Research of OSCAR Parallelizing Compiler for High Performance and Low Power Green Computing

Hironori Kasahara

Professor, Dept. of Computer Science & Engineering

Director, Advanced Multicore Processor Research Institute

Waseda University, Tokyo, Japan

IEEE Computer Society Board of Governors

IEEE Computer Society Multicore STC Chair

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

Green Computing Systems R&D Center

Waseda University

Supported by METI (Mar. 2011 Completion)

＜R & D Target＞

Hardware, Software, Application for Super Low-Power Manycore Processors

- **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, etc

＜Ripple Effect＞

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

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

Beside Subway Waseda Station, Near Waseda Univ. Main Campus

2

Industry-government-academia collaboration in R&D and target practical applications

Protect lives

For smart life

Toyota

Denso

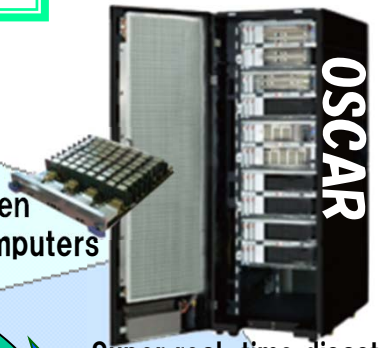
Robots

On-board vehicle technology (navigation systems, integrated controllers, infrastructure coordination)

Protect environment

Waseda University :R&D
Many-core system technologies with ultra-low power consumption
OSCAR many-core chip

NEC | Renesas | ESOL | OSCAR Many-core Chip | OSCAR Compiler | OSCAR Technology | OSCAR API | Hitachi | Fujitsu | 11 Companies 3 Univ.



Green supercomputers

Super real-time disaster simulation (tectonic shifts, tsunami), tornado, flood, fire spreading)

Green cloud servers

Stock trading

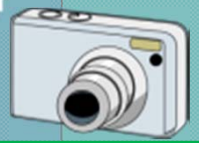
Tokyo Stock Exchange

Consumer electronic Internet TV/DVD

Camcorders

Capsule inner cameras

Cameras



Olympus

Solar Powered Smart phones

Operation/recharging by solar cells

Fujitsu

Medical servers

National Institute of Radiological Sciences

Heavy particle radiation planning, cerebral infarction)

Mitsubishi Electric

Cool desktop servers

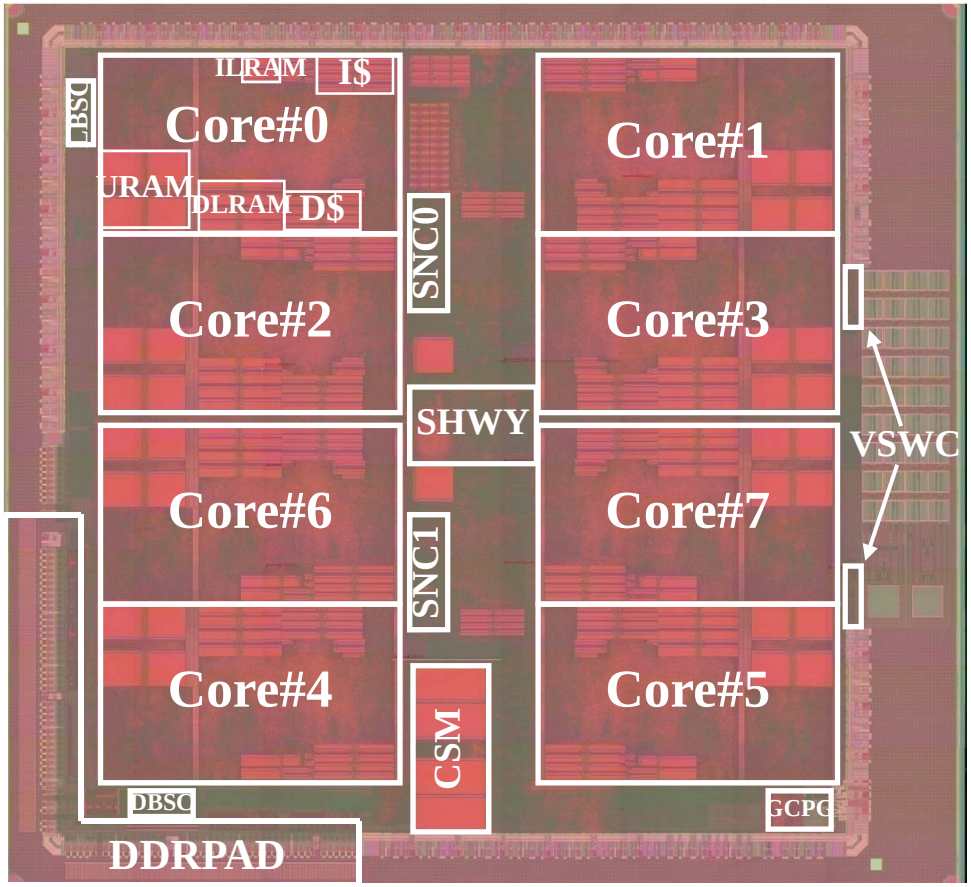
Non-fan, cool, quiet servers designed for server

Industry

Intelligent home appliances

Supercomputers and servers

Renesas-Hitachi-Waseda Low Power 8 core RP2 Developed in 2007 in METI/NEDO project



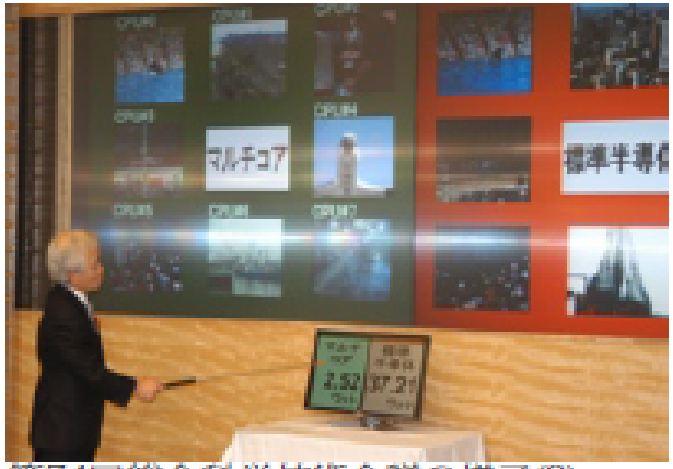
Process Technology	90nm, 8-layer, triple-Vth, CMOS
Chip Size	104.8mm^2 (10.61mm x 9.88mm)
CPU Core Size	6.6mm^2 (3.36mm x 1.96mm)
Supply Voltage	1.0V–1.4V (internal), 1.8/3.3V (I/O)
Power Domains	17 (8 CPUs, 8 URAMs, common)

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"

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

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)

CSTP Members

Prime Minister:
Mr. Y. FUKUDA

Minister of State for Science, Technology and Innovation Policy:
Mr. F. KISHIDA

Chief Cabinet Secretary:
Mr. N. MACHIMURA

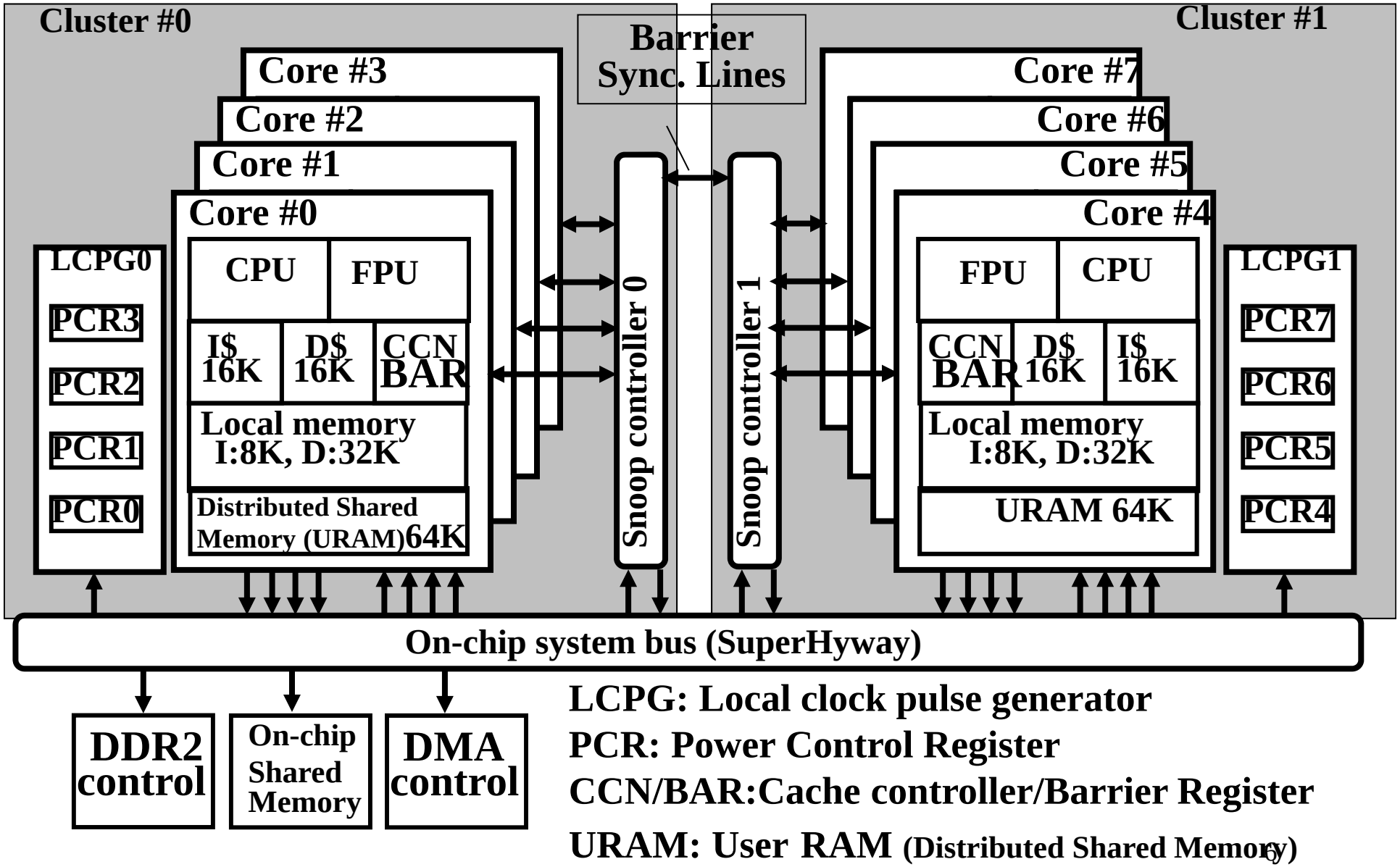
Minister of Internal Affairs and Communications :
Mr. H. MASUDA

Minister of Finance :
Mr. F. NUKAGA

Minister of Education, Culture, Sports, Science and Technology:
Mr. K. TOKAI

Minister of Economy,Trade and Industry:
Mr. A. AMARI

8 Core RP2 Chip Block Diagram



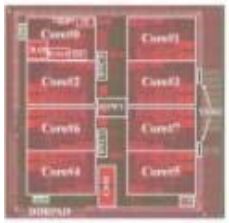
LCPG: Local clock pulse generator

PCR: Power Control Register

CCN/BAR:Cache controller/Barrier Register

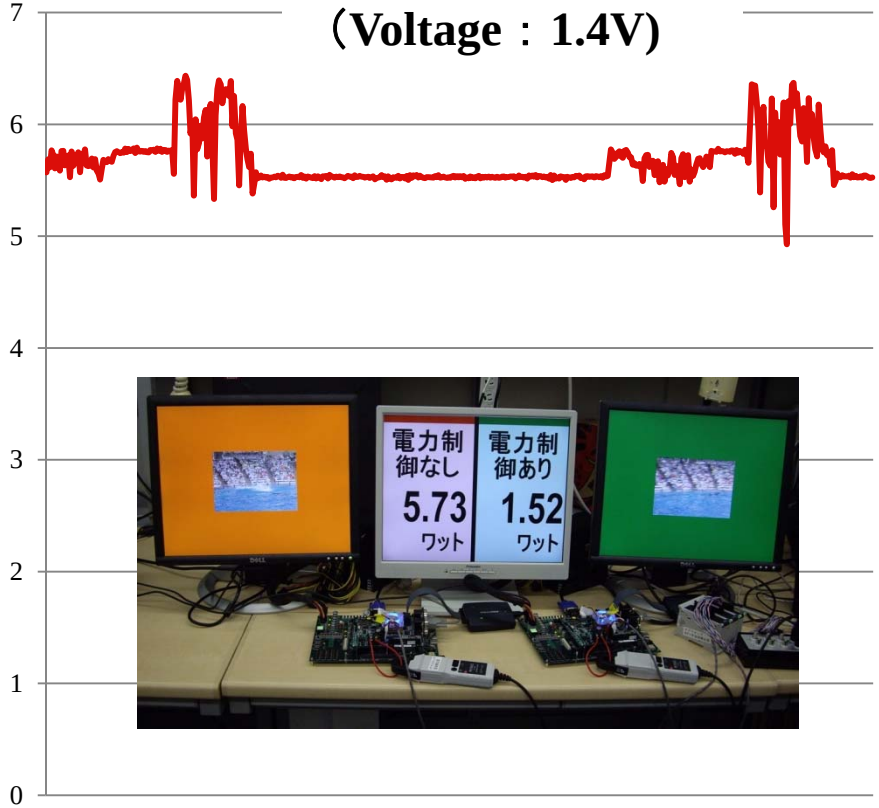
URAM: User RAM (Distributed Shared Memory)

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

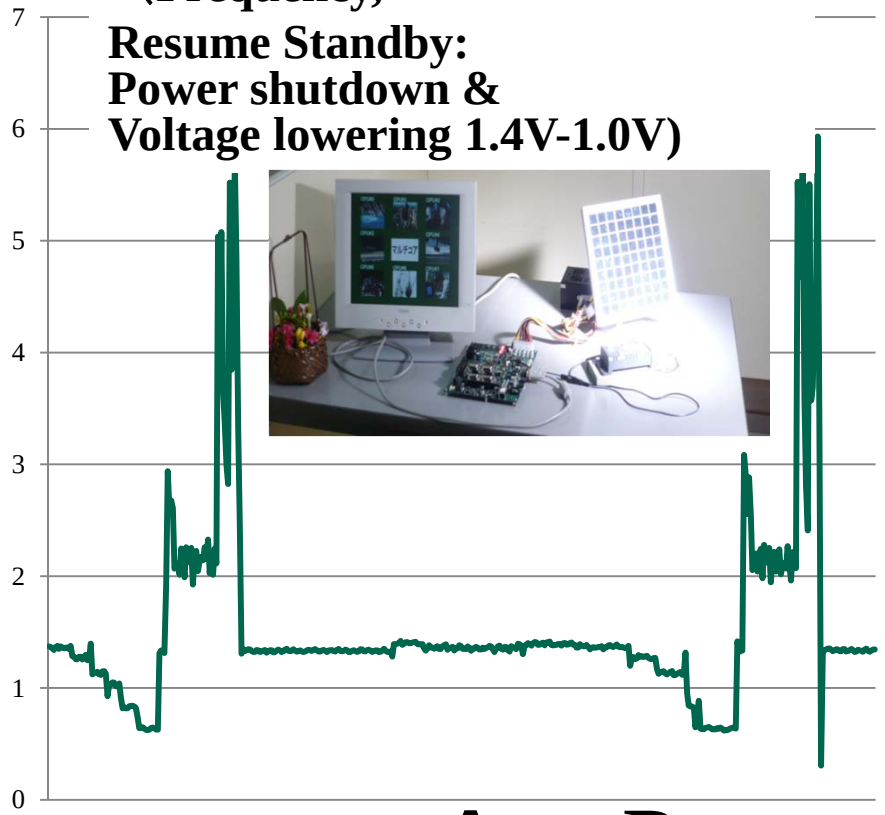


MPEG2 Decoding with 8 CPU cores

Without Power Control （Voltage：1.4V)



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

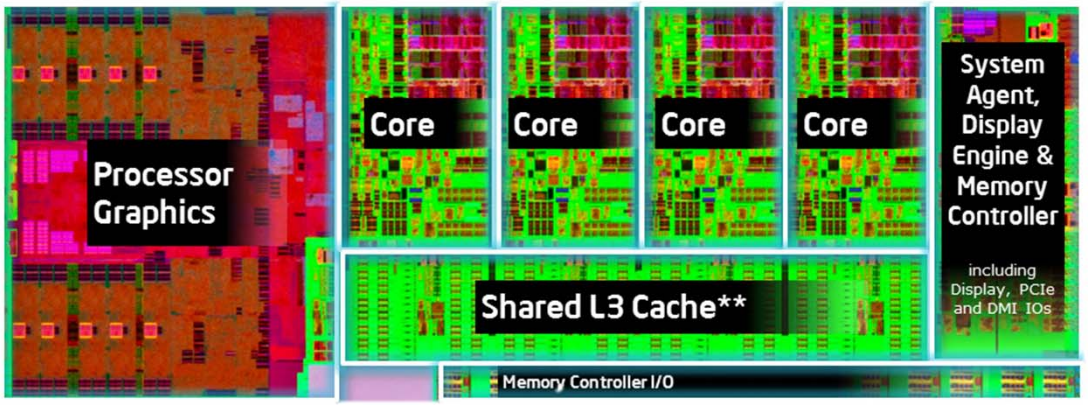


Avg. Power 5.73 [W]

73.5% Power Reduction

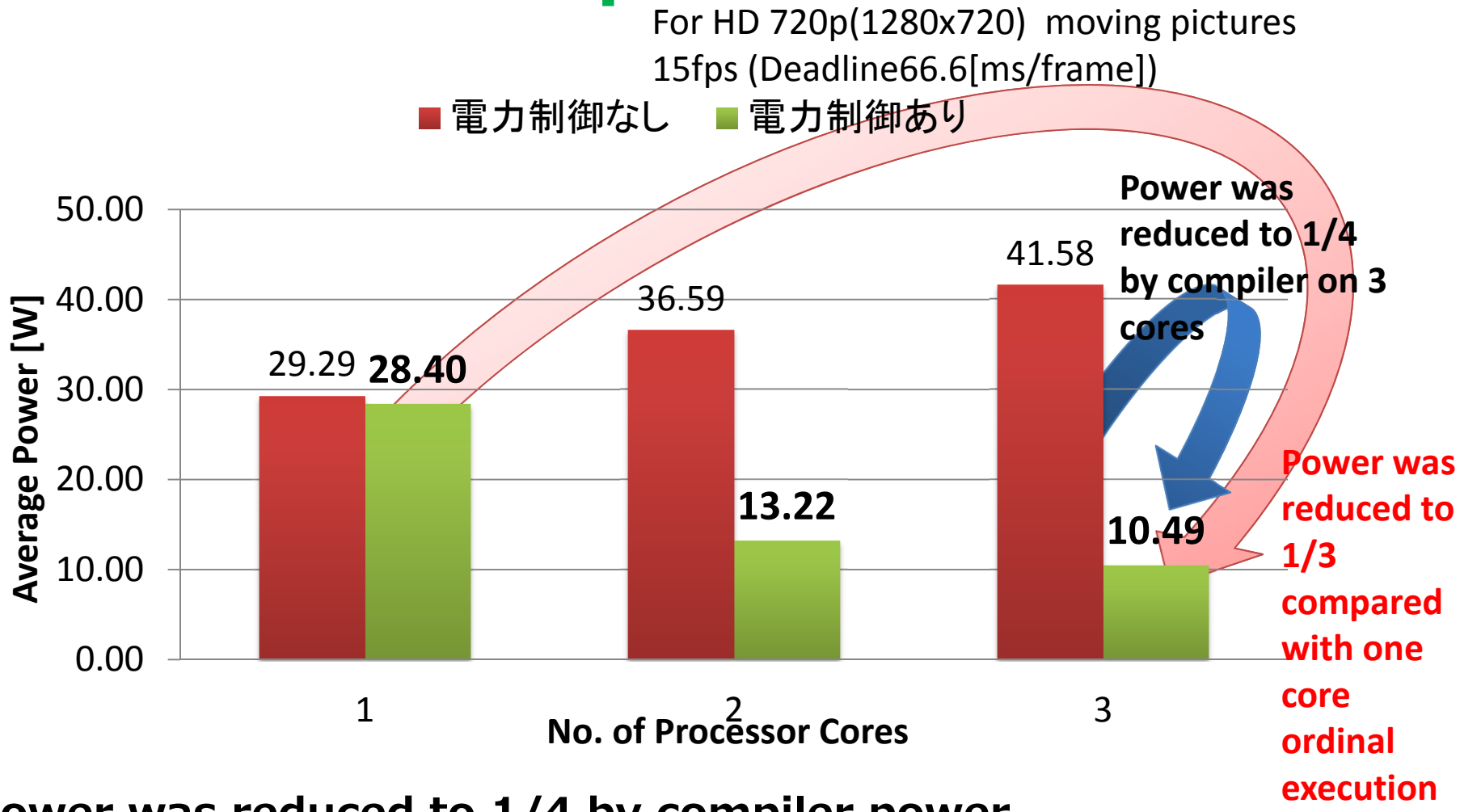
Avg. Power 1.52 [W]

Automatic Power Reduction on 4 core Intel Haswell



- Haswell Processor
  - OS Ubuntu 13.10
  - Intel CPU Core i7 4770K
    - 4 cores
    - L1 Cache: Load 64Bytes/cycle, Store 32Bytes/cycle
    - L2 Cache 64Bytes/cycle
    - L3 Cache 8 MB
    - Frequency 3.5GHz~0.8MHz
  - Memory 16GB (8GB×2)

Power Reduction on Intel Haswell for Real-time Optical Flow



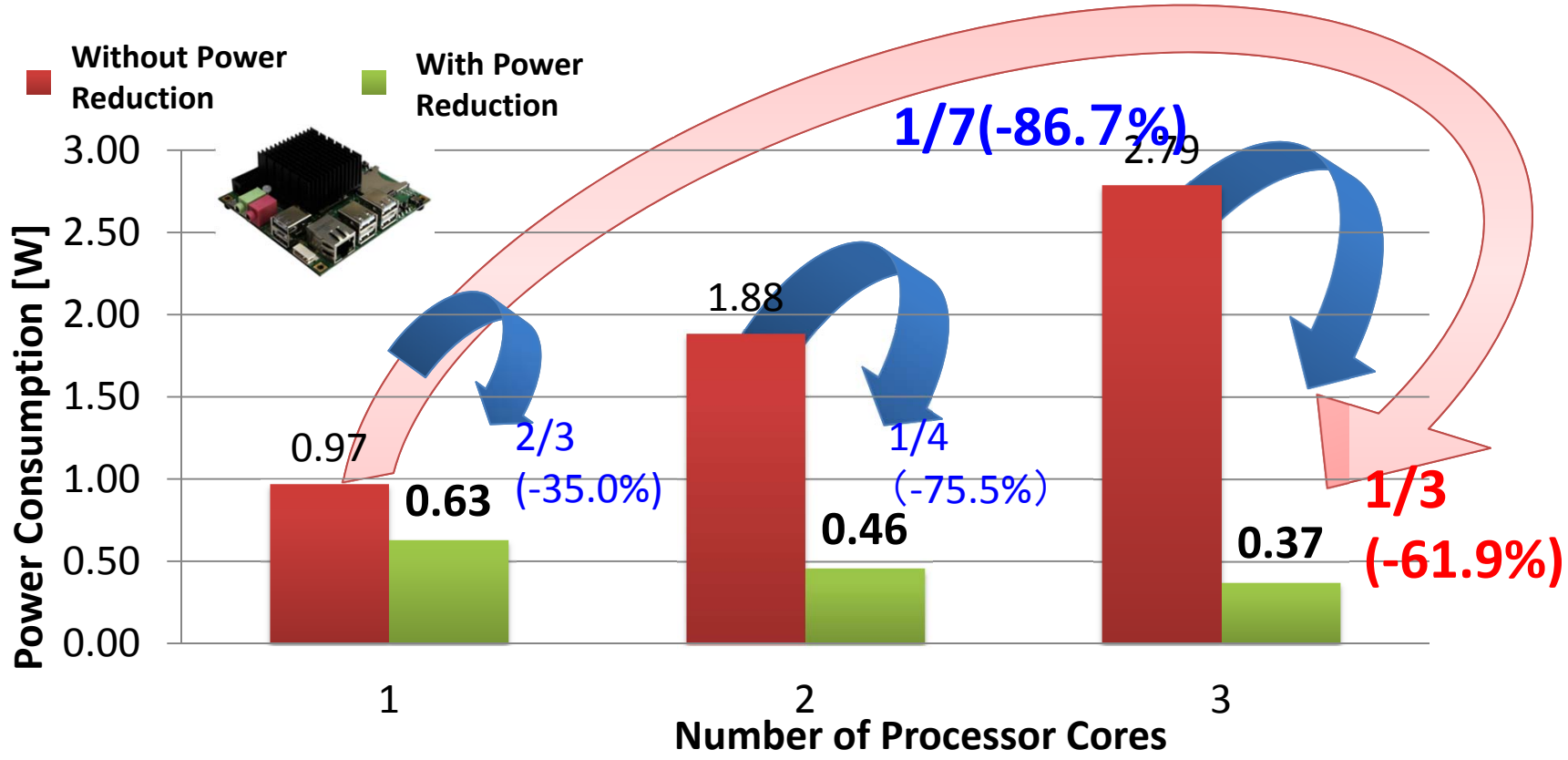
Power was reduced to 1/4 by compiler power optimization on the same 3 cores.

The power with 3 core was reduced to 1/3 against 1 core.

Automatic Power Reduction for MPEG2 Decode on Android Multicore

ODROID X2 ARM Cortex-A9 4 cores

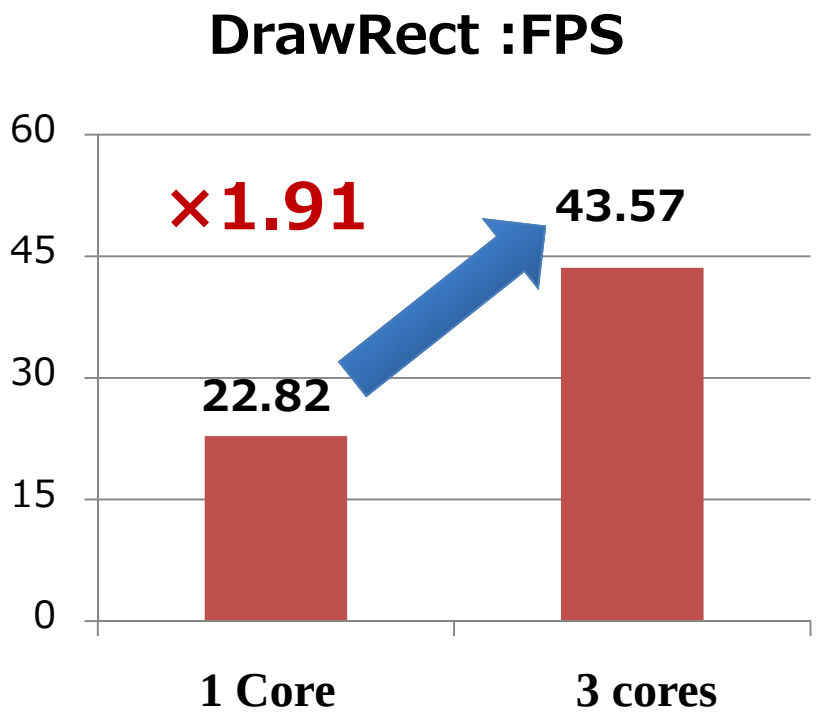
http://www.youtube.com/channel/UCS43lNYEIkC8i_KIgFZYQBQ

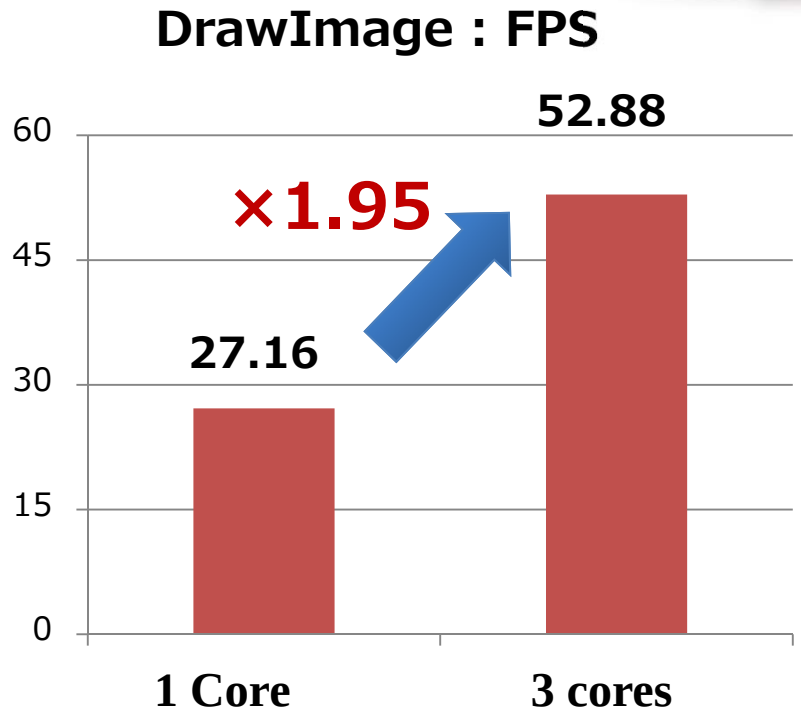


- On 3 cores, Automatic Power Reduction successfully reduced power to 1/7 against without Power Reduction.
- 3 cores with power reduction reduced power to 1/3 against ordinary 1 core execution.

Parallelization of 2D Rendering Engine SKIA on 3 cores of Google NEXUS7

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



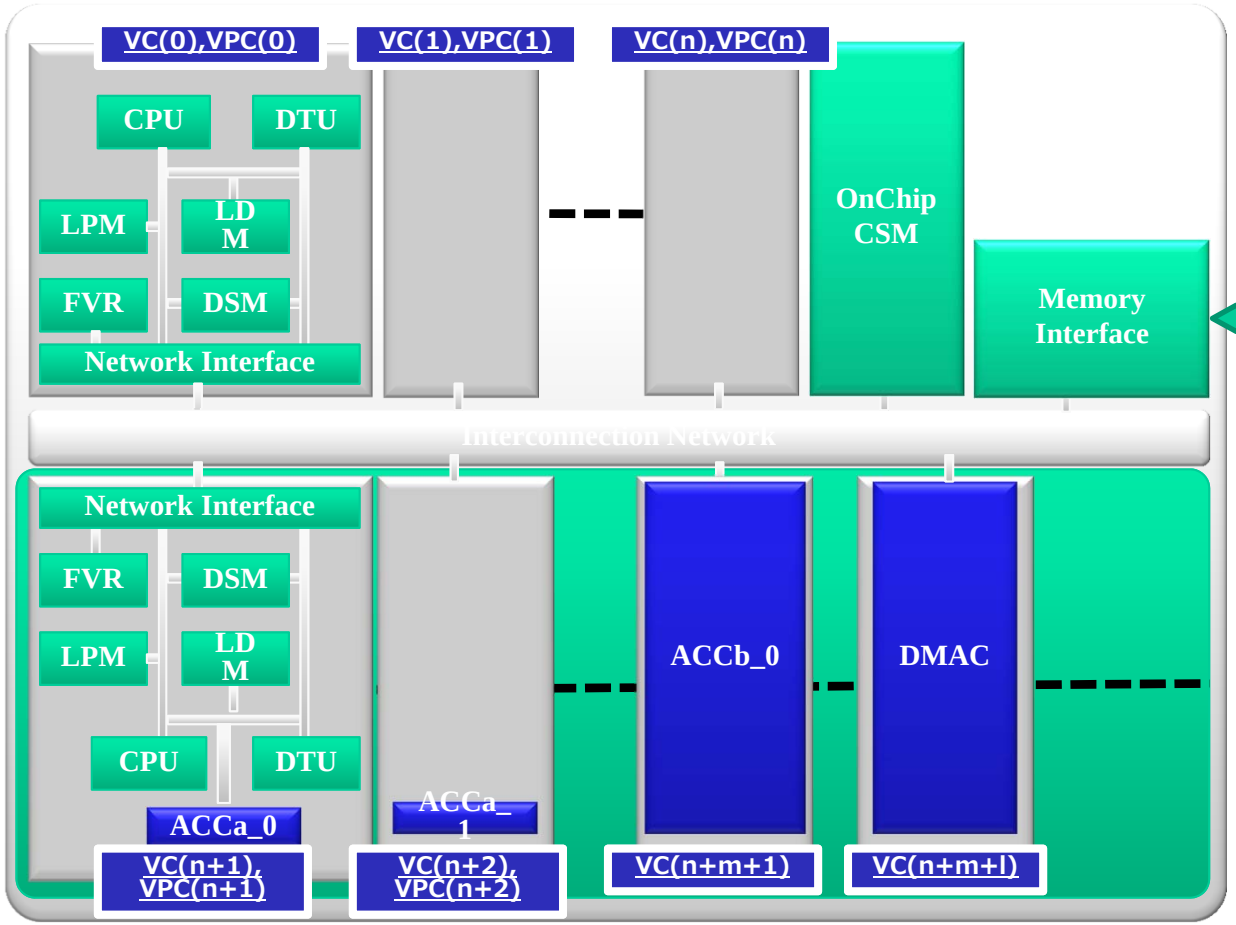


On Nexus7, 3 core parallelization gave us

for DrawRect **1.91** speedup

for DrawImage **1.95** speedup

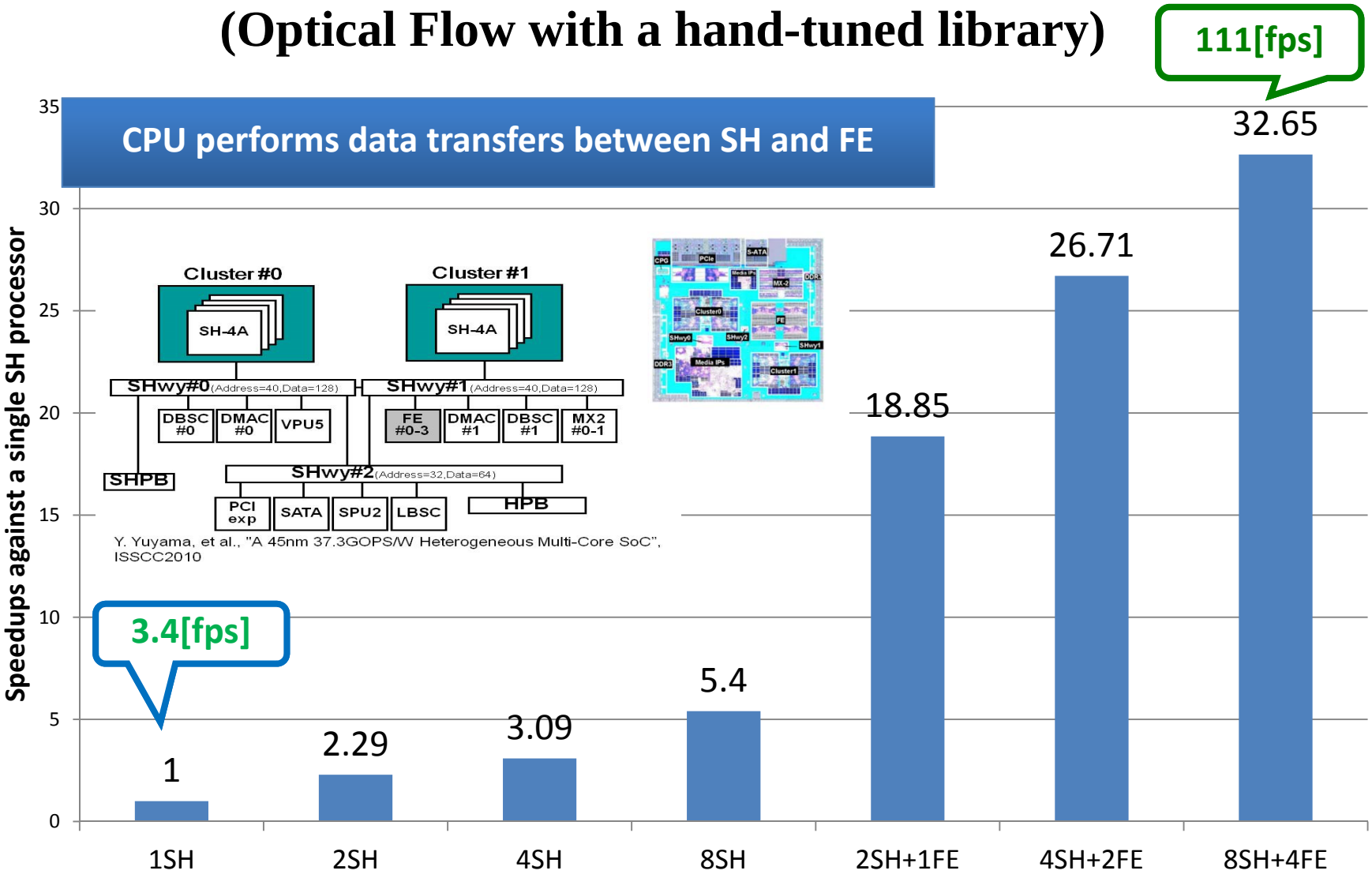
OSCAR API-Applicable Heterogeneous Multicore Architecture



- 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

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)

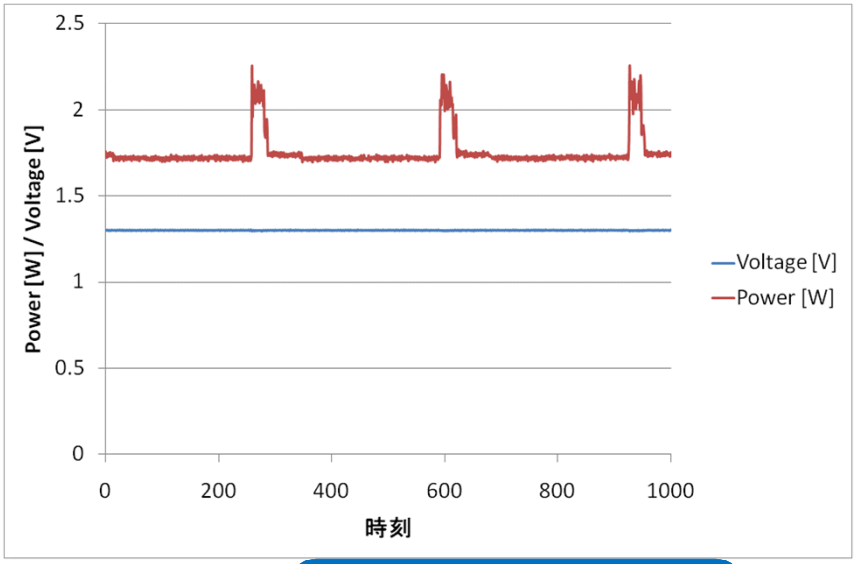
Without Power Reduction

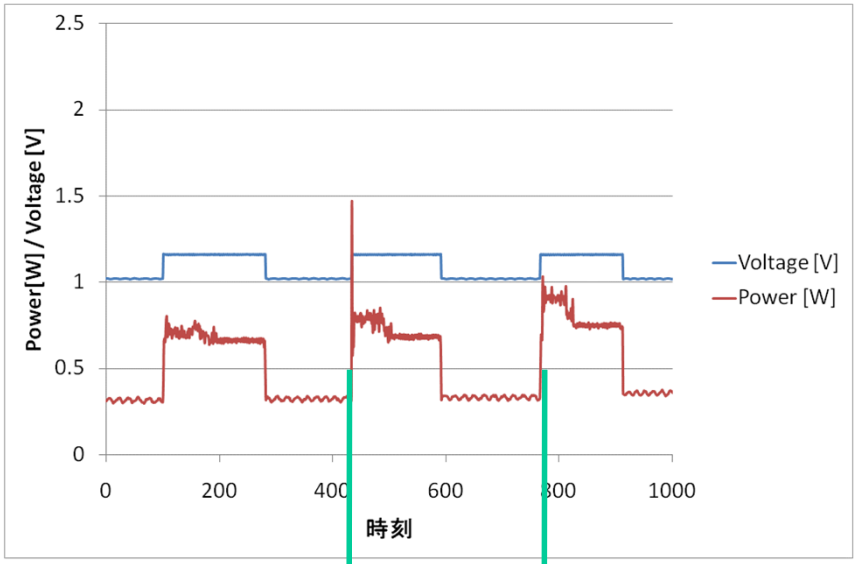
With Power Reduction by OSCAR Compiler

70% of power reduction

Average:1.76[W]

Average:0.54[W]





1cycle : 33[ms]
→30[fps]

OSCAR Parallelizing Compiler

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

Multigrain Parallelization

coarse-grain parallelism among loops and subroutines, near fine grain parallelism among statements in addition to loop parallelism

Data Localization

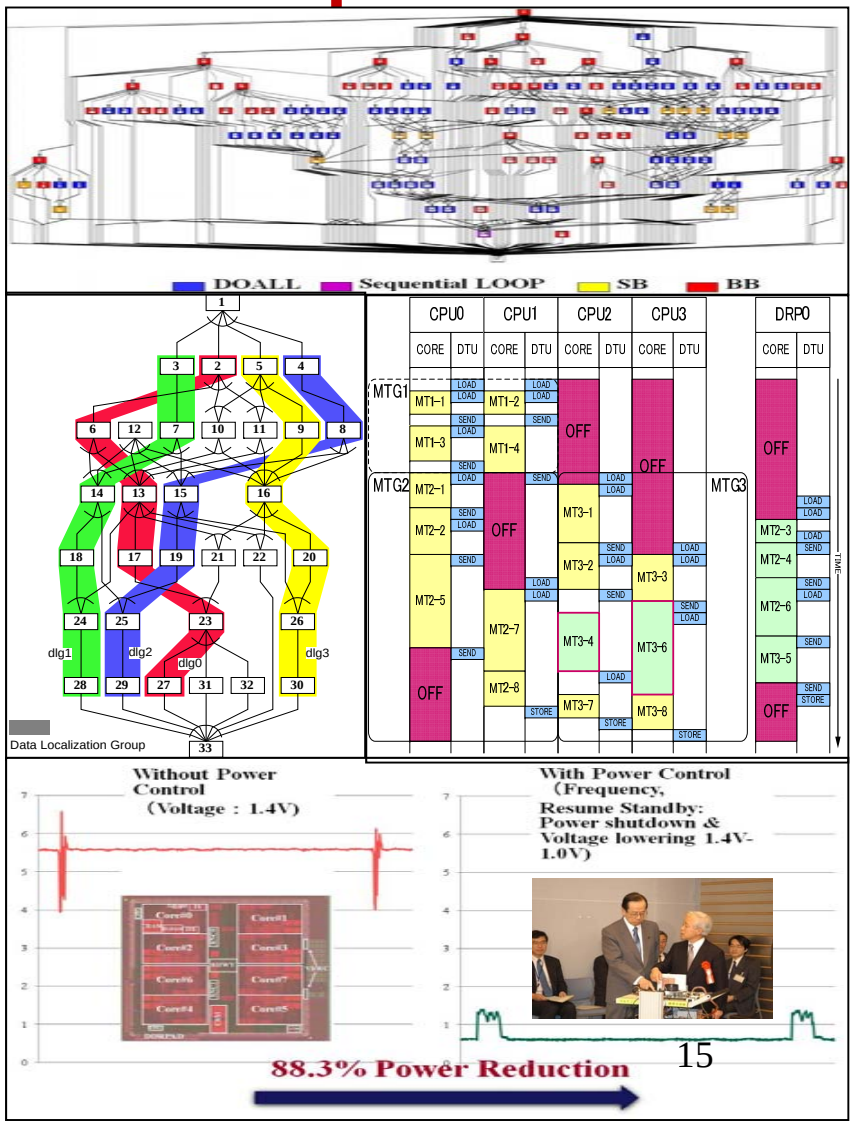
Automatic data management for distributed shared memory, cache and local memory

Data Transfer Overlapping

Data transfer overlapping using Data Transfer Controllers (DMAs)

Power Reduction

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

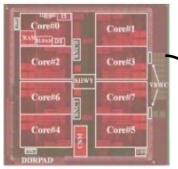
Accelerator 2
Code

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

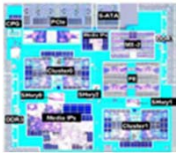
Low Power Heterogeneous Multicore Code Generation
API Analyzer (Available from Waseda) | Existing sequential compiler

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

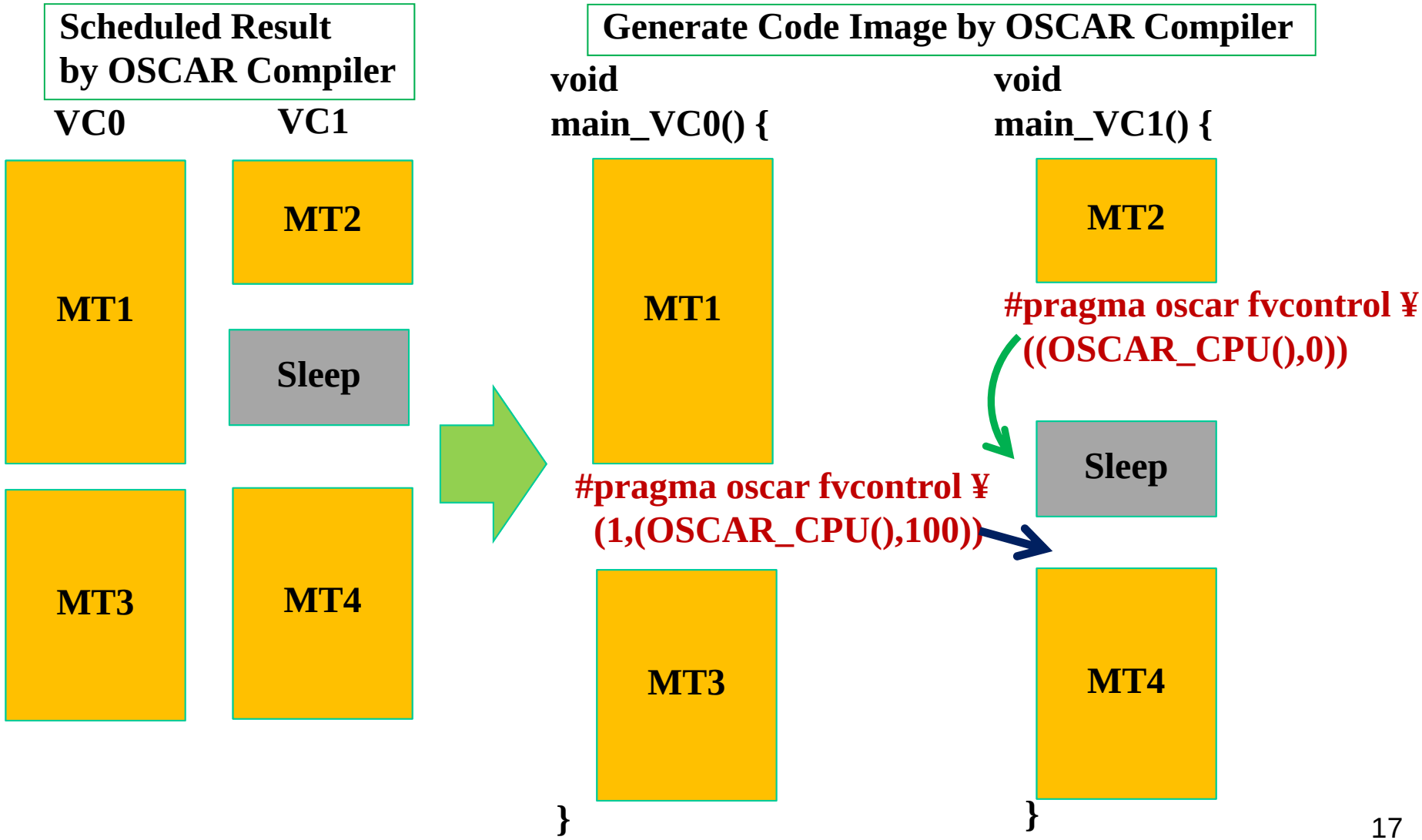
Shred memory servers

Executable on various multicores

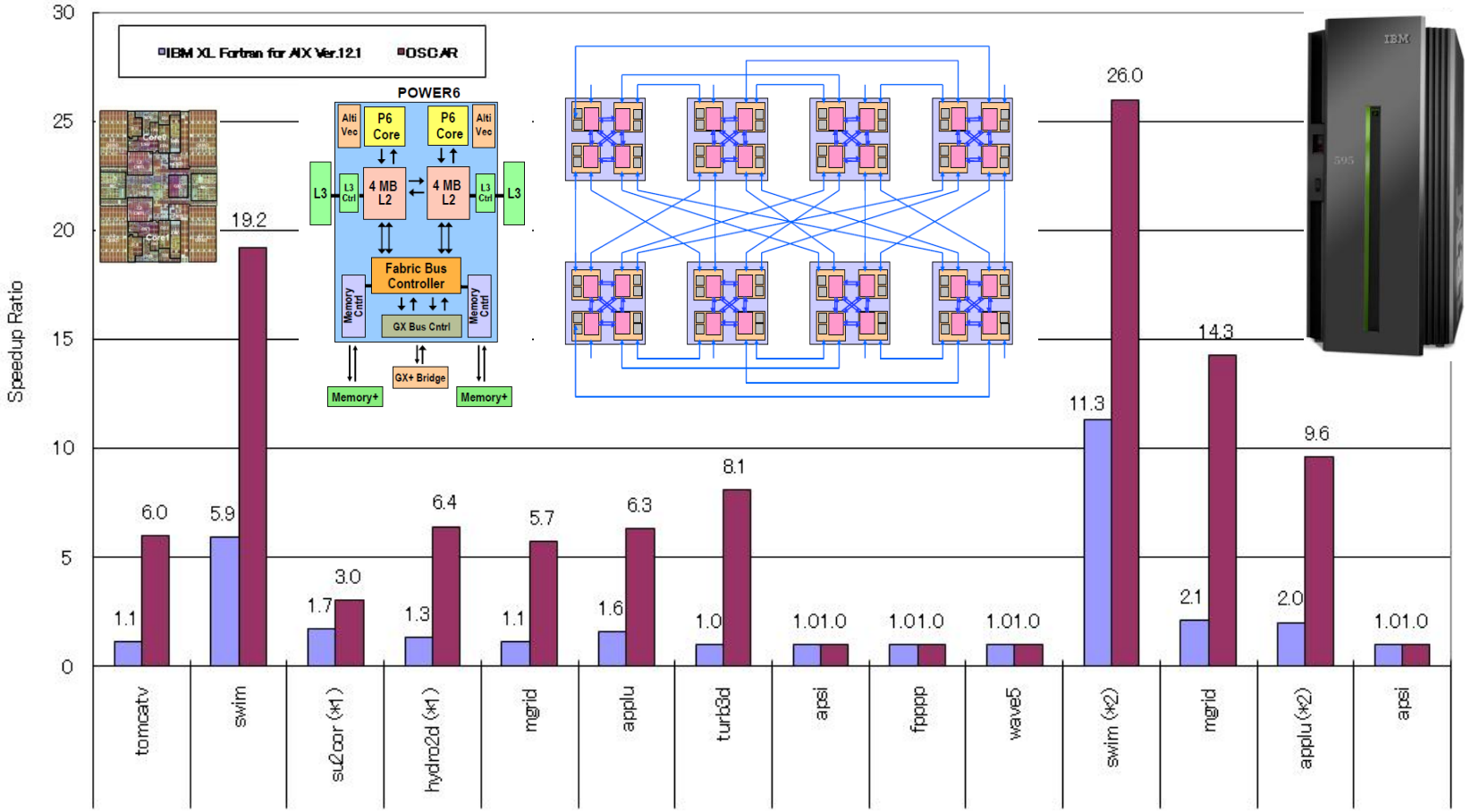
Hitachi, Renesas, NEC, Fujitsu, Toshiba, Denso, Olympus, Mitsubishi, Esol, Cats, Gaio, 3 univ.

OSCAR: Optimally Scheduled Advanced Multiprocessor
API : Application Program Interface

Low-Power Optimization with OSCAR API



Performance of OSCAR Compiler on IBM p6 595 Power6 (4.2GHz) based 32-core SMP Server

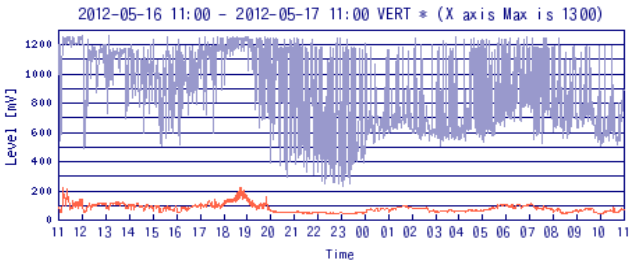


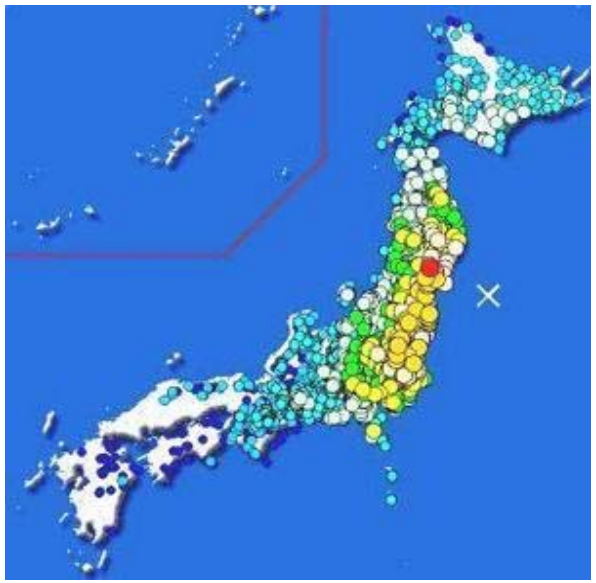
OpenMP codes generated by OSCAR compiler accelerate IBM XL Fortran for AIX Ver.12.1 about 3.3 times on the average

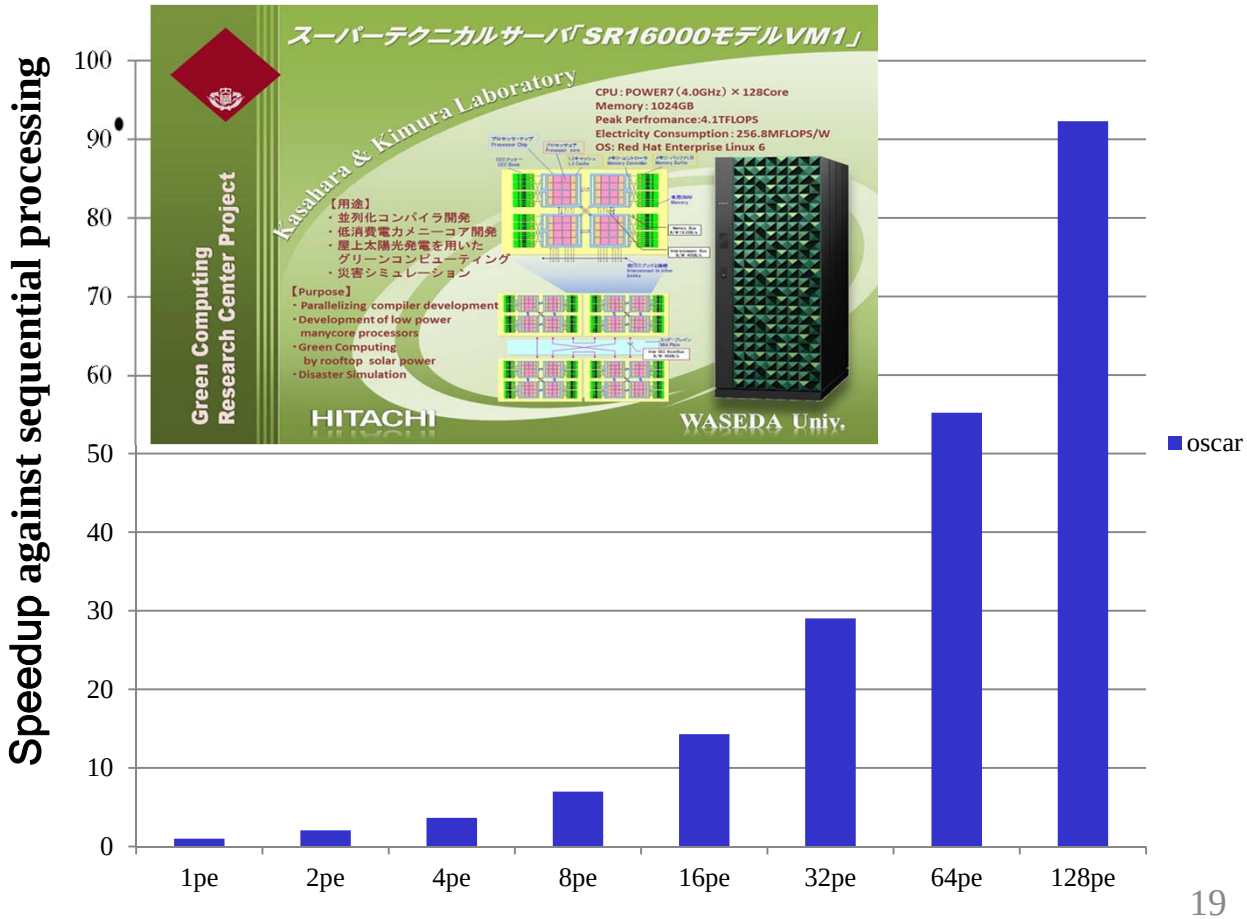
Compile Option:
(*1) Sequential: -O3 –qarch=pwr6, XLF: -O3 –qarch=pwr6 –qsmp=auto, OSCAR: -O3 –qarch=pwr6 –qsmp=noauto
(*2) Sequential: -O5 -q64 –qarch=pwr6, XLF: -O5 –q64 –qarch=pwr6 –qsmp=auto, OSCAR: -O5 –q64 –qarch=pwr6 –qsmp=noauto
(Others) Sequential: -O5 –qarch=pwr6, XLF: -O5 –qarch=pwr6 –qsmp=auto, OSCAR: -O5 –qarch=pwr6 –qsmp=noauto

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

（Power7 Based 128 Core Linux SMP）

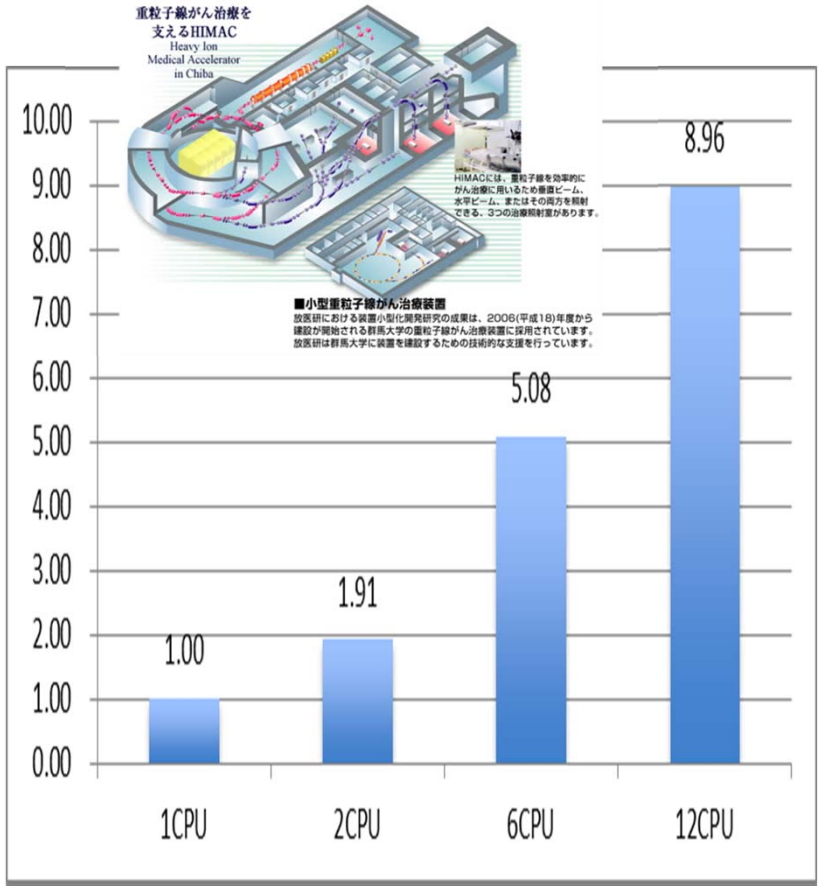


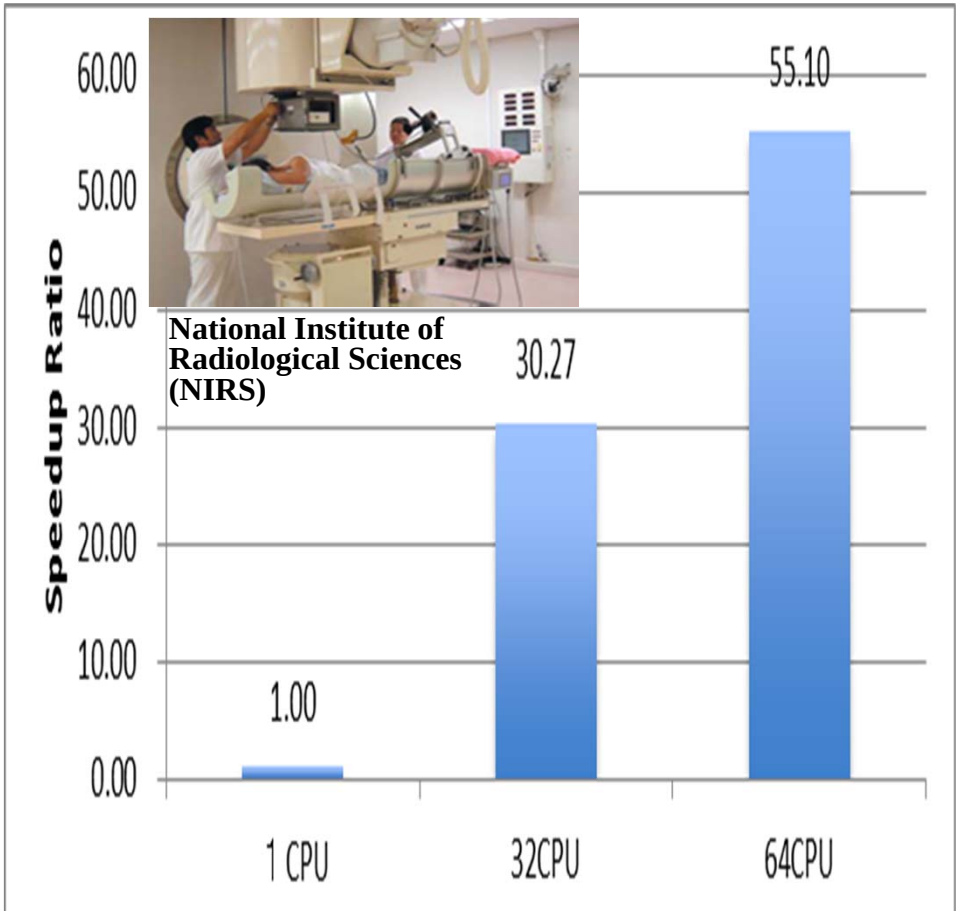




Cancer Treatment
Carbon Ion Radiotherapy

(Previous best was 2.5 times speedup on 16 processors with hand optimization)





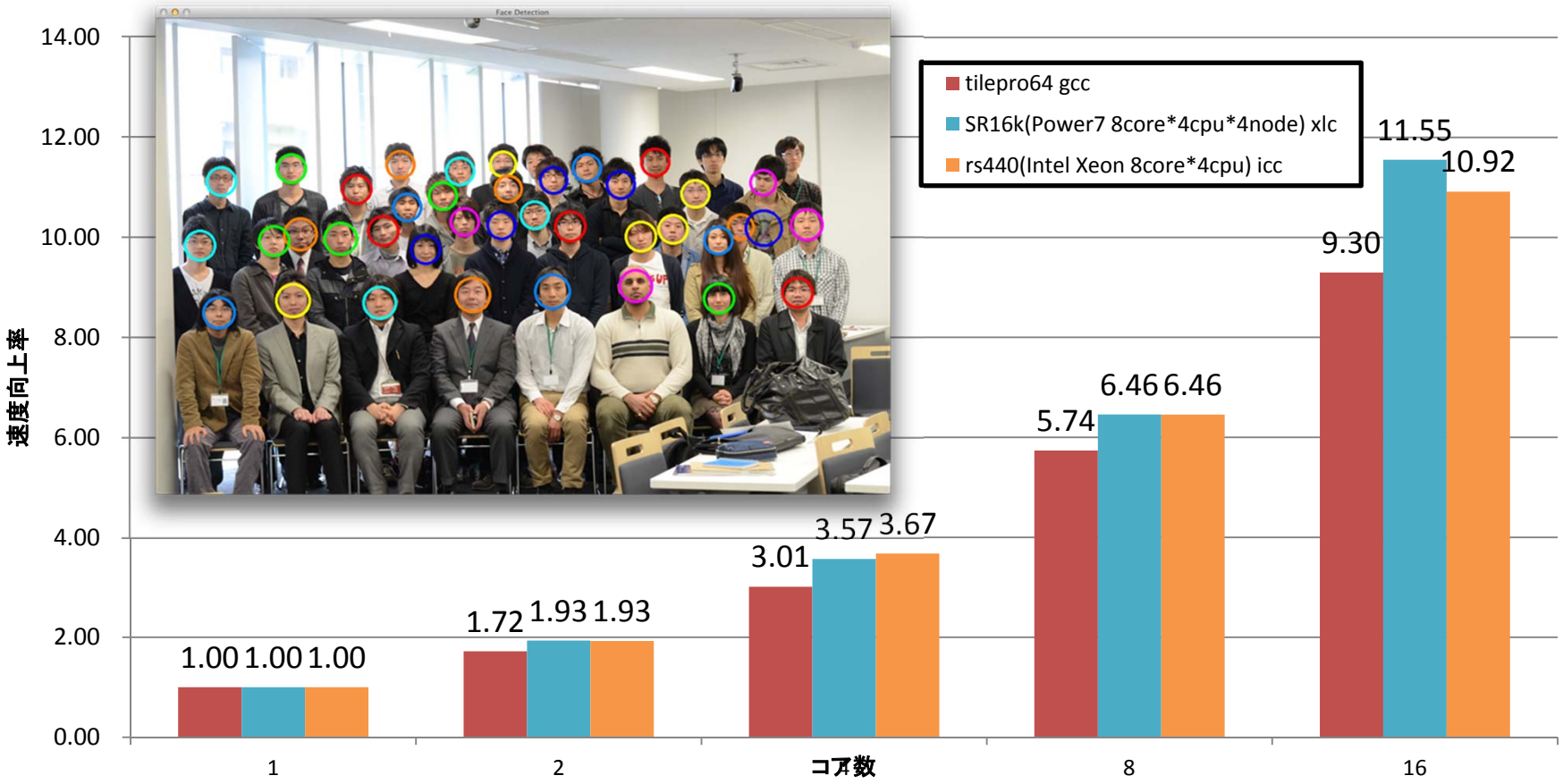
8.9times speedup by 12 processors

Intel Xeon X5670 2.93GHz 12 core SMP (Hitachi HA8000)

55 times speedup by 64 processors

IBM Power 7 64 core SMP (Hitachi SR16000)

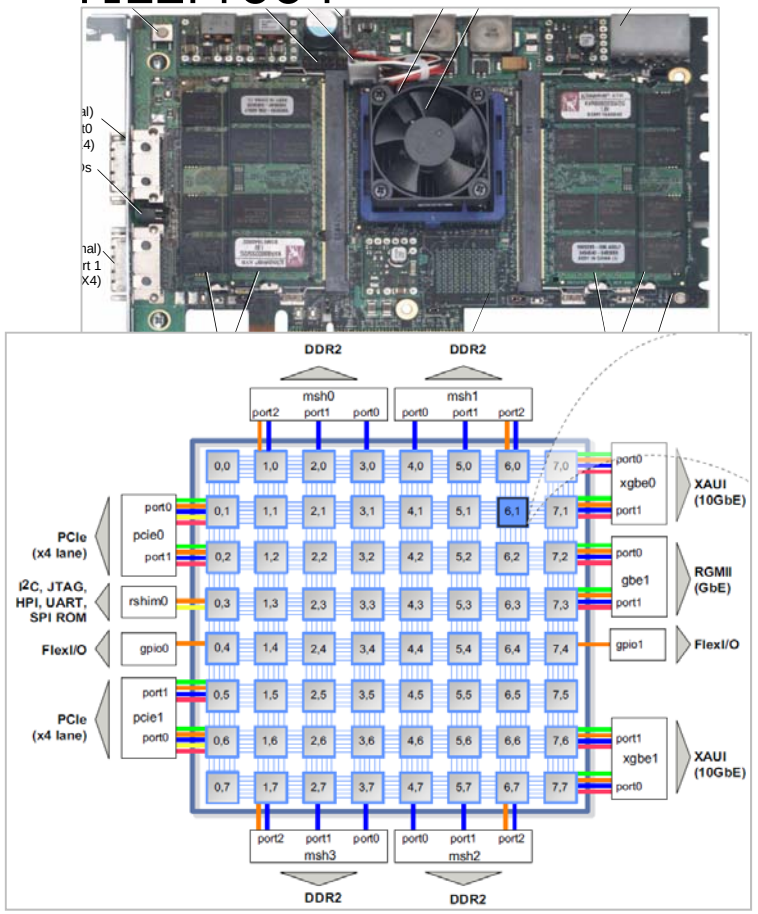
Parallel Processing of Face Detection on Manycore, Highend and PC Server

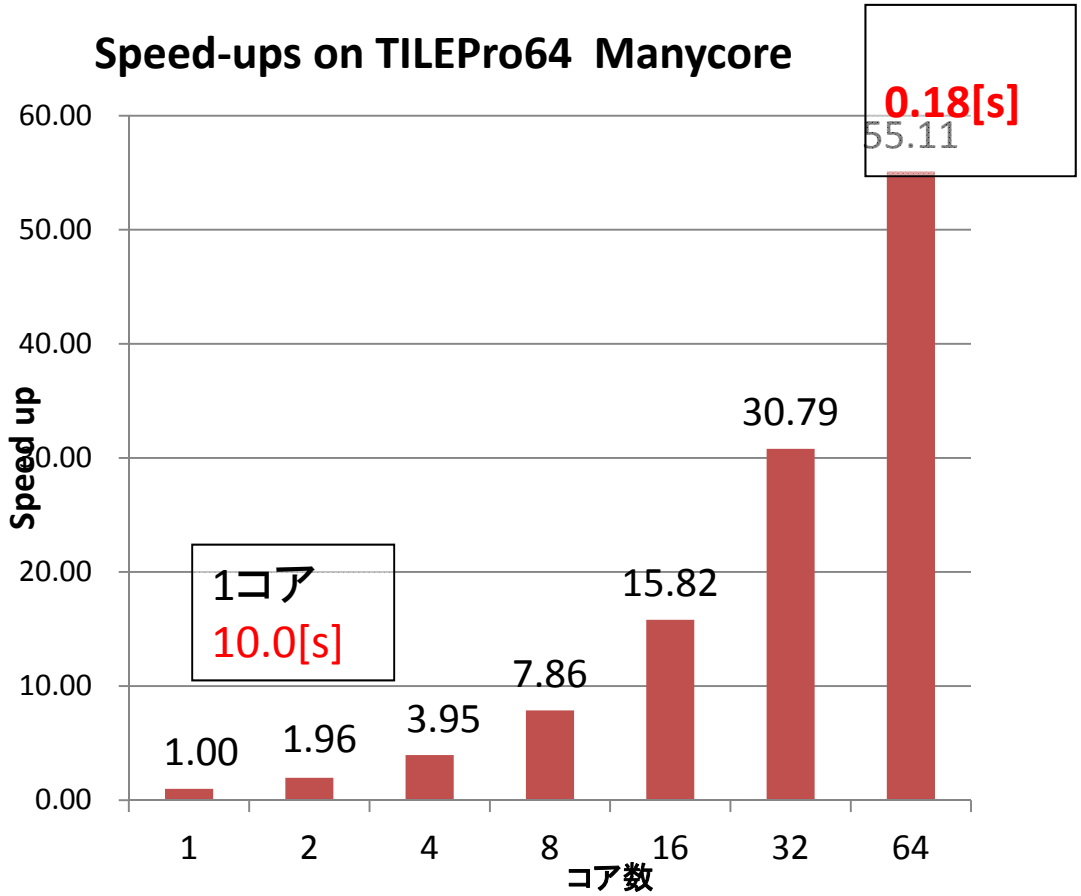


- OSCAR compiler gives us 11.55 times speedup for 16 cores against 1 core on SR16000 Power7 highend server.

Automatic Parallelization of Still Image Encoding Using JPEG-XR for the Next Generation Cameras and Drinkable Inner Camera

- TILEPro64



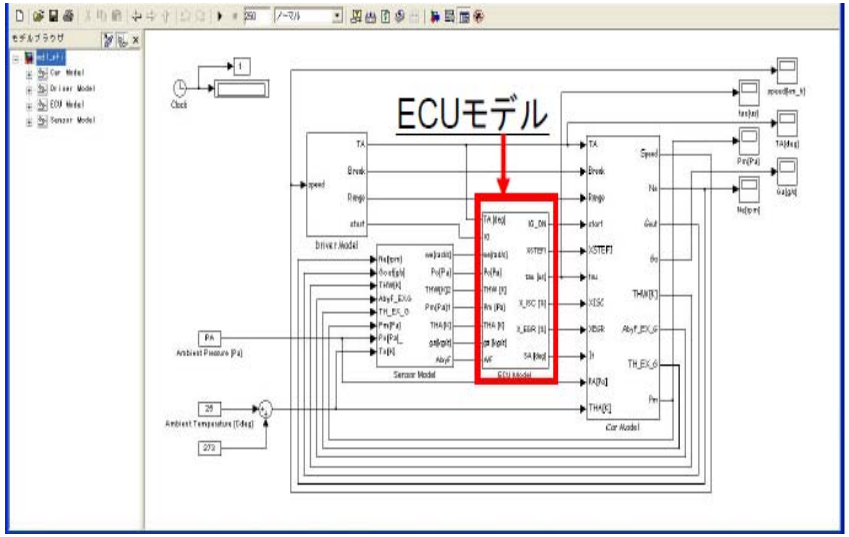


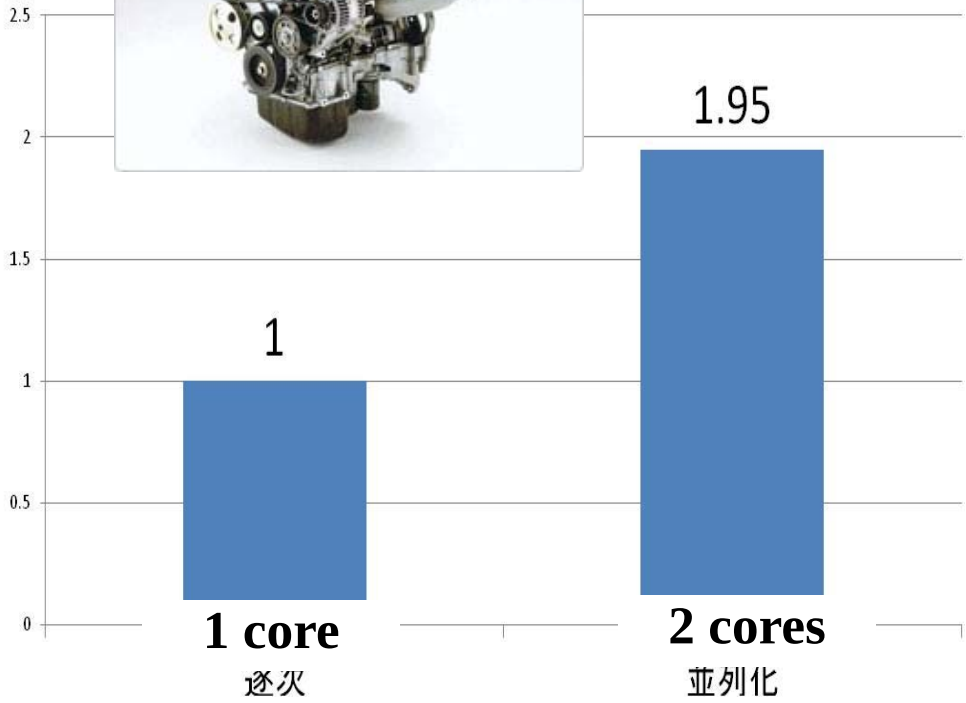
55 times speedup with 64 cores against 1 core

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





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