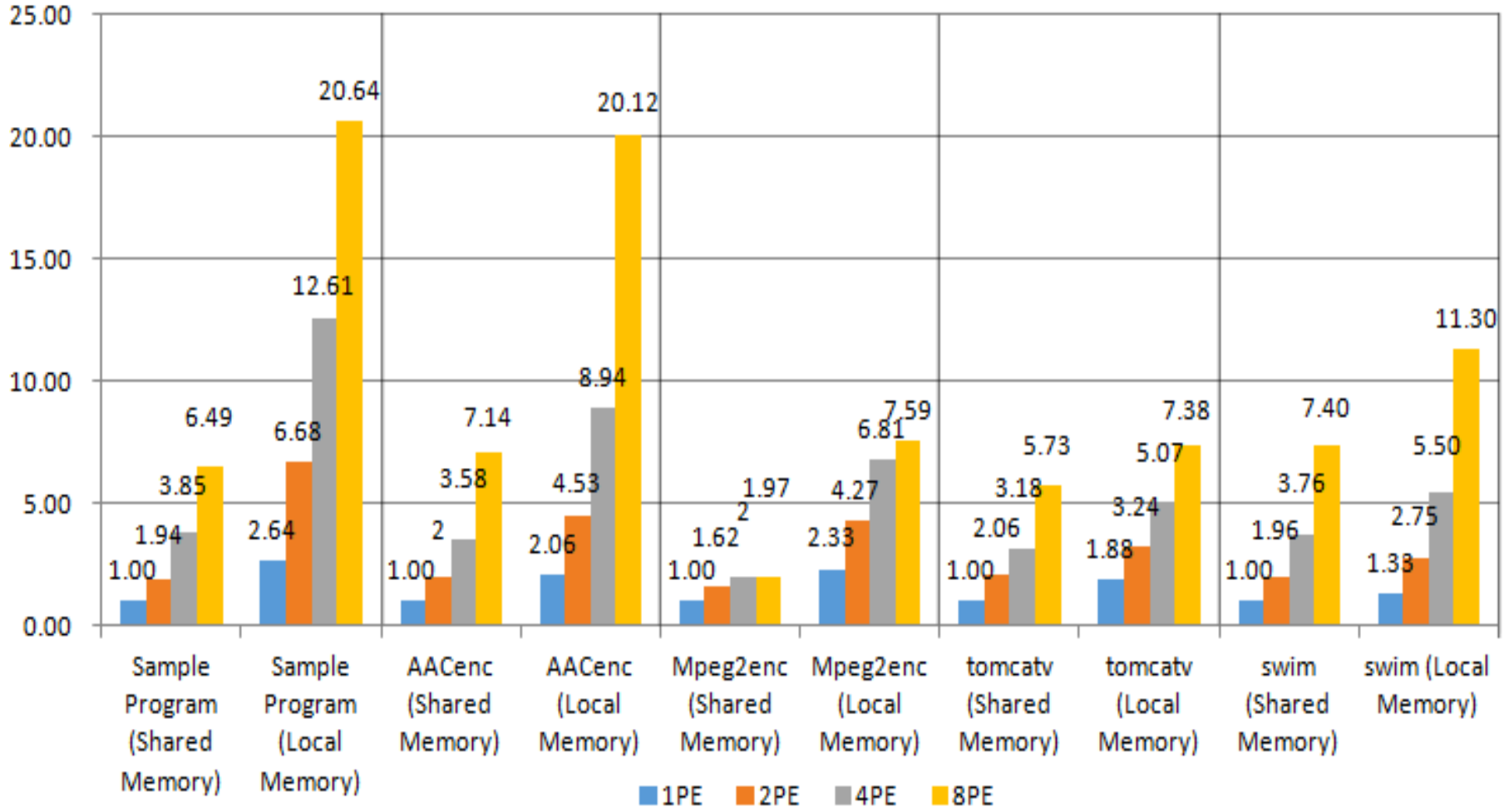


LDM

Block Replacement Policy

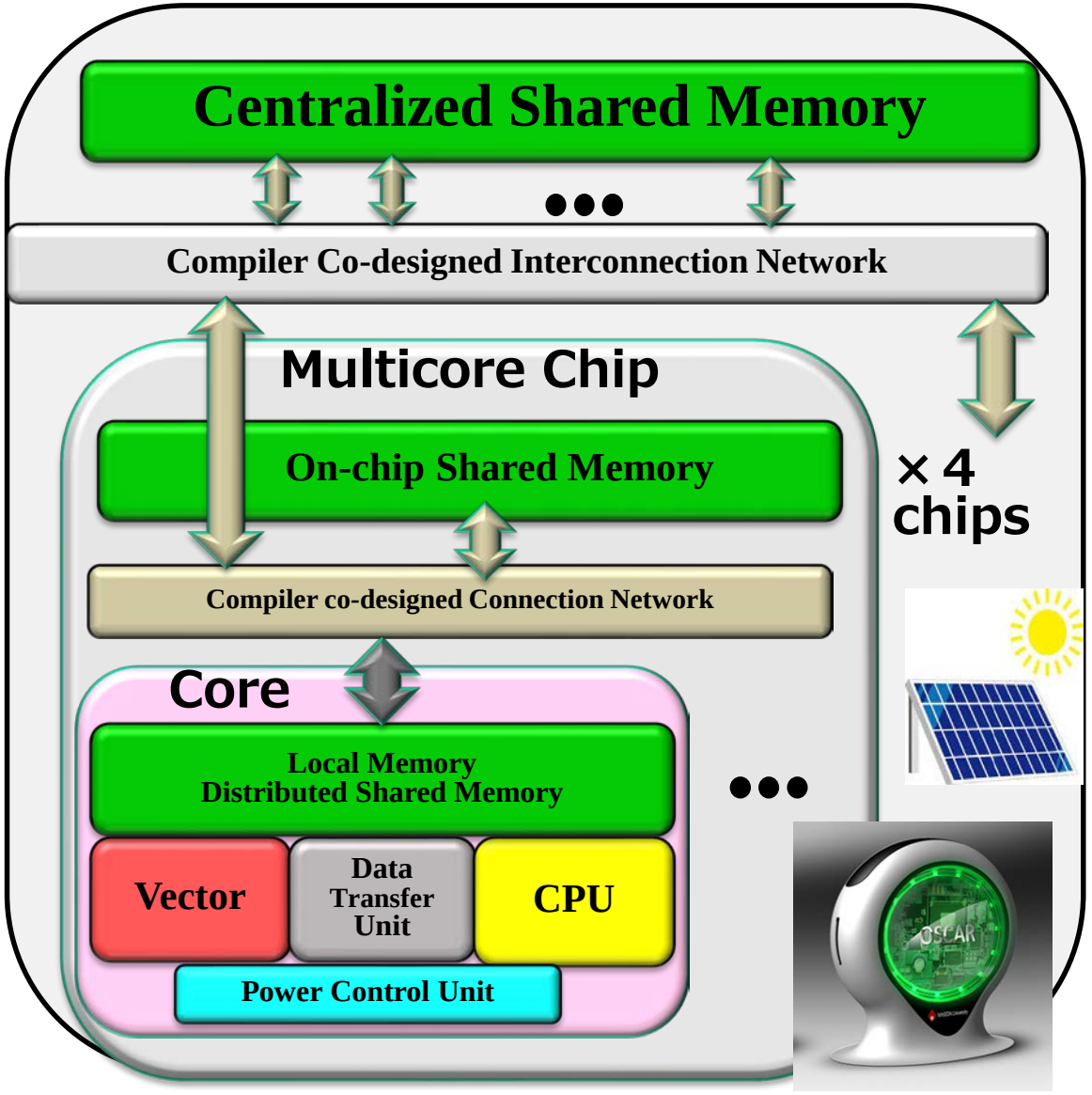
- Compiler Control Memory block Replacement
  - using live, dead and reuse information of each variable from the scheduled result
  - different from LRU in cache that does not use data dependence information
- Block Eviction Priority Policy
  1. (Dead) Variables that will not be accessed later in the program
  2. Variables that are accessed only by other processor cores
  3. Variables that will be later accessed by the current processor core
  4. Variables that will immediately be accessed by the current processor core

Speedups by the Proposed Local Memory Management Compared with Utilizing Shared Memory on Benchmarks Application using RP2



20.12 times speedup for 8cores execution using local memory against sequential execution using off-chip shared memory of RP2 for the AACenc

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



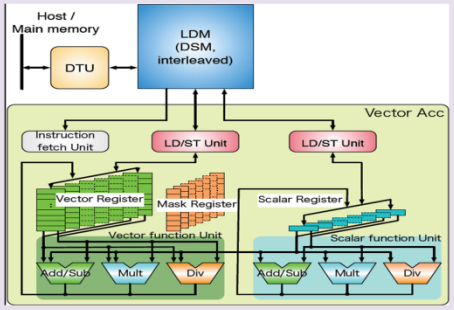
Target:

- **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

Future Multicore Products

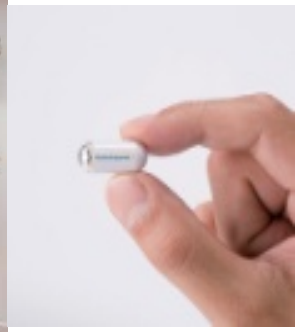
Next Generation Automobiles

- 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

IEEE Computer Society

More than 60,000 computer scientist and IT professionals, in 168 countries driving technological innovation.

Our Vision:

To be the leading provider of technical information, community services, and personalized services to the world's computing professionals.

IEEE Computer Society Student Membership

IEEE + Computer Society

Japan = US $35（Half year: US$17）

- Already an IEEE member = US$8 (Half year: US$4)

Full member benefits

- Included Access to Computer Society Digital Library

13 Magazines | 20 Journals | 600,000 Searchable Articles | 569,000 Authors | 9,000 Conference Publications

- Discounts on all IEEE Computer Society Conferences & Events
- Scholarships
- Opportunities to publish and be recognized
- Skill development
- Refine your interests
- Make connections and assemble your network
- Growing IEEE young professional network