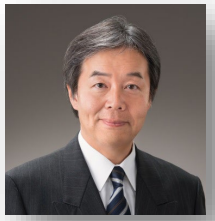
Green Multicore Computing



Prof. Hironori Kasahara, IEEE Life-Fellow, IPSJ Fellow

Senior Executive Vice President (2018-2022), Waseda University

IEEE Computer Society President 2018

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., 1989-90 Research Scholar: U. of Illinois, Urbana-Champaign, Center for Supercomputing R&D, 1997 Prof.,
2004 Director, Advanced Multicore Research Institute,
2017member: the Engineering Academy of Japan (2020- Board Mem) and the Science Council of Japan
2018 IEEE Computer Society President
Senior Exectutive Vice President, Waseda U. (2018-22)

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 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, , 2023 IEEE Life Fellow

Reviewed Papers: 243, Invited Talks: 244,
Granted Patents: 70 (Japan, US, GB, DE, China),
Articles in News Papers, Web News, TV etc.: 709

Committees in Societies and Government 300
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 Development Member (Top 1:NWT, Earth Simulator, K)
JST SPRING Ph.D Promotion Chair, Boost AI Ph.D. Boosting Chair, SBIR Phase 1 Chair, Moonshot Project G3 Robot & AI Advisor & Int'l Symp Vice Chair,
【COCN】Council of Competitiveness Nippon Ex-Board

PRESENTED BY

TOP 500

The List.

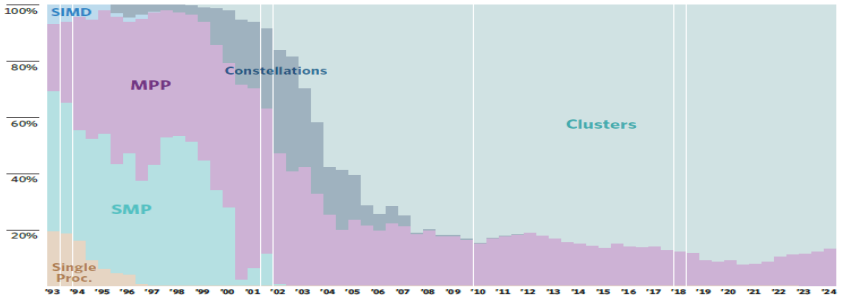
MAY 2024

FIND OUT MORE AT top500.org

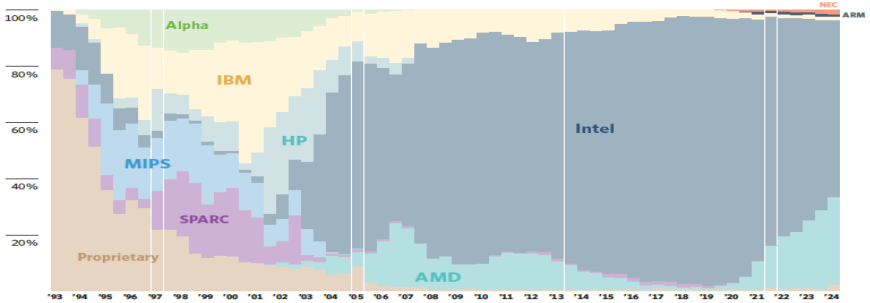
https://www.top500.org/

			SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	**Frontier**	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	DOE/SC/ORNL	USA	8,699,904	1,206.0	22.7
2	**Aurora**	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 (52C 2.4GHz), Intel Data Center GPU Max, Slingshot-11	DOE/SC/ANL	USA	9,264,128	1,012.0	38.7
3	**Eagle**	Microsoft NDv5, Xeon Platinum 8480C (48C 2GHz), NVIDIA H100, NVIDIA Infiniband NDR	Microsoft Azure	USA	1,123,200	561.2	
4	**Fugaku**	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
5	**LUMI**	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	EuroHPC/CSC	Finland	2,220,288	379.7	6.01

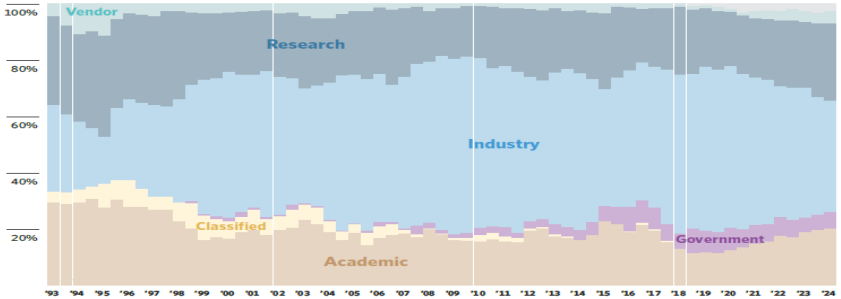
ARCHITECTURES



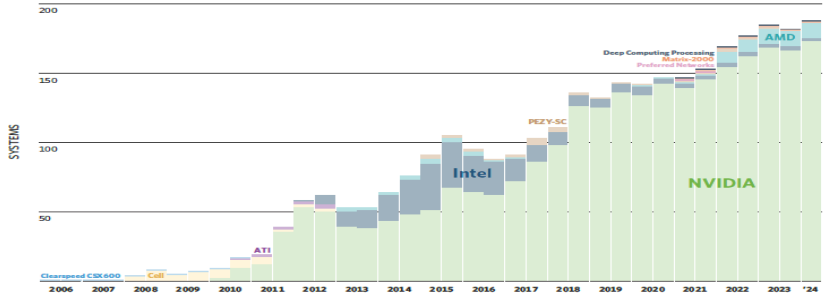
CHIP TECHNOLOGY



INSTALLATION TYPE



ACCELERATORS/CO-PROCESSORS



https://olcf.ornl.gov/wp-content/uploads/Frontiers-Architecture-Frontier-Training-Series-final.pdf

OLCF Systems by the numbers

System	Titan (2012)	Summit (2017)	Frontier (2021)
Peak	27 PF	200 PF	2.0 EF
# nodes	18,688	4,608	9,408
Node	1 AMD Opteron CPU 1 NVIDIA Kepler GPU	2 IBM POWER9™ CPUs 6 NVIDIA Volta GPUs	1 AMD EPYC "Trento" CPU 4 AMD Instinct MI250X GPUs
Memory	0.6 PB DDR3 + 0.1 PB GDDR	2.4 PB DDR4 + 0.4 HBM + 7.4 PB On-node storage	4.6 PB DDR4 + **4.6 PB HBM2e** + 36 PB On-node storage, 75 TB/s Read 38 Write
On-node interconnect	PCI Gen2 No coherence across the node	NVIDIA NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node
System Interconnect	Cray Gemini network 6.4 GB/s	Mellanox Dual-port EDR IB 25 GB/s	Four-port Slingshot network 100 GB/s
Topology	3D Torus	Non-blocking Fat Tree	Dragonfly
Storage	32 PB, 1 TB/s, Lustre Filesystem	250 PB, 2.5 TB/s, IBM Spectrum Scale™ with GPFS™	695 PB HDD+11 PB Flash Performance Tier, 9.4 TB/s and 10 PB Metadata Flash Lustre
Power	9 MW	13 MW	29 MW

Amazon buys nuclear-powered data center from Talen

Thu, Mar 7, 2024, 10:01PM | Nuclear News

Susquehanna nuclear plant in Salem Township, Penn., along with the data center in foreground. (Photo: Talen Energy)

Talen Energy announced its sale of a 960-megawatt data center campus to cloud service provider Amazon Web Services (AWS), a subsidiary of Amazon, for $650 million.

The data center, Cumulus Data Assets, sits on a 1,200-acre campus in Pennsylvania and is directly powered by the adjacent Susquehanna Steam Electric Station, which generates 2.5 gigawatts of power.

From Smartphones, Smart Cars to HPC use Multicores & Accelerators

Cores: ARM, IBM, Renesas, Infineon, SPARC, RISC V, **Accelerators:** NVIDIA, Intel, AMD, etc.

https://www.apple.com/jp/shop/buy-iphone

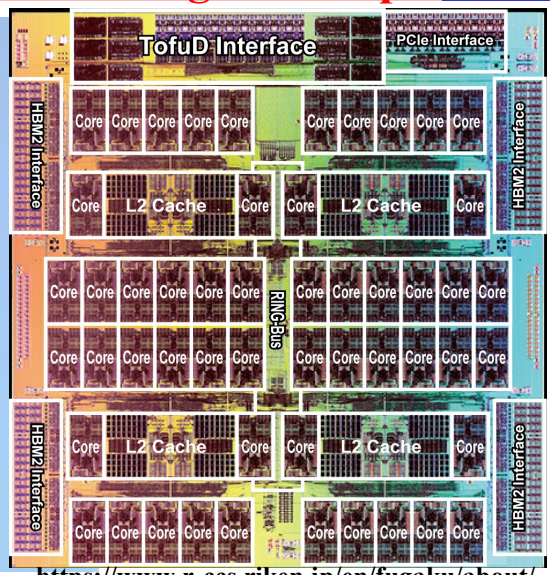
64-bit
iPhone 13
2021

Launched September 14, 2021
Designed by Apple Inc.
Common manufacturer(s) : TSMC
Max. CPU clock rate to 3.23 GHz in iPhone 13 Pro
Technology node: 5 nm
6 Cores: 2 "Avalanche"High Perf. & 4 "Blizzard" Low Power
Instruction set: ARM A64, Transistors: 15 billion (15億個)
GPU(s): Apple-designed 5 core GPU in iPhone 13

https://en.wikipedia.org/wiki/Apple_A15

Riken Fugaku : Top1 June, 2020-Nov.2021

https://fugaku100kei.jp/fugaku/



https://www.r-ccs.riken.jp/en/fugaku/about/

RIKEN Center for Computational Science,
Fujitsu（arm based processor）
Cores:7,299,072； Memory:4,866,048GB；
Processor:A64FX 48Cores, 2.2GHz
Interconnect: Tofu interconnect D
Linpack (Rmax)415,530 TFlop/s;
Theoretical Peak (Rpeak): 513PFLOPS
HPCG [TFlop/s]13,366.4; Power: 28.3MW

48cores/chip, 2.2GHz, 7 nm FinFET,
約7百30万コア, 28MW
理論最高性能:51京回浮動小数点演算/秒,
2020年6月時点

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

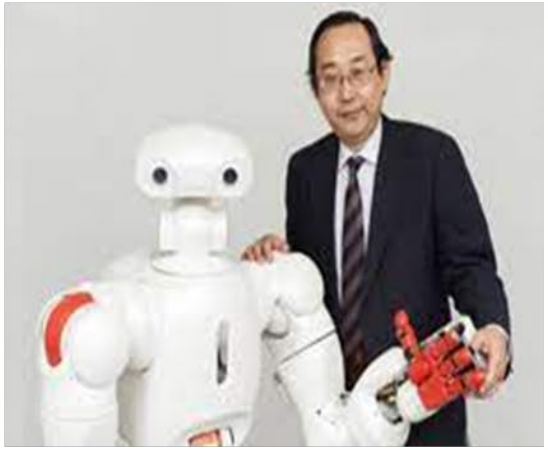
https://www8.cao.go.jp/cstp/english/moonshot/sub3_en.html

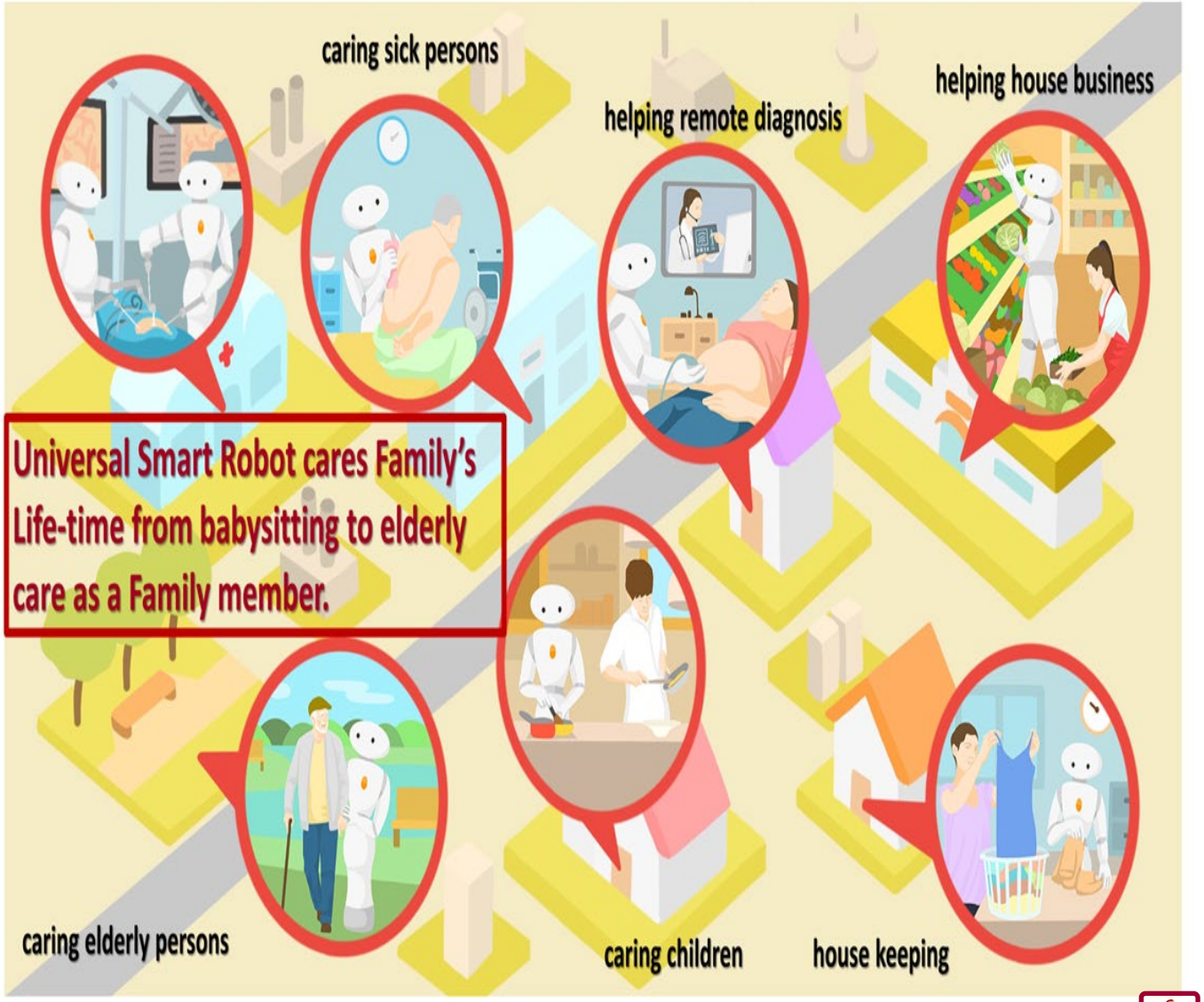
Cabinet Office

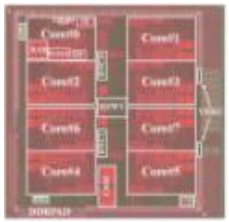
MOONSHOT
RESEARCH & DEVELOPMENT PROGRAM

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

Supported by Japanese Government "Moonshot" Project from 2020

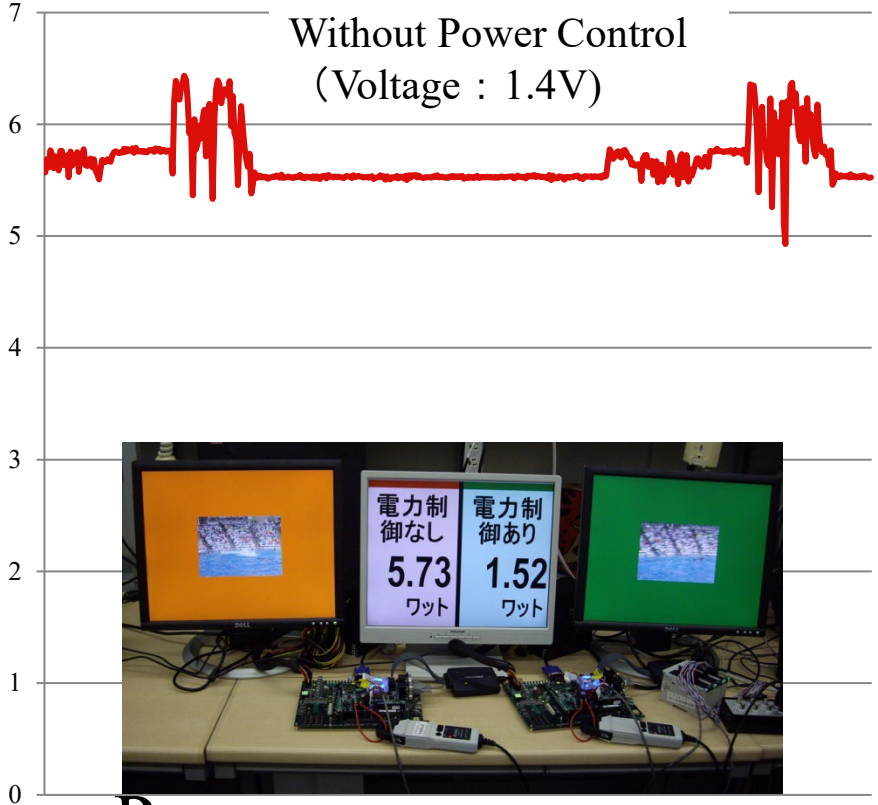


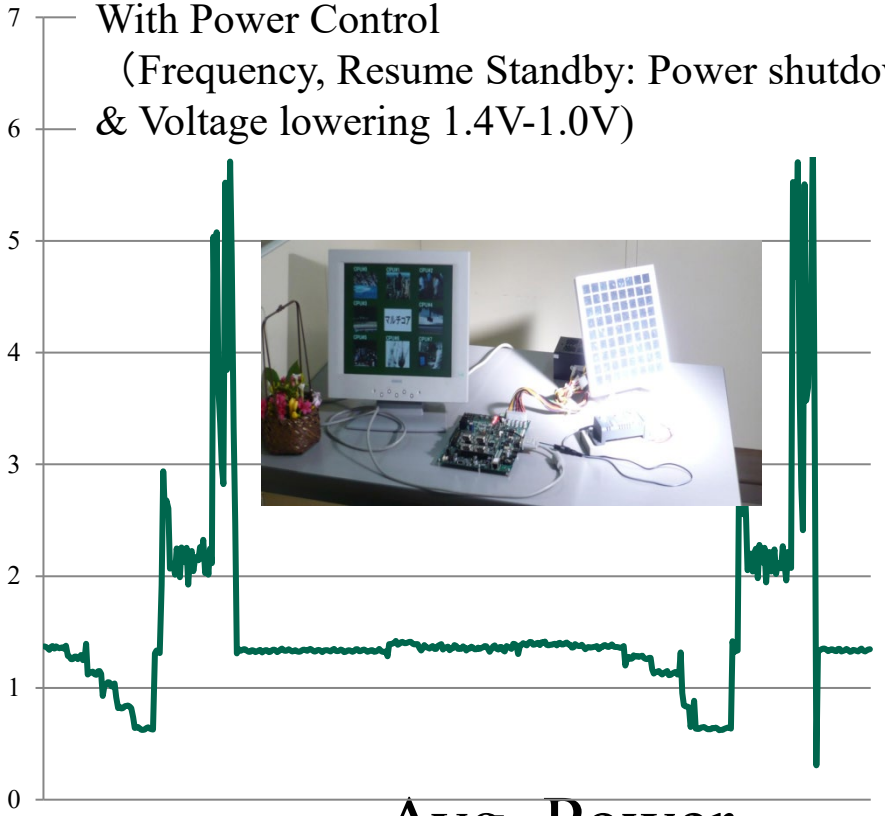




Power Reduction of MPEG2 Decoding to 1/4 on 8 Core Homogeneous Multicore RP-2 by OSCAR Parallelizing Compiler

MPEG2 Decoding with 8 CPU cores





Avg. Power
5.73 [W]

73.5% Power Reduction

Avg. Power
1.52 [W]

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)

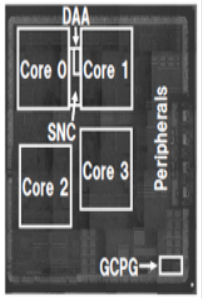
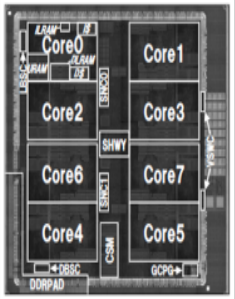
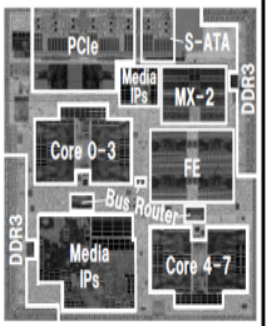
第74回総合科学技術会議の様子(2)


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
97.6 mm^2 (9.88 x 9.88 mm)	104.8 mm^2 (10.61 x 9.88 mm)	153.8 mm^2 (12.4 x 12.4 mm)
1.0V (internal), 1.8/3.3V (I/O)	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)	18.3 GOPS/W (32b換算)	37.3 GOPS/W (32b換算)

Prime Minister FUKUDA is touching our multicore chip during execution.

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

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

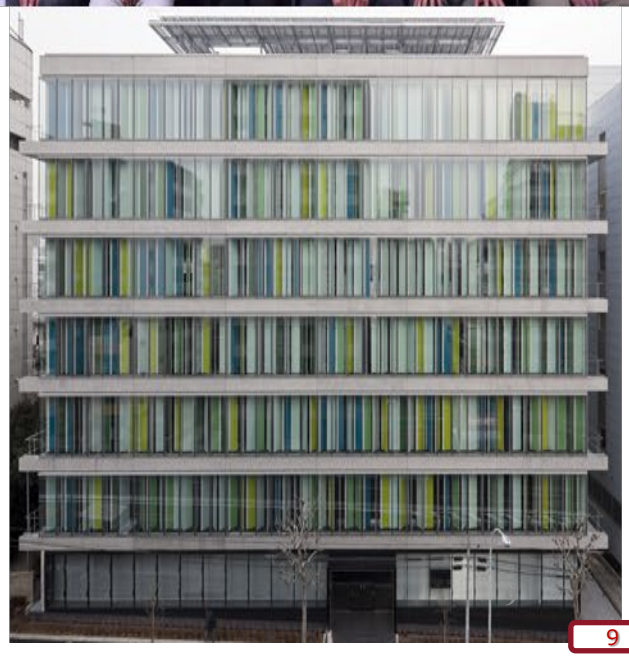
Keynote: Paul Messina ANL, USA

- 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

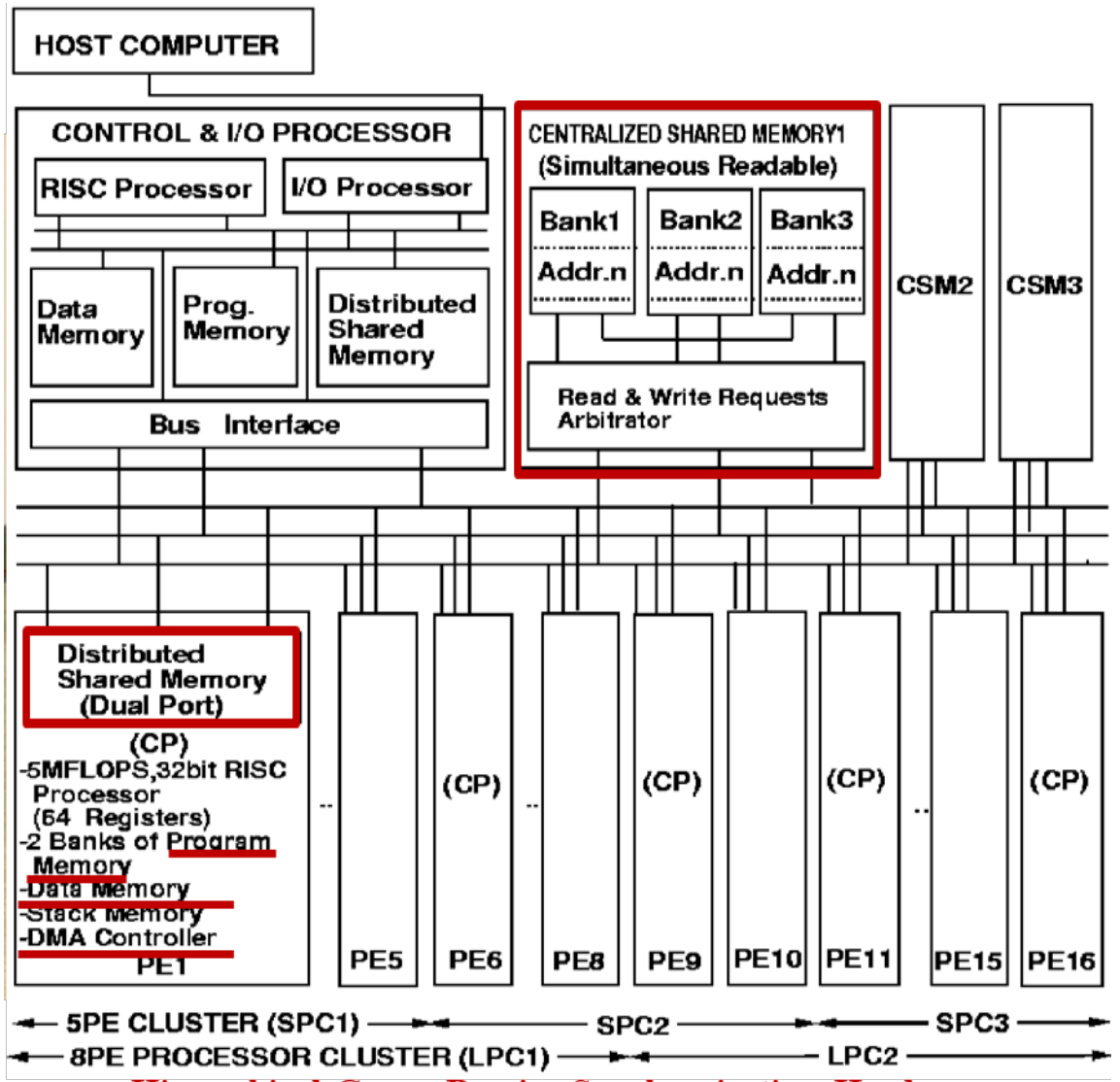


The First Compiler Codesigned Multiprocessor

OSCAR (Optimally Scheduled Advanced Multiprocessor) in 1987

AMD29325 32-bit Floating-point unit

H. Kasahara, "OSCAR Fortran Multigrain Compiler", Stanford University, Hosted by Professor John L. Hennessy and Professor Monica Lam, May. 15. 1995.



AMD29325 32-bit Floating-point unit

Hierarchical Group Barrier Synchronization Hardware

https://ieeexplore.ieee.org/document/130111/similar#similar

Parallel processing of near fine grain tasks using static scheduling on OSCAR (optimally scheduled advanced multiprocessor)

Publisher: IEEE | Cite This | PDF

H. Kasahara ; H. Honda ; S. Narita All Authors

Published in: Supercomputing '90:Proceedings of the 1990 ACM/IEEE Conference on Supercomputing

https://ieeexplore.ieee.org/document/1653942

Journals & Magazines > Computer > Volume: 15 Issue: 2

A Second Opinion on Data Flow Machines and Languages

Publisher: IEEE | Cite This | PDF

Gajski ; Padua ; Kuck ; Kuhn All Authors

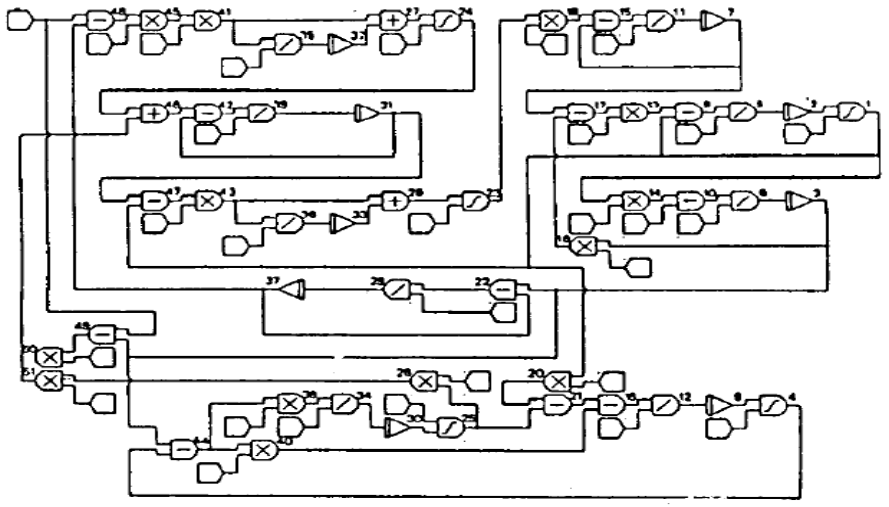
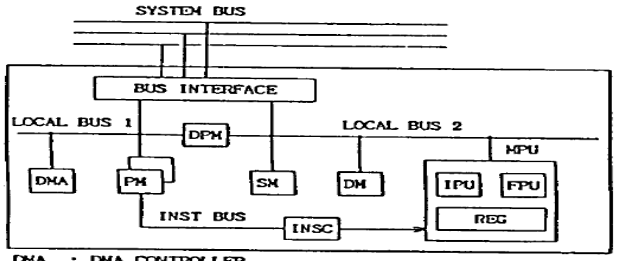


Fig.10 A kind of dataflow graph with near fine grain tasks



DMA : DMA CONTROLLER
PM : PROGRAM MEMORY (128KW×2BANK)
INSC : INSTRUCTION CONTROL UNIT
DPM : DUAL PORT MEMORY (32bit×2KW)
SM : STACK MEMORY (32bit×4KW)
DM : DATA MEMORY (32bit×256KW)
MPU : MAIN PROCESSING UNIT
IPU : INTEGER PROCESSING UNIT
FPU : FLOWTING PROCESSING UNIT
REG : REGISTER FILE (32bit×64lines)
INST BUS : INSTRUCTION BUS

Fig.2 OSCAR's processor element

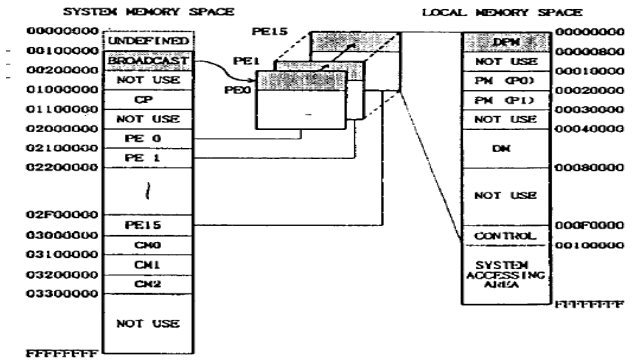


Fig.3 OSCAR's memory space

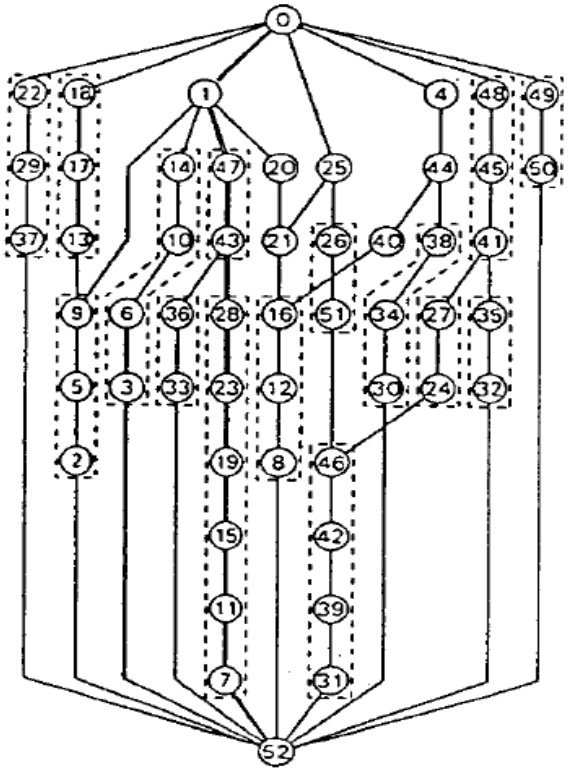
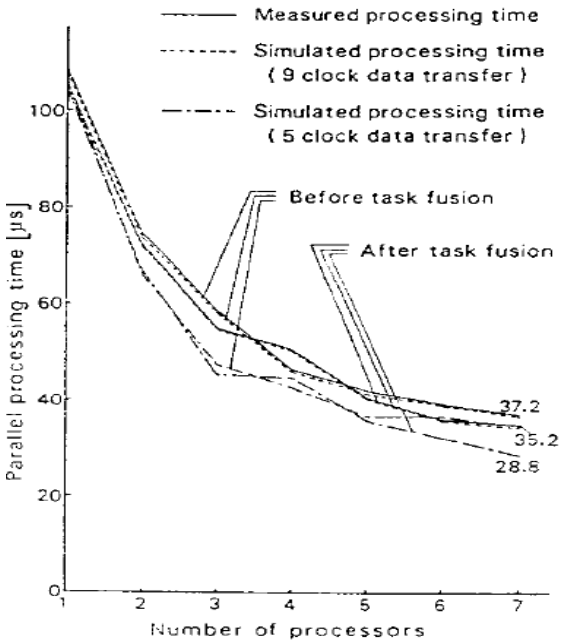


Fig.11 Task graph for Fig. 10



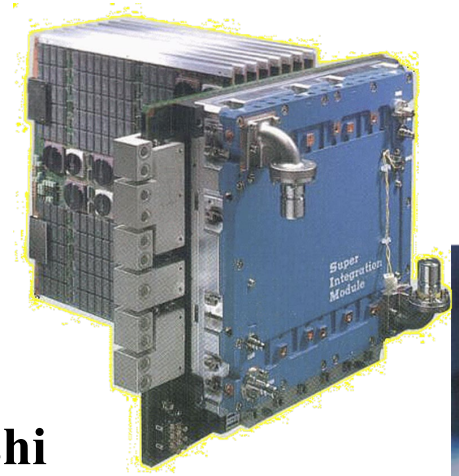
Parallel processing time of fig.10 on OSCAR

tioned in Section 3.1. The measured processing times were reduced from 108.7 [us] and 105.8 [us] for one PE to 37.2 [us] (1/2.92) and 35.2 [us](1/3.01) for seven PEs respectively.

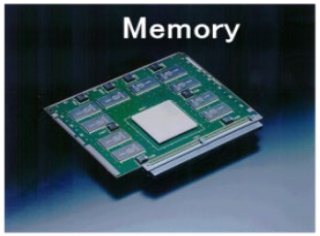
NWT(VPP500) First Japanese Supercomputer became Top1, 1993

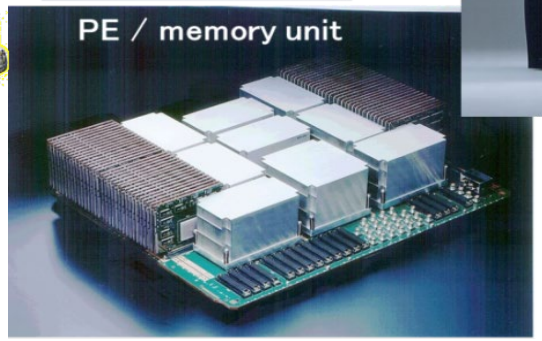
ACM/IEEE SC '94: Washington, D.C. November, 1994

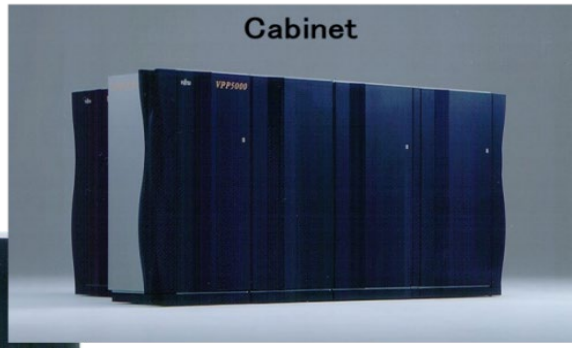
Mr. Hajim Miyoshi

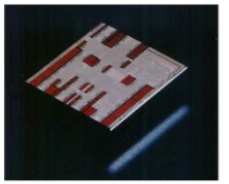


スパーコンピュータNWTの外観



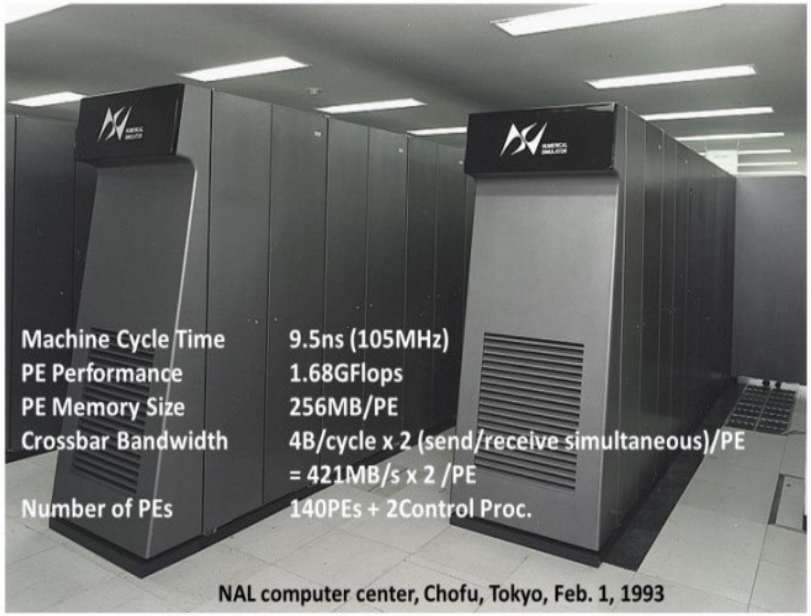


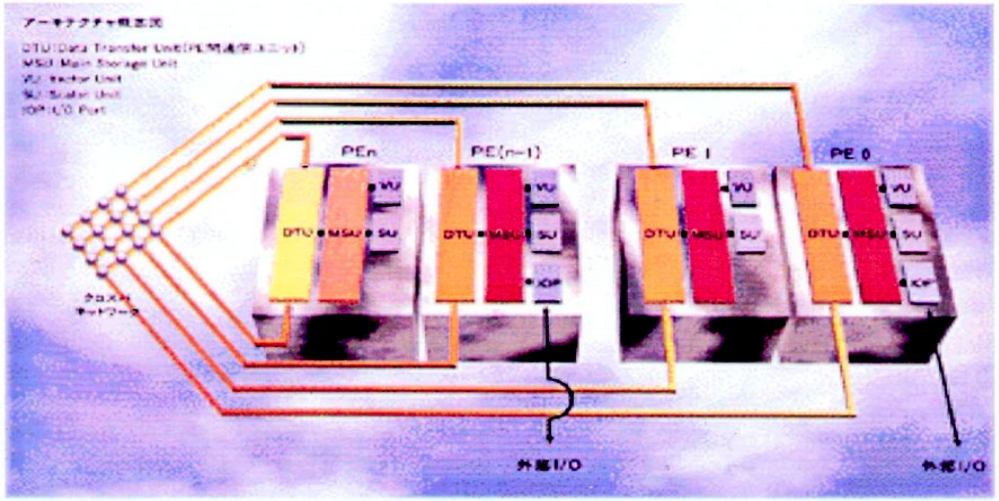




CMOS LSI

商用VPP5000 (仏気象庁他)





Earth Simulator

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

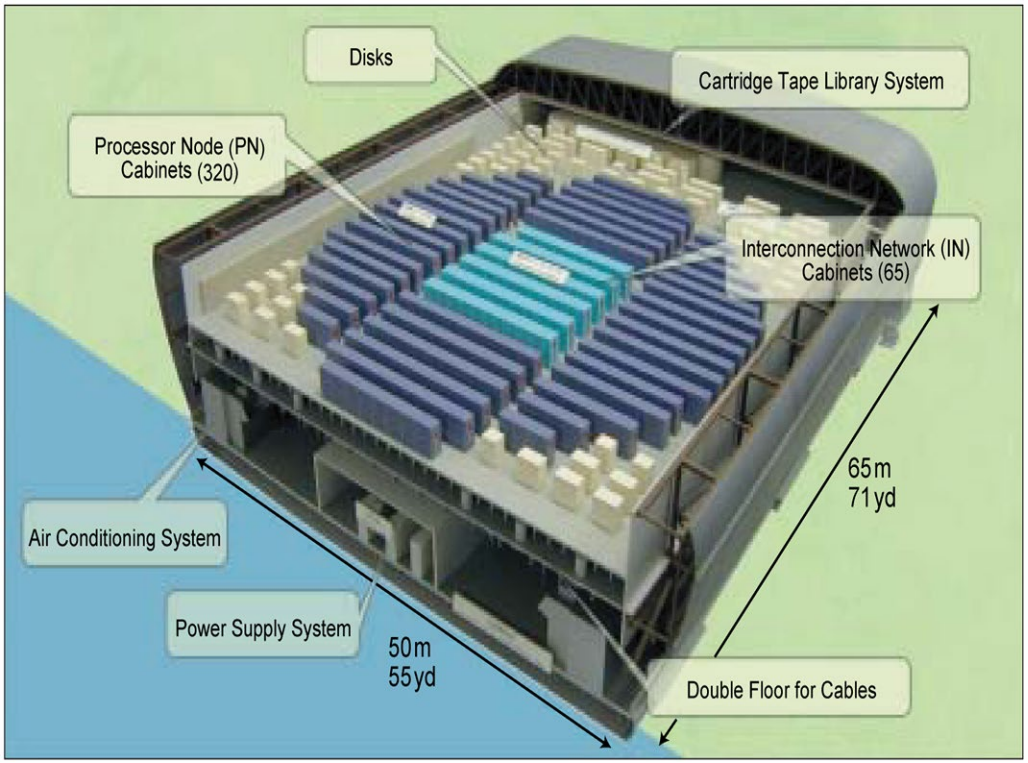
- Earth Environmental simulation like Global Warming, El Nino, PlateMovement for the all lives onr this planet.
- Developed in Mar. 2002 by STA (MEXT) and NEC with 400 M$ investment under Dr. Miyoshi's direction.

(Dr.Miyoshi: Passed away in Nov.2001. NWT, VPP500, SX6)

Mr. Hajime Miyoshi

Image of Earth Simulator

4 Tennis Courts



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

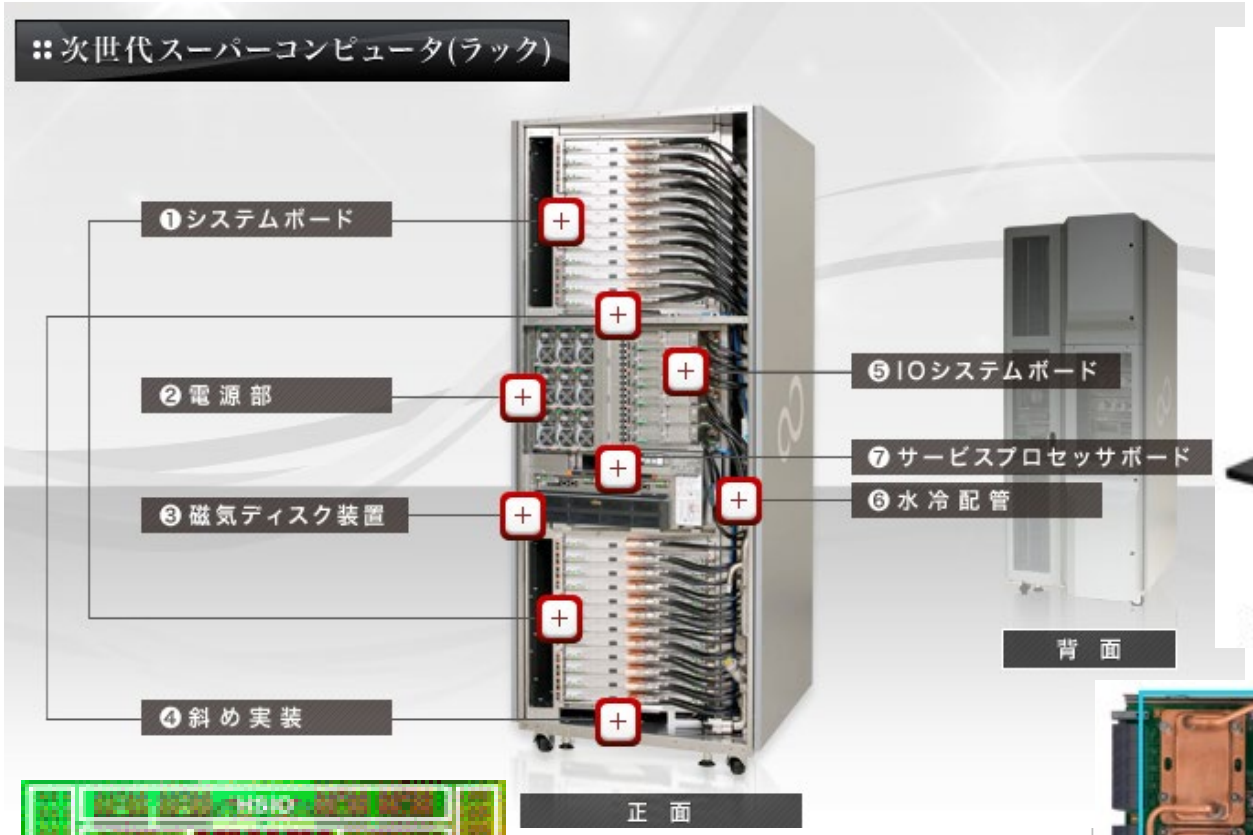
35.6 TFLOPS Linpack

June 2002 Top1

Cores: 5,120, Rmax:35.86TFlop/s

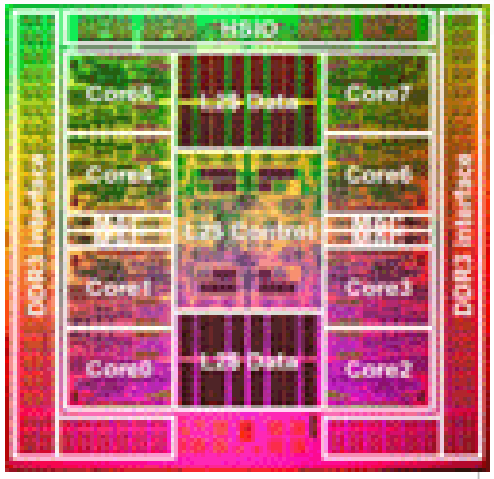
Rpeak: 40.96TFlop/s, Power: 3.2MW

“K” Supercomputer by Riken, No.1 in TOP500, June 20 & Nov.2,2011

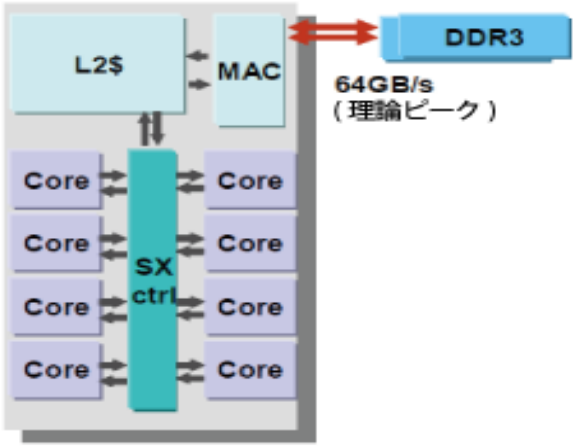


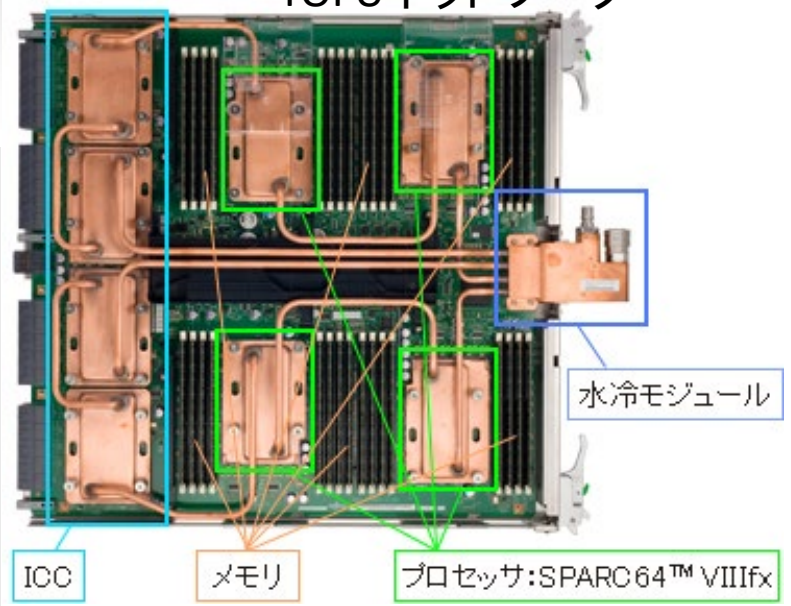
6次元メッシュ/トーラス(概念模型)

TOFUネットワーク



SPARC64™ VIIIfx
(提供:富士通(株))





IEEE Computer Society

The first President from outside North America in 72 years history of IEEE CS

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

847,000+
Articles in CSDL

6
New Standards

230
Active Standards

IEEE754, 802

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

Cornel Univ. Prof. Steven Squyres: Mars Exploration、CalTech. Dr. Katie Bouman: Visualization of Black Hole

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)

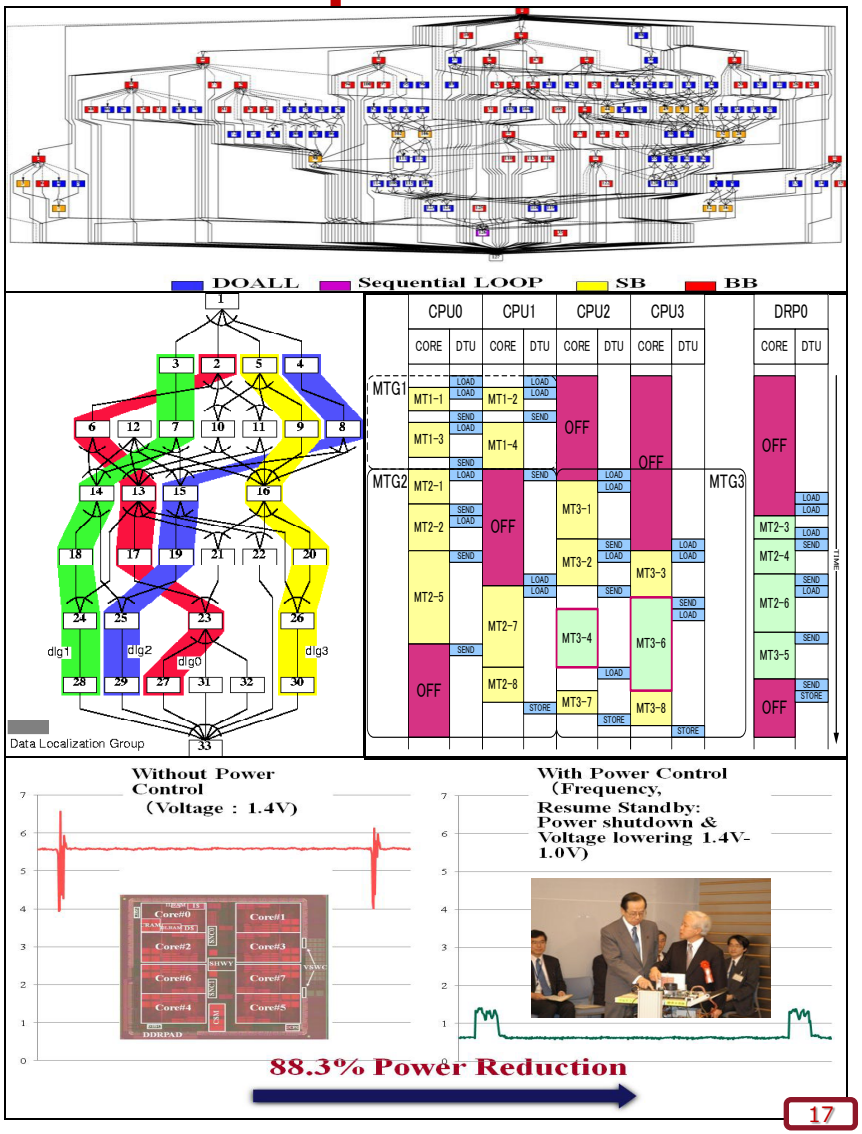
Data Transfer Overlapping(2016 partially)

Data transfer overlapping using Data Transfer Controllers (DMAs)

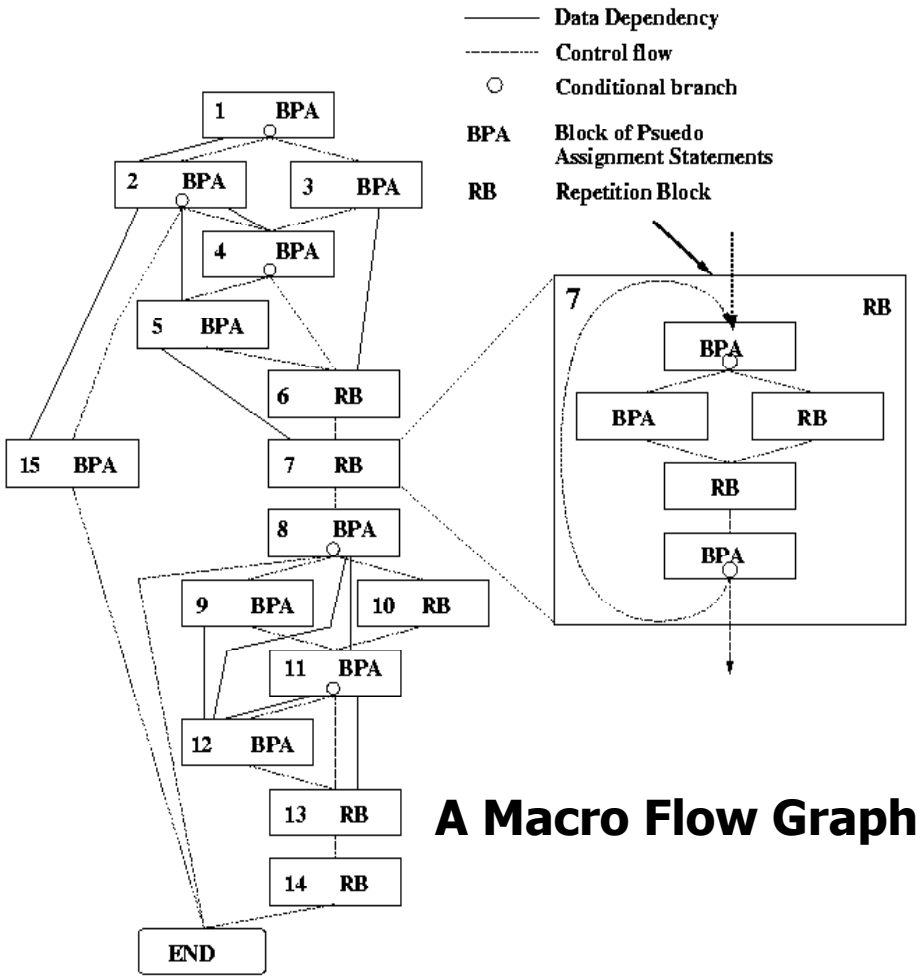
Power Reduction

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

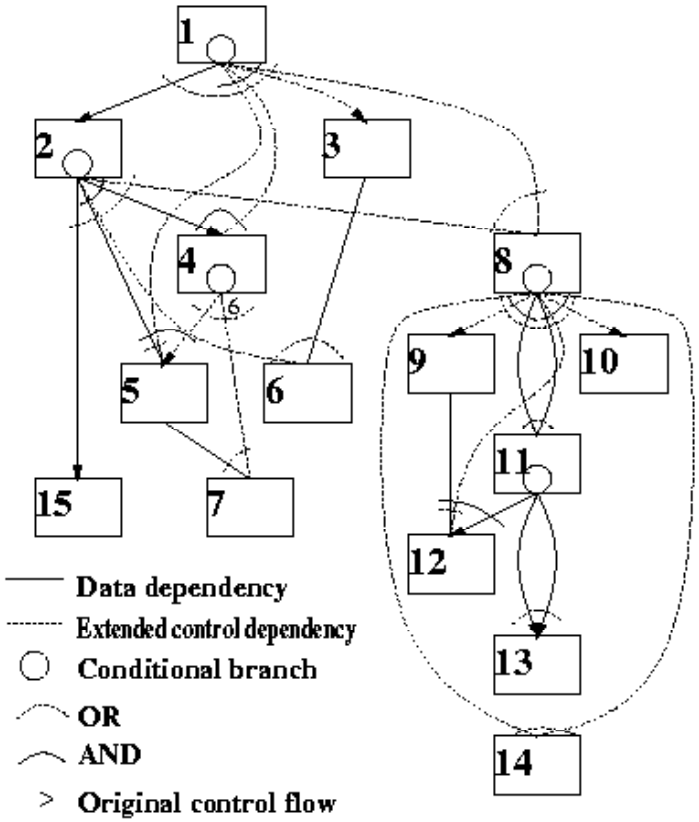
Reduction of consumed power by compiler control DVFS and Power gating with hardware supports.



Earliest Executable Condition Analysis for coarse grain tasks (Macro-tasks)



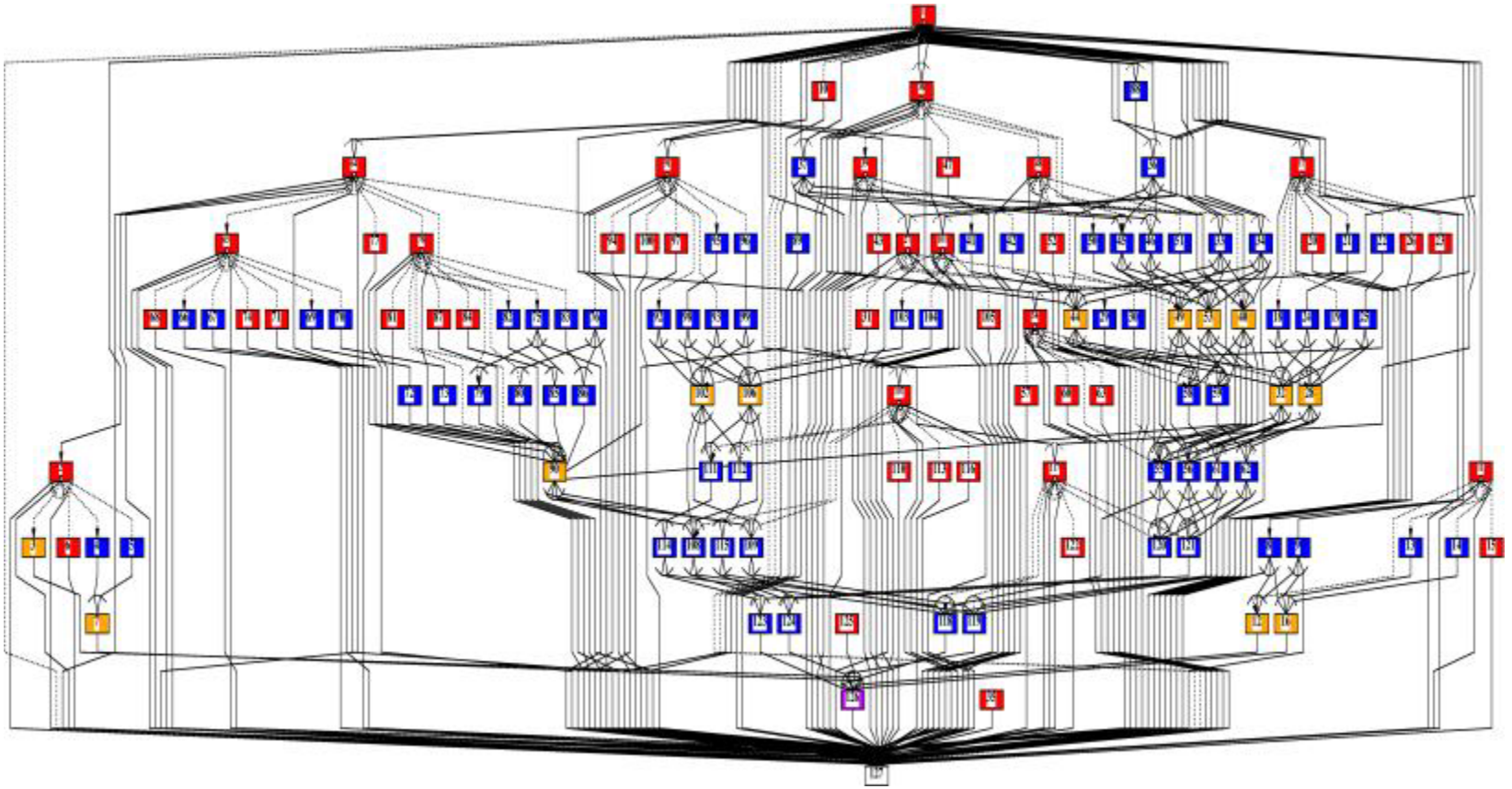
A Macro Flow Graph



A Macro Task Graph

MTG of Su2cor-LOOPS-DO400

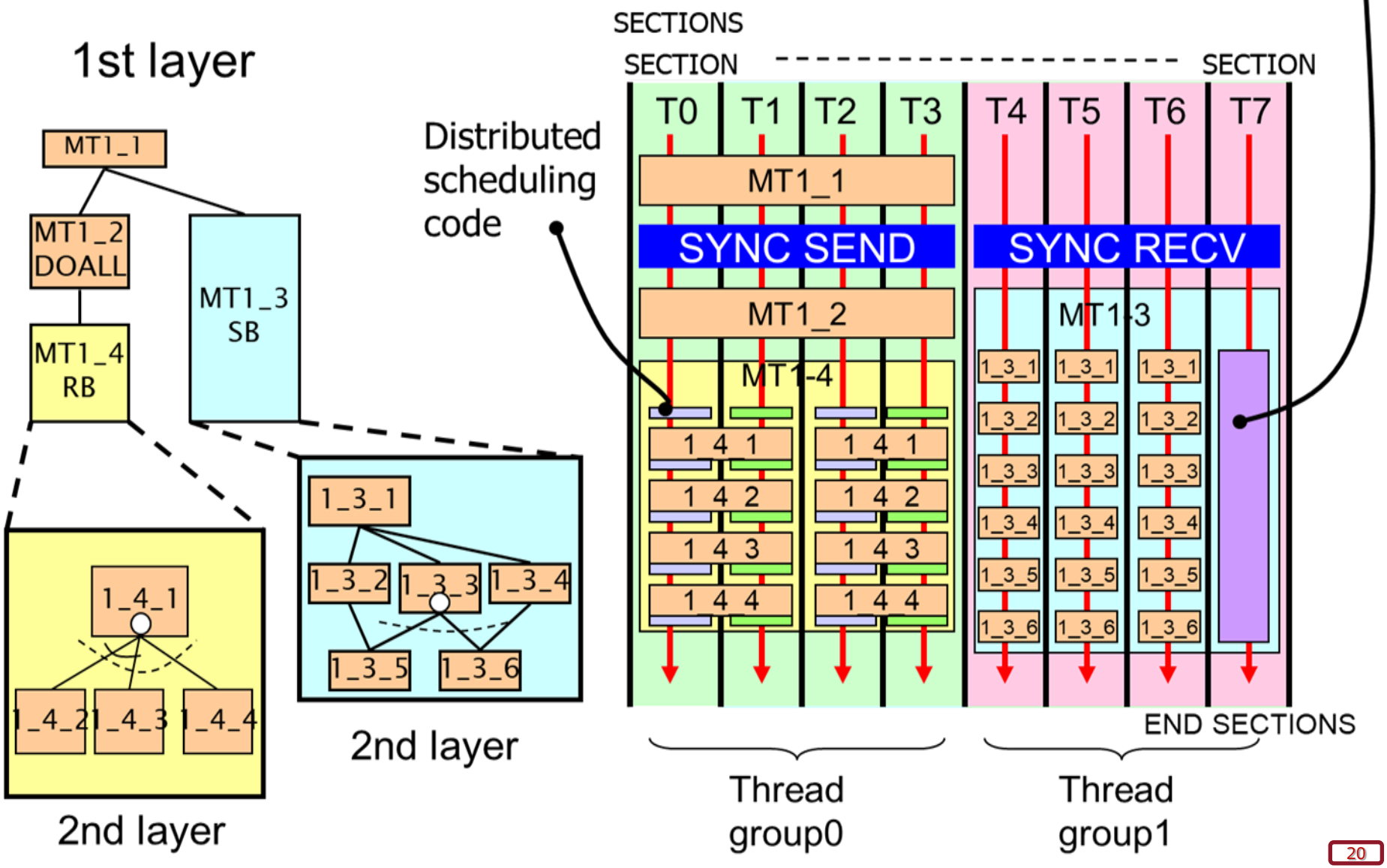
- Coarse grain parallelism PARA_ALD = 4.3



DOALL Sequential LOOP SB BB

Generated Multigrain Parallelized Code

(The nested coarse grain task parallelization is realized by only OpenMP "section", "Flush" and "Critical" directives.)



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

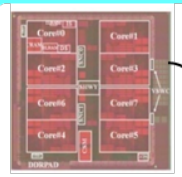
Accelerator 2
Code

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

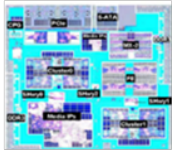
Low Power Heterogeneous Multicore Code Generation
API Analyzer (Available from Waseda) | Sequential compiler for CPU and Acc. or linking ACC Library

Server Code Generation
OpenMP Compiler

Generation of parallel machine codes using sequential compilers



Homegeneous Multicore s from Vendor A (SMP servers)



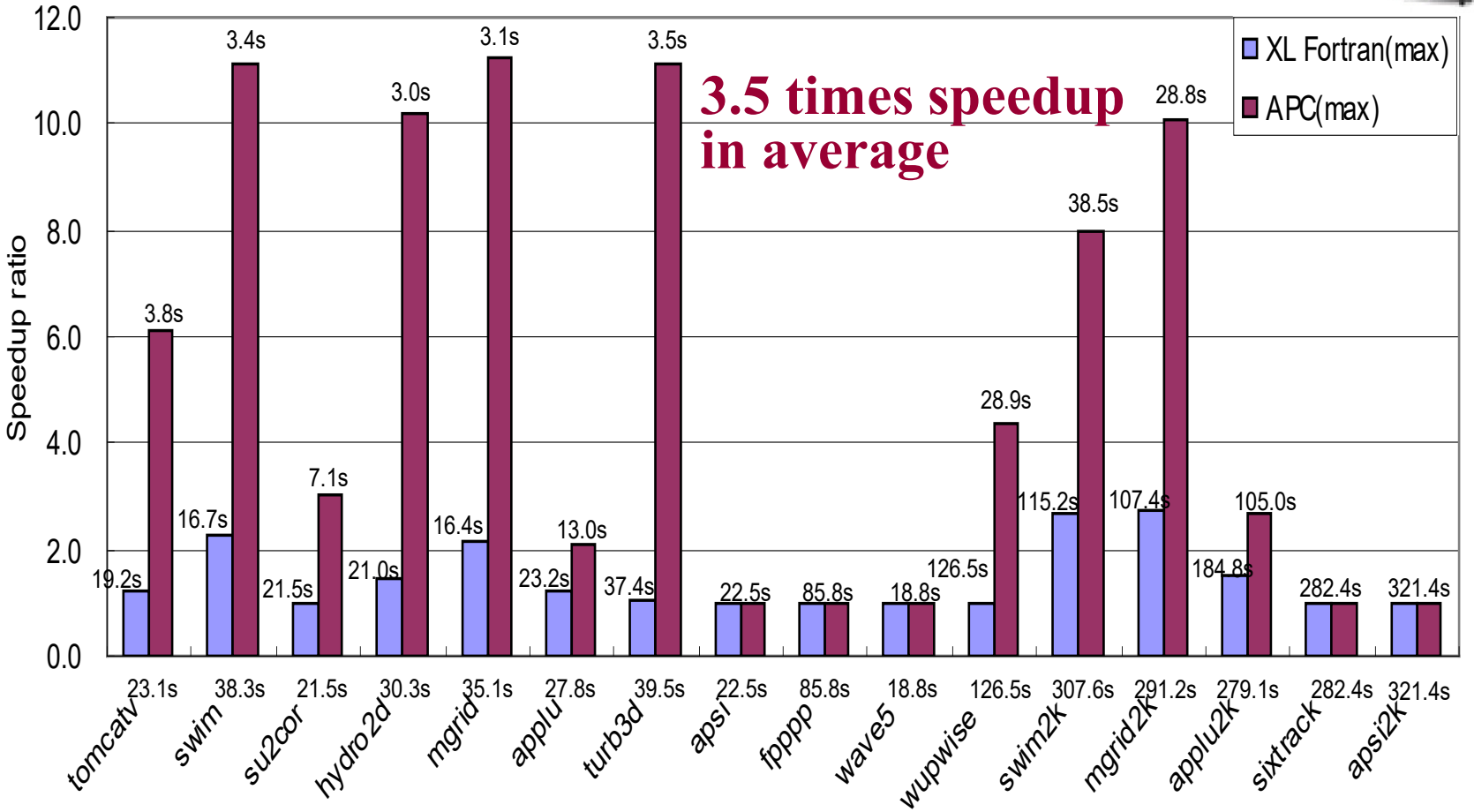
Heterogeneous Multicores from Vendor B

CC-NUMA servers

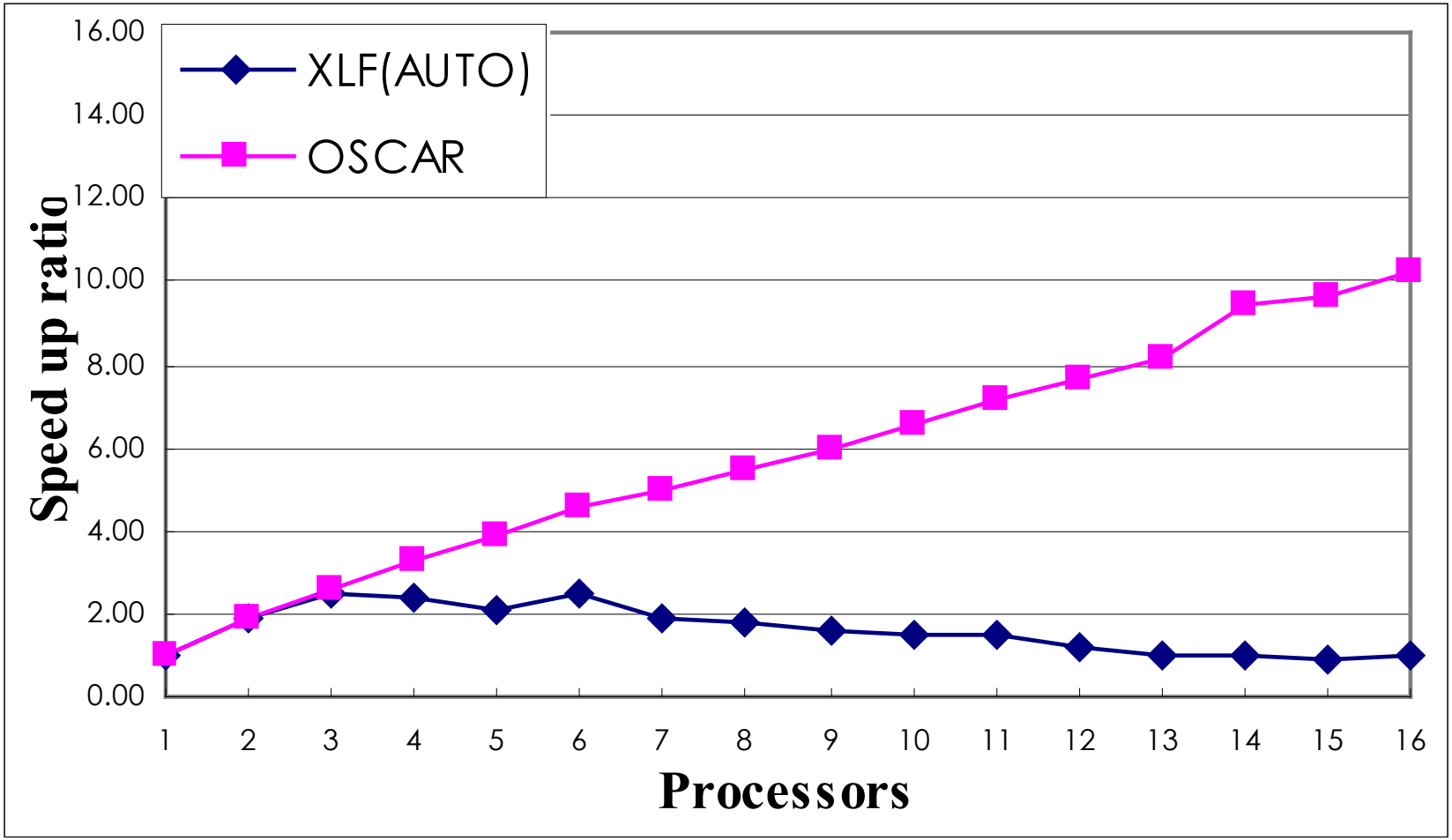
Executable on various multicores

Performance of APC Compiler on IBM pSeries690 16 Processors High-end Server

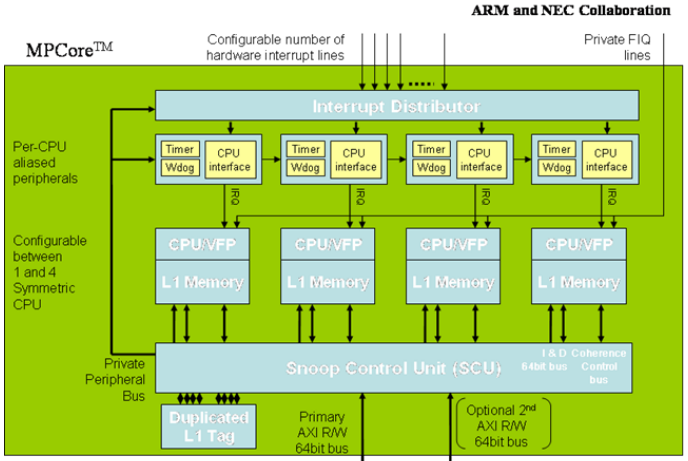
- IBM XL Fortran for AIX Version 8.1
  - Sequential execution : -O5 -qarch=pwr4
  - Automatic loop parallelization : -O5 -qsmp=auto -qarch=pwr4
  - OSCAR compiler : -O5 -qsmp=noauto -qarch=pwr4 (su2cor: -O4 -qstrict)

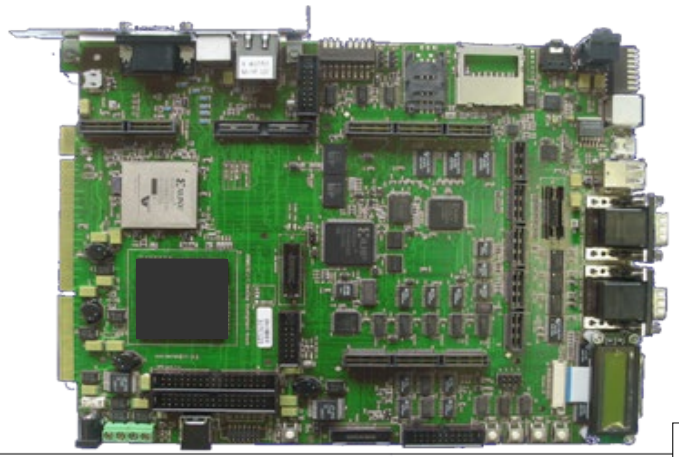


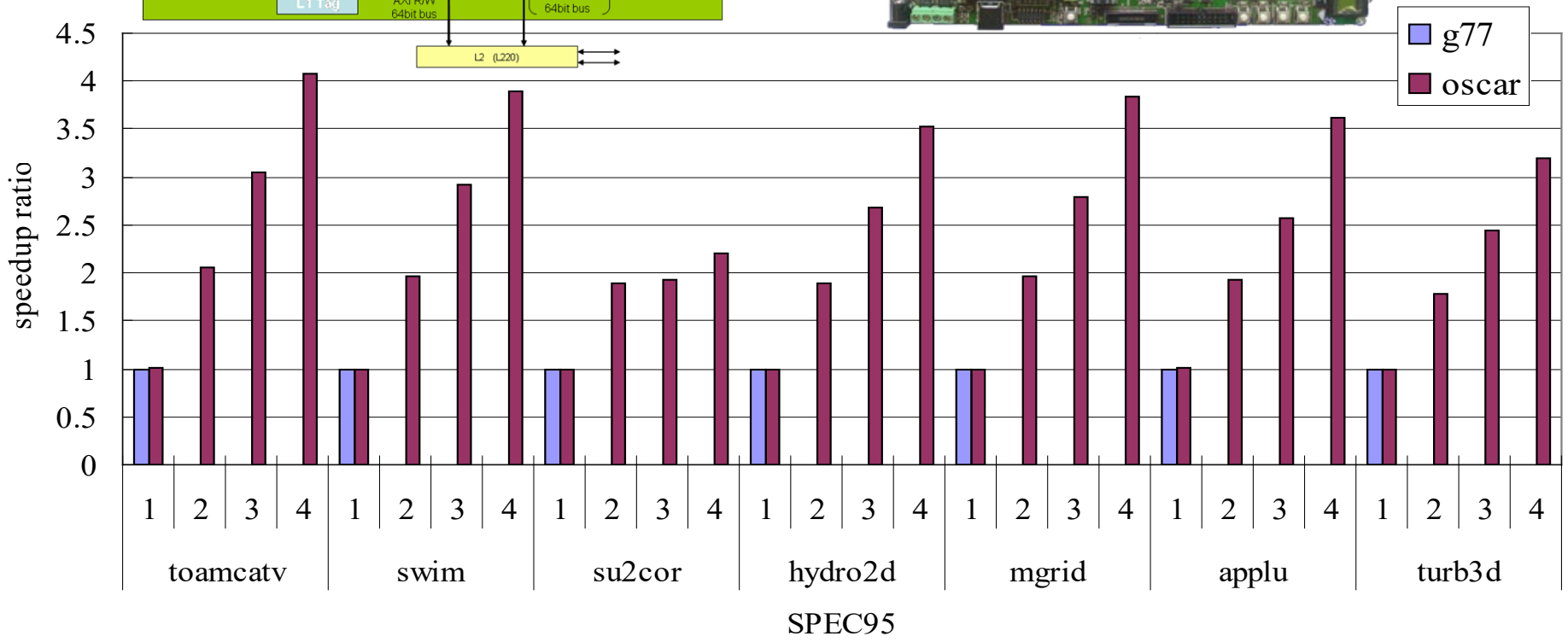
Performance of Multigrain Parallel Processing for 102.swim on IBM pSeries690



NEC/ARM MPCore Embedded 4 core SMP

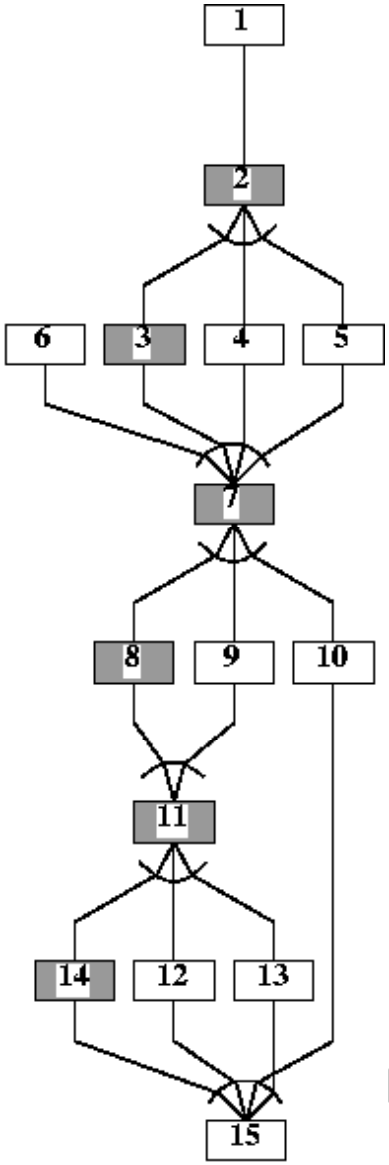




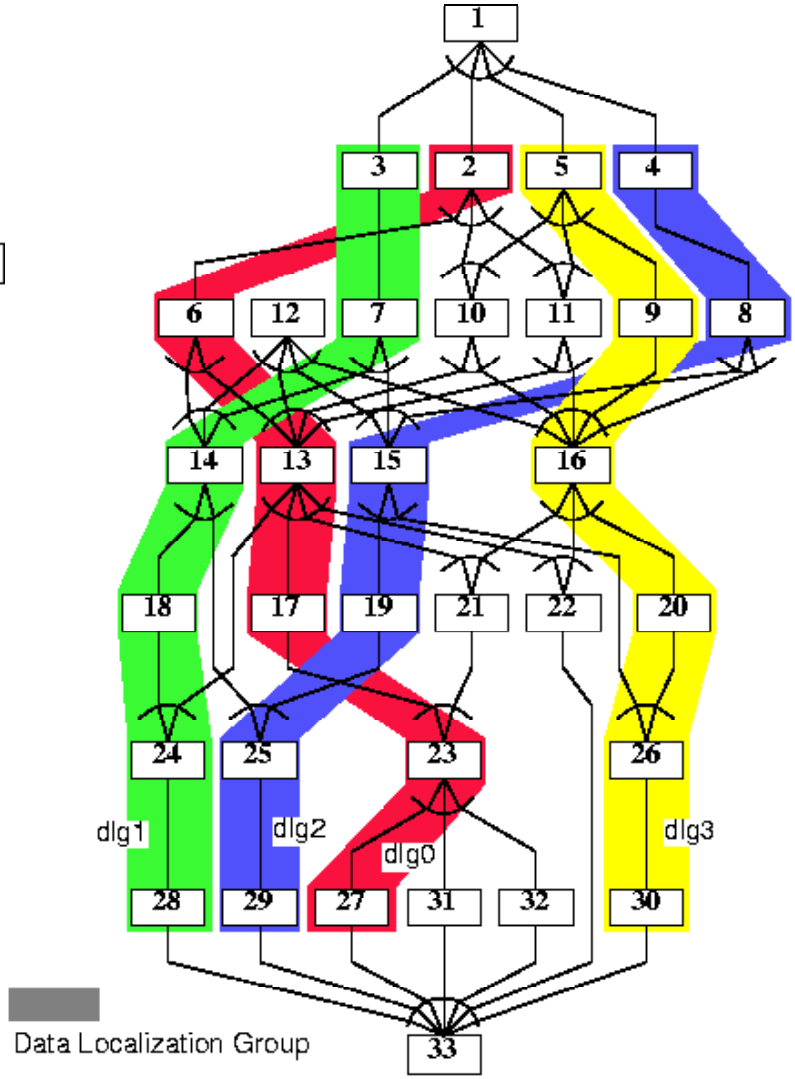


3.48 times speedup by OSCAR compiler against sequential processing

Data Localization



MTG



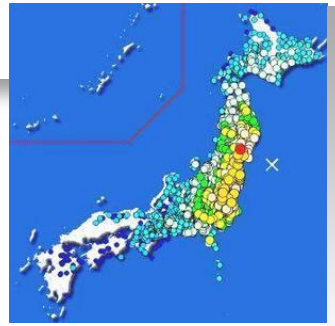
MTG after Division

A schedule for two processors

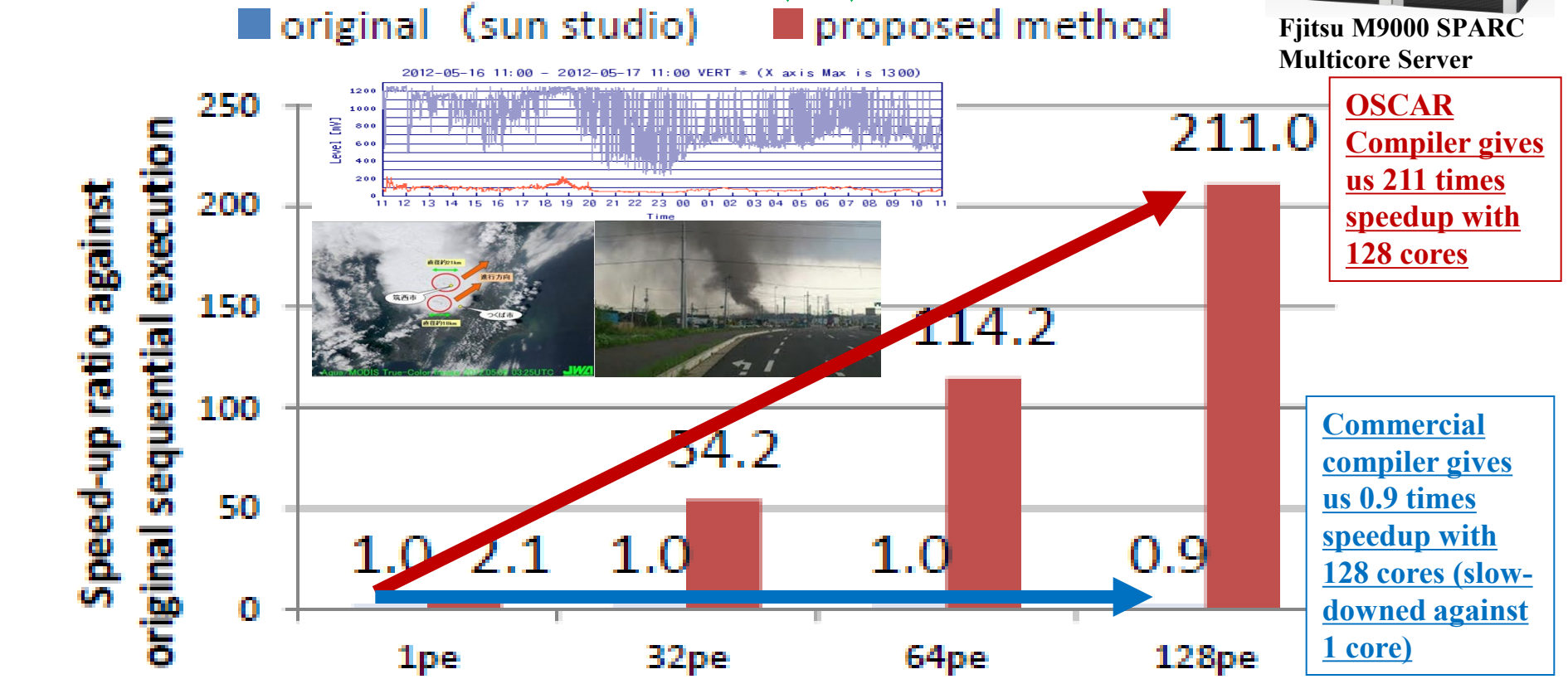
地震波伝搬シミュレーションの高速化：富士通スパコン128プロセッサコア上で、従来1コアより211倍の高速化に成功

- 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

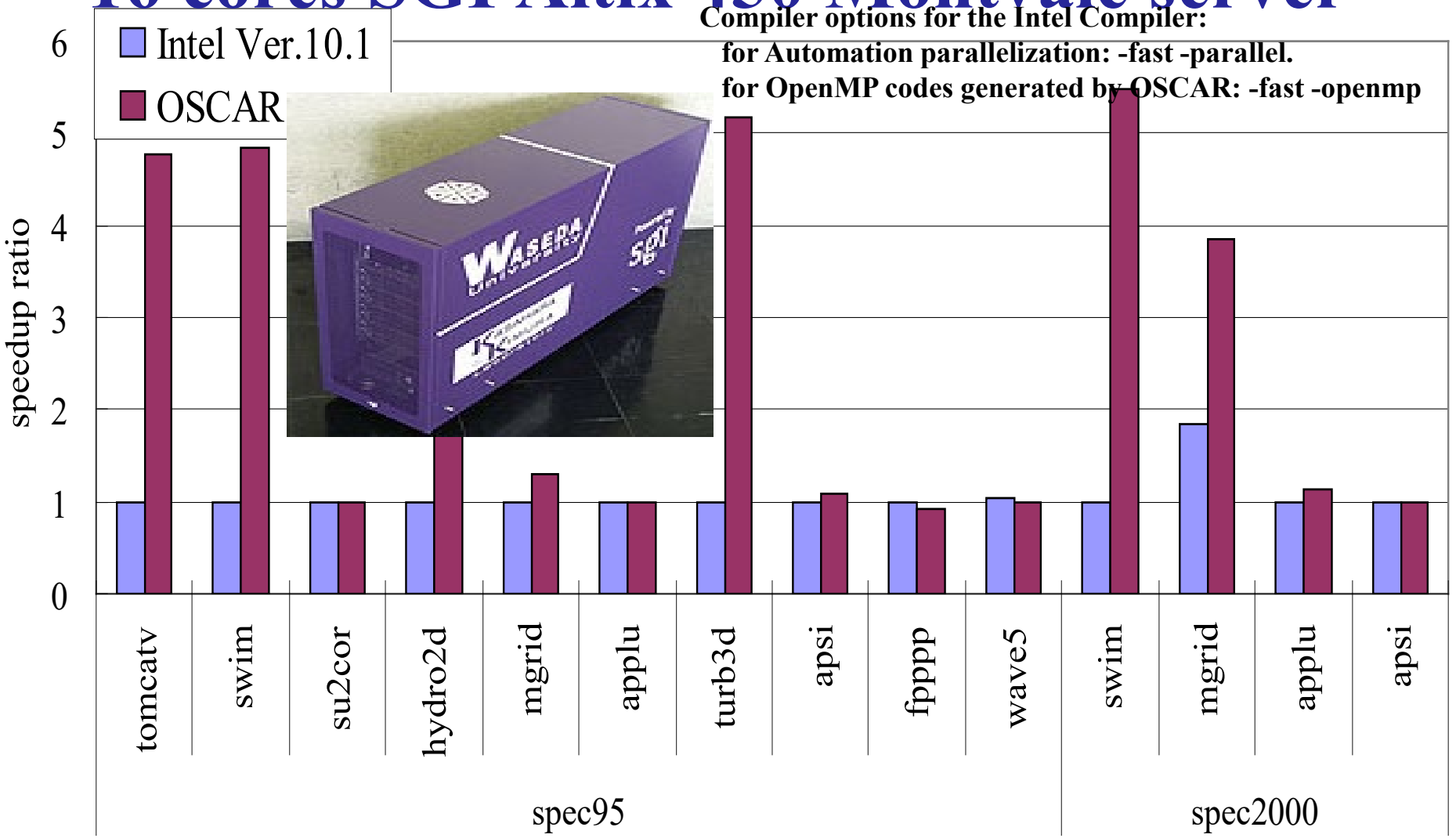


OSCAR Compiler gives us 211 times speedup with 128 cores

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

Performance of OSCAR compiler on 16 cores SGI Altix 450 Montvale server



- **OSCAR compiler gave us 2.32 times speedup against Intel Fortran Itanium Compiler revision 10.1**

スーパーテクニカルサーバ「SR16000モデルVM1」

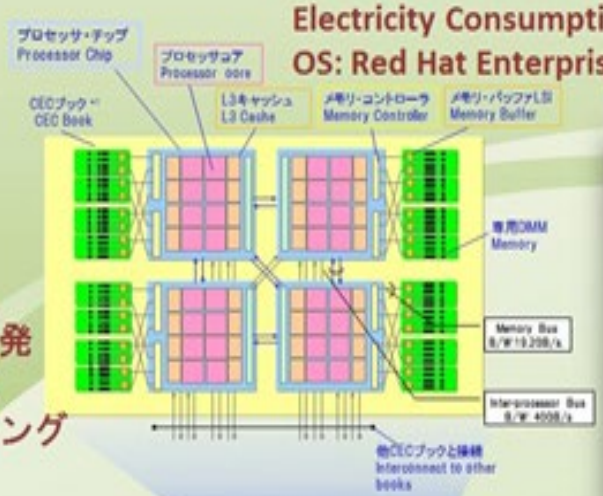
Kasahara & Kimura Laboratory

Green Computing
Research Center Project

CPU : POWER7（4.0GHz）× 128Core
Memory : 1024GB
Peak Perfromance:4.1TFLOPS
Electricity Consumption : 256.8MFLOPS/W
OS: Red Hat Enterprise Linux 6

【用途】
・並列化コンパイラ開発
・低消費電力メニーコア開発
・屋上太陽光発電を用いた
グリーンコンピューティング
・災害シミュレーション

【Purpose】
・Parallelizing compiler development
・Development of low power
manycore processors
・Green Computing
by rooftop solar power
・Disaster Simulation

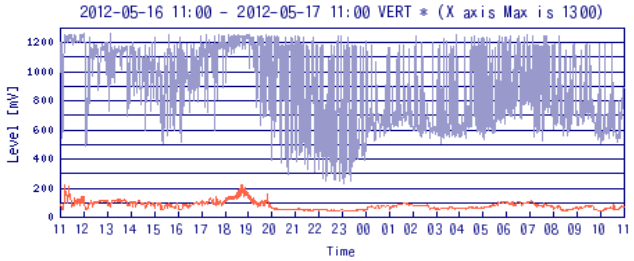


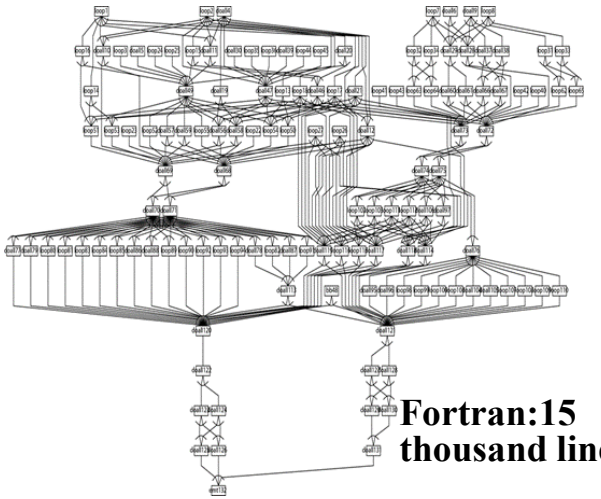
HITACHI

WASEDA Univ.

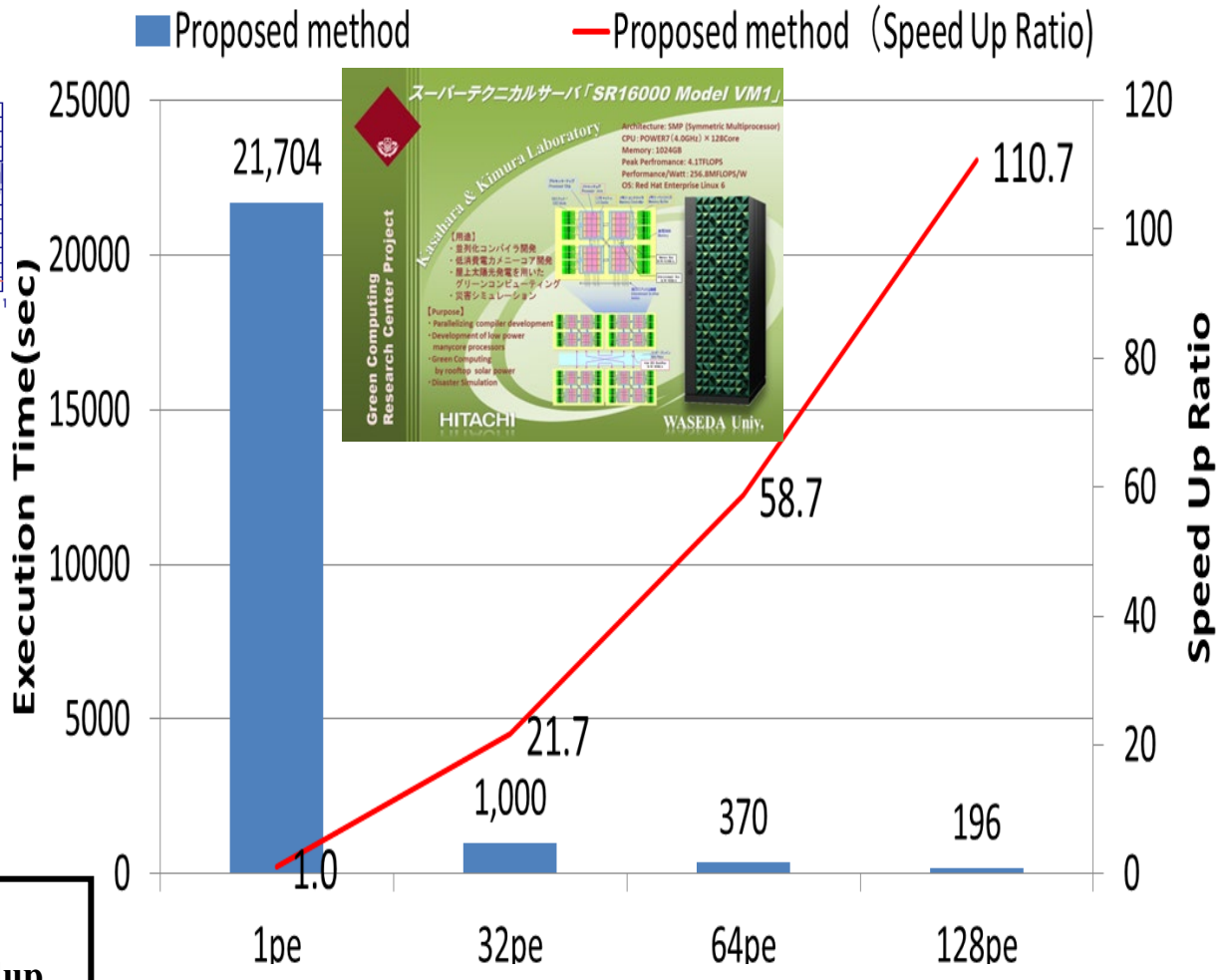
110 Times Speedup against the Sequential Processing for GMS Earthquake Wave Propagation Simulation on Hitachi SR16000

（Power7 Based 128 Core Linux SMP） (LCPC2015)





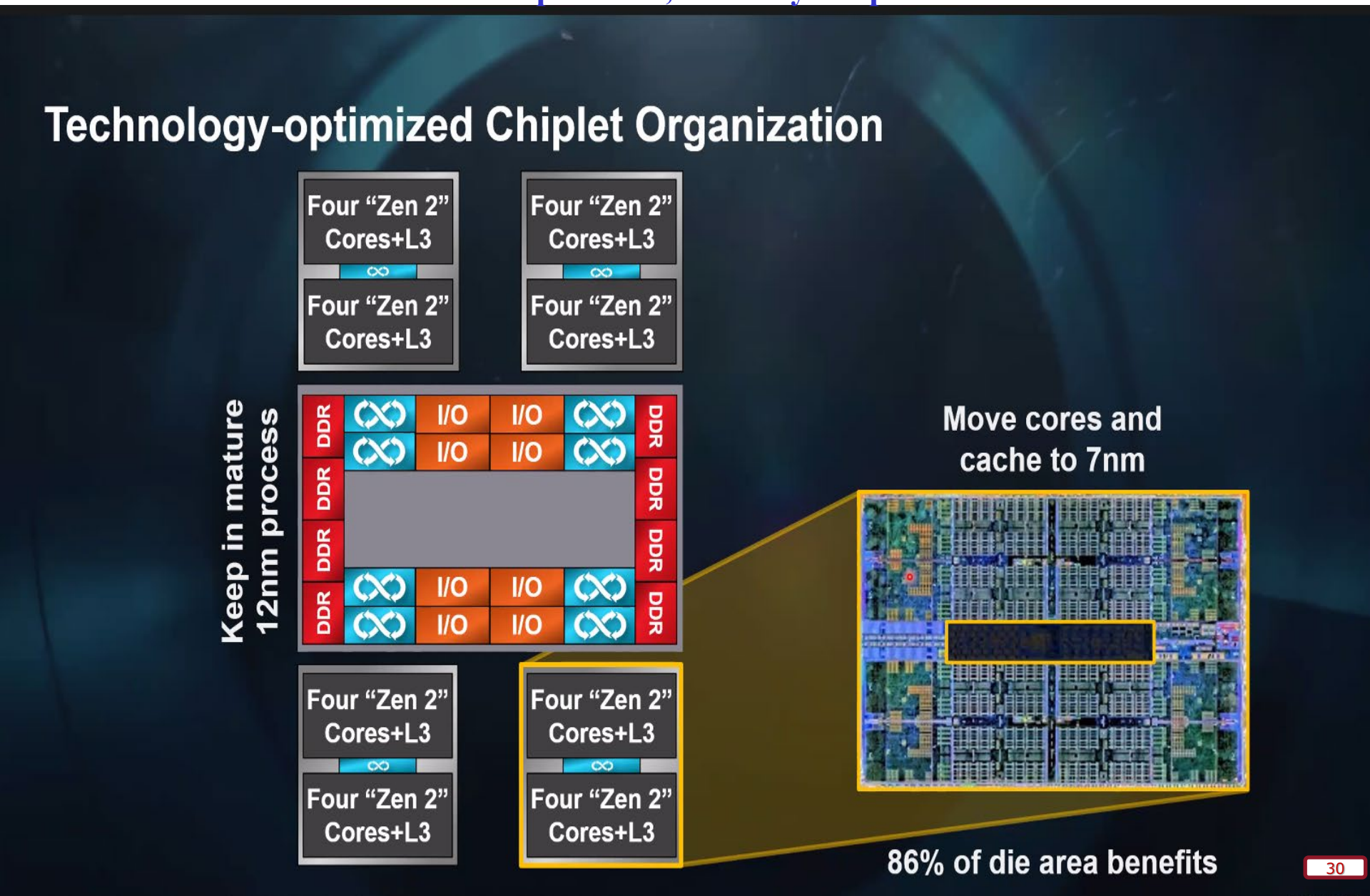
Fortran:15 thousand lines



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

AMD Chiplet for 64 cores: Zen2

Processor Chiplet 7nm, Memory Chiplet: 12nm

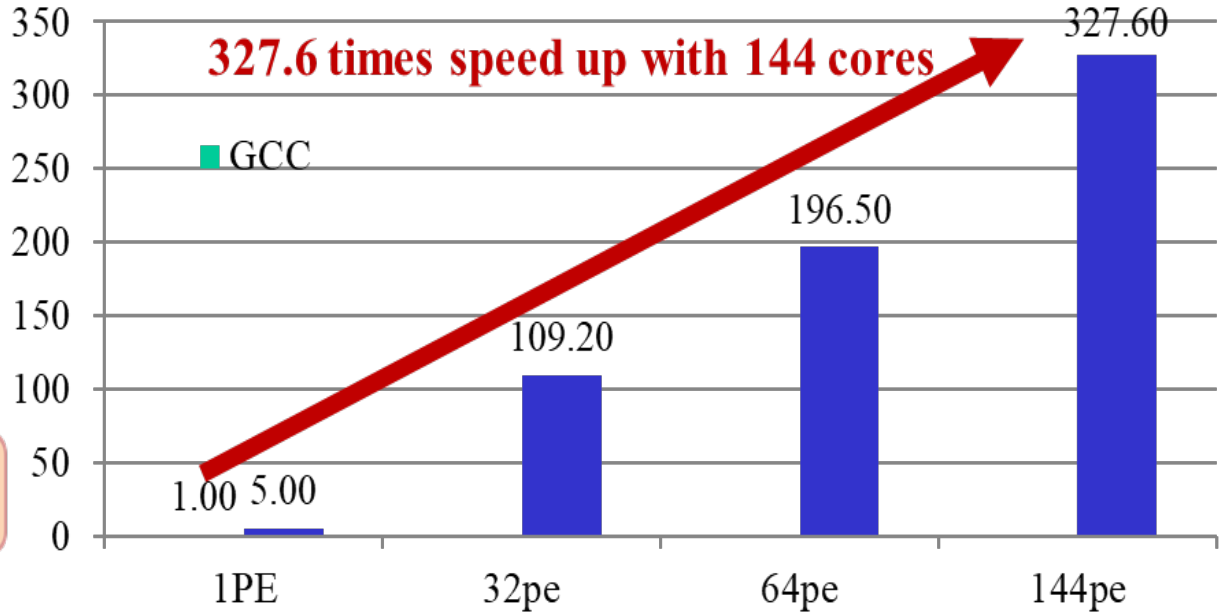


AMD

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

327 times speedup on 144 cores

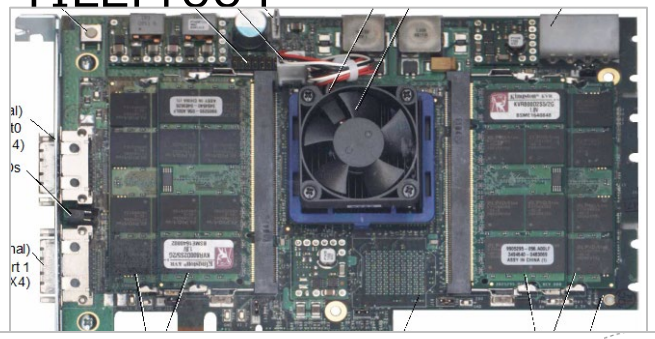
Hitachi 144cores SMP Blade Server BS500:
Xeon E7-8890 V3(2.5GHz　18core/chip) x8 chip

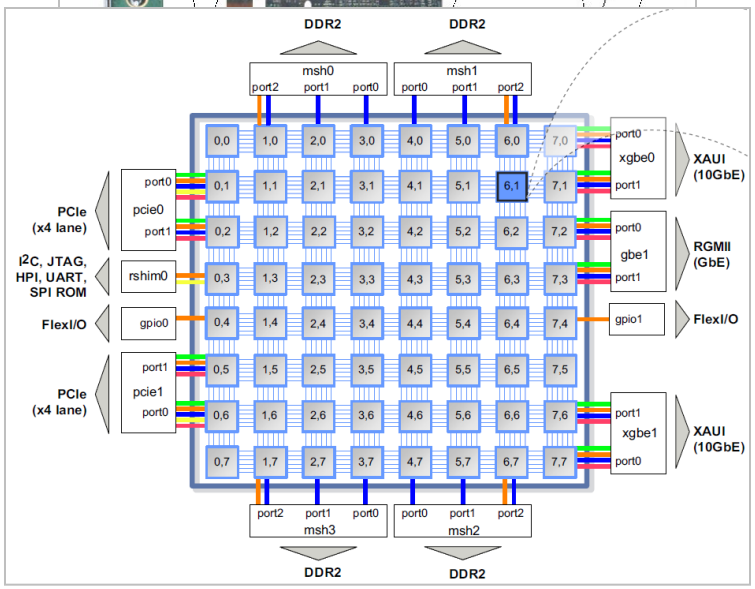


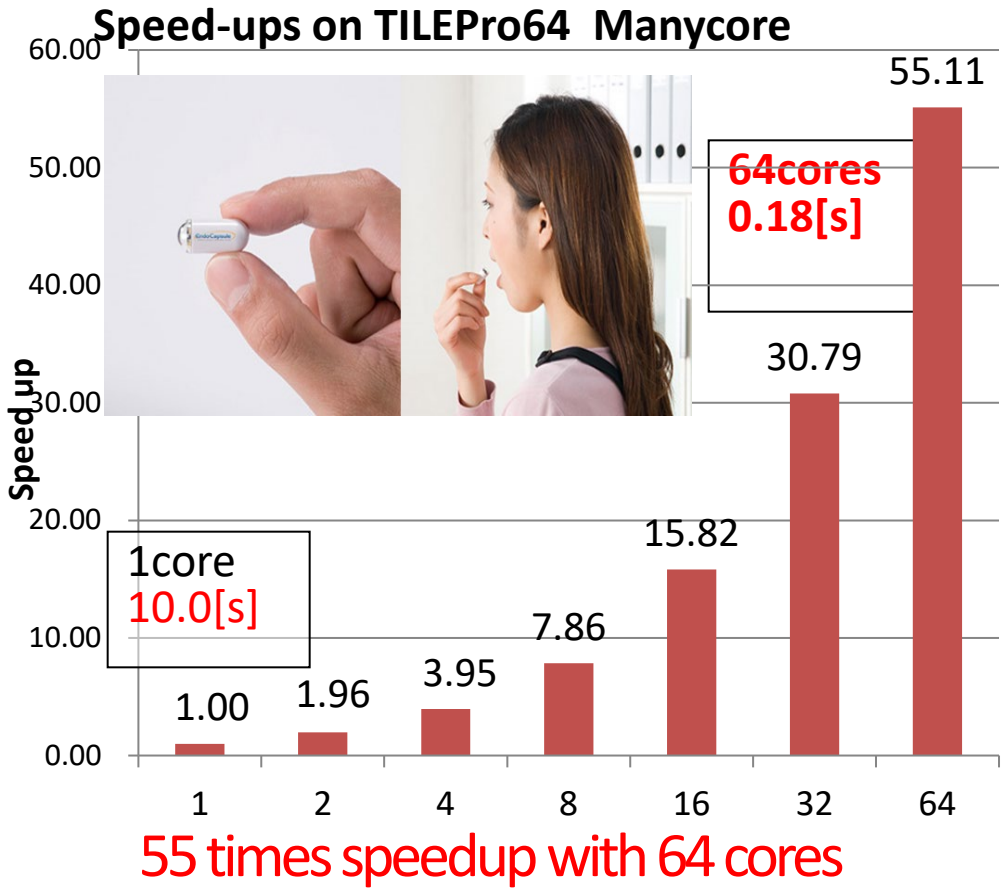
- Original sequential execution time 2948 sec（50 minutes）using GCC was reduced to 9 sec with 144 cores（327.6 times speedup）
  - Reduction of treatment cost and reservation waiting period is expected

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

- TILEPro64







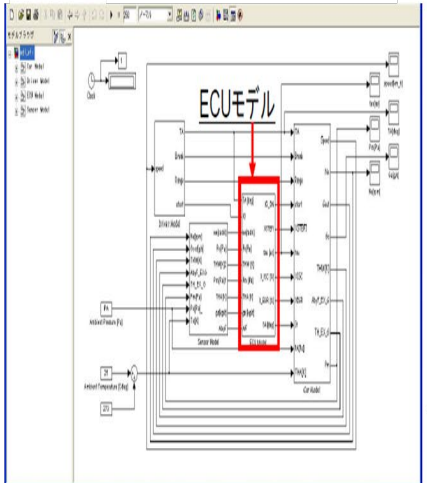
日本乗用車のエンジン制御計算をデンソー2コアECU上で、1．95倍の速度向上に成功。(見神、梅田）

欧州農耕作業車エンジン制御計算をインフィニオン2コアプロセッサ上で8.7倍の高速化に成功。

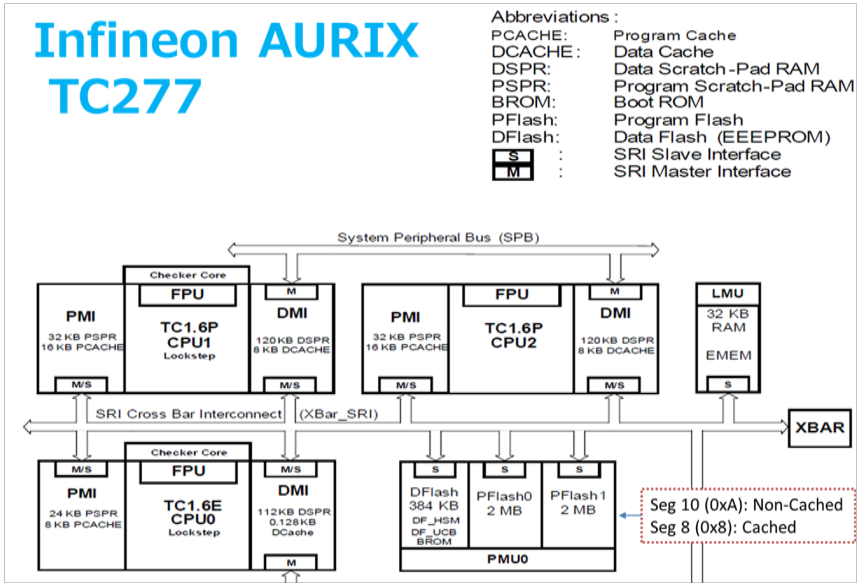
Engine Control by multicore with Denso

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



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

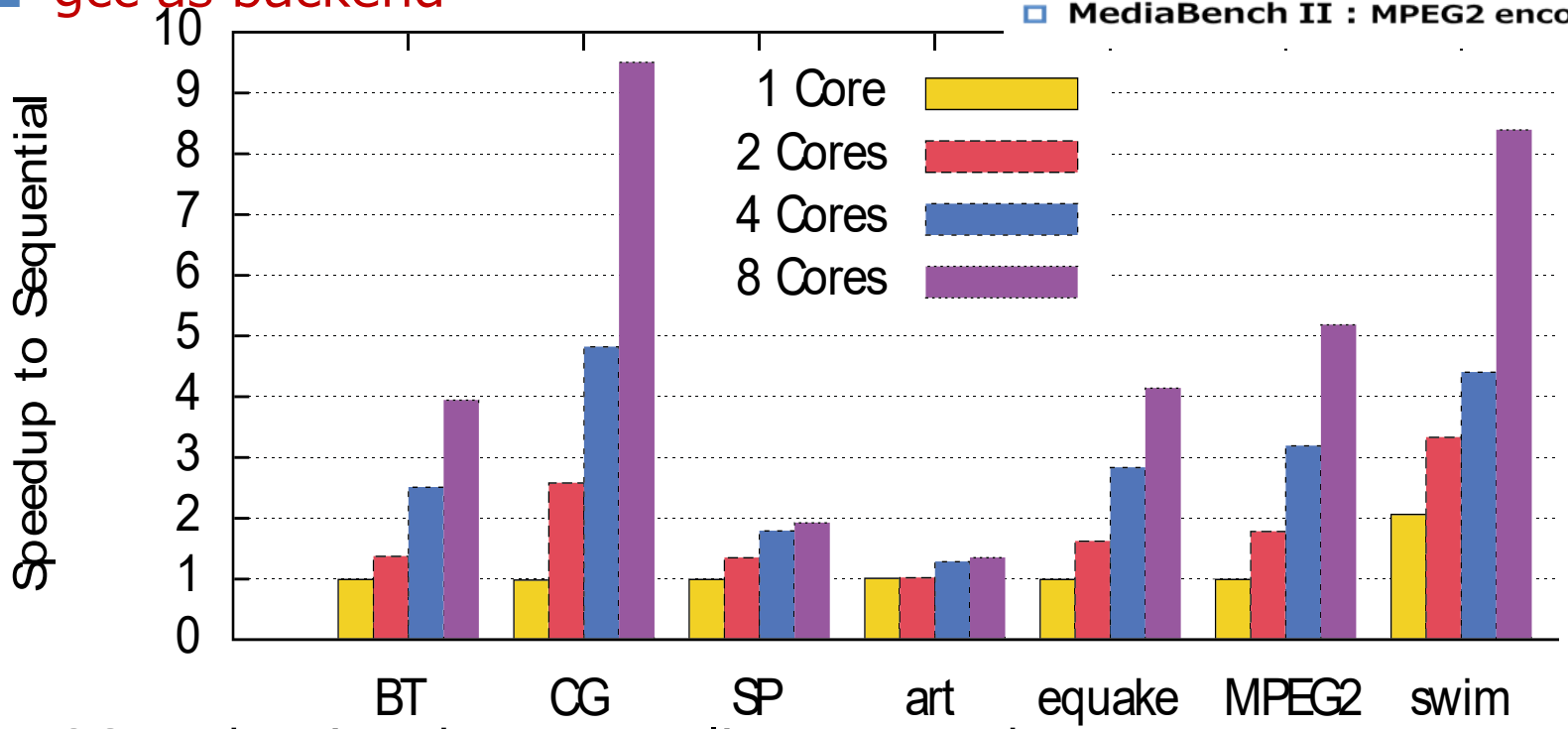


AMD EPYC 7702P – Benchmark results on upto 8 cores

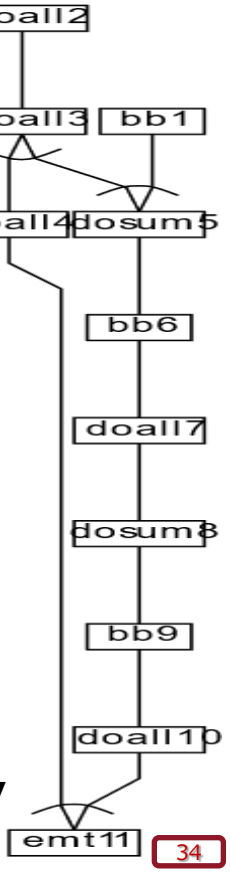
- x86-64 based Architecture, 64 Cores, 2.0 GHz – 3.35 GHz
- 16 MiB L3 cache per 4 core cluster,
- shared within the cluster

- NAS parallel benchmark suite
  - BT: Block Tri-diagonal solver
  - CG: Conjugate Gradient computation
  - SP: Scalar Penta-diagonal solver
- SPEC2000
  - art: Image recognition / Neural networks
  - equake: Seismic wave propagation simulation
  - swim: Shallow water modelling - Fortran 77
- MediaBench II : MPEG2 encoding

- speedup to sequential version
- gcc as backend



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

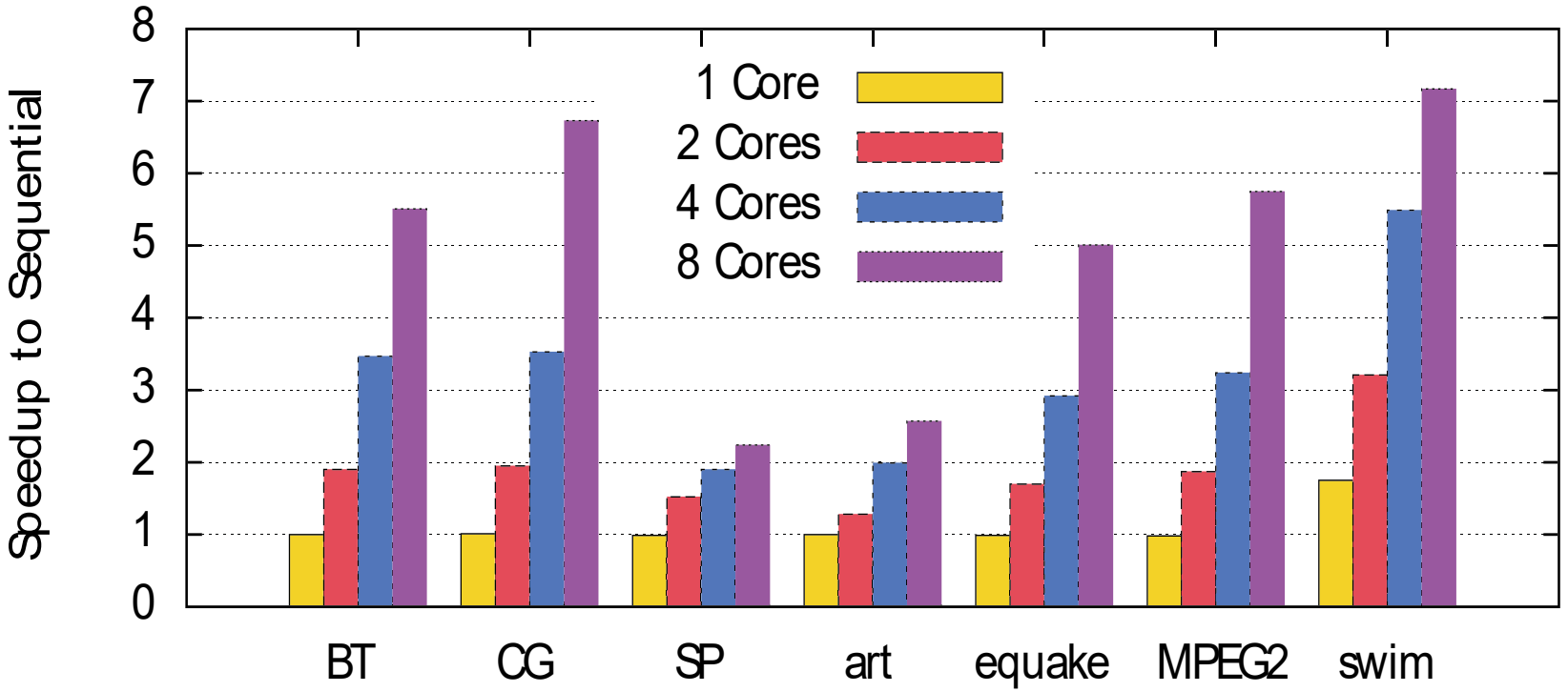


CG macrotask graph
L3 cache Optimization by Data Localization

Intel Xeon E5-2650v4 – Benchmark results on upto 8 cores

- x86-64 based Architecture, 12 Cores, 2.2 GHz – 2.9 GHz
- 30 MiB shared L3 cache,
- L3 Cache: Shared by all cores
- speedup to sequential version
- gcc as backend

- **NAS parallel benchmark suite**
  - **BT: Block Tri-diagonal solver**
  - **CG: Conjugate Gradient computation**
  - **SP: Scalar Penta-diagonal solver**
- **SPEC2000**
  - **art: Image recognition / Neural networks**
  - **equake: Seismic wave propagation simulation**
  - **swim: Shallow water modelling - Fortran 77**
- **MediaBench II : MPEG2 encoding**

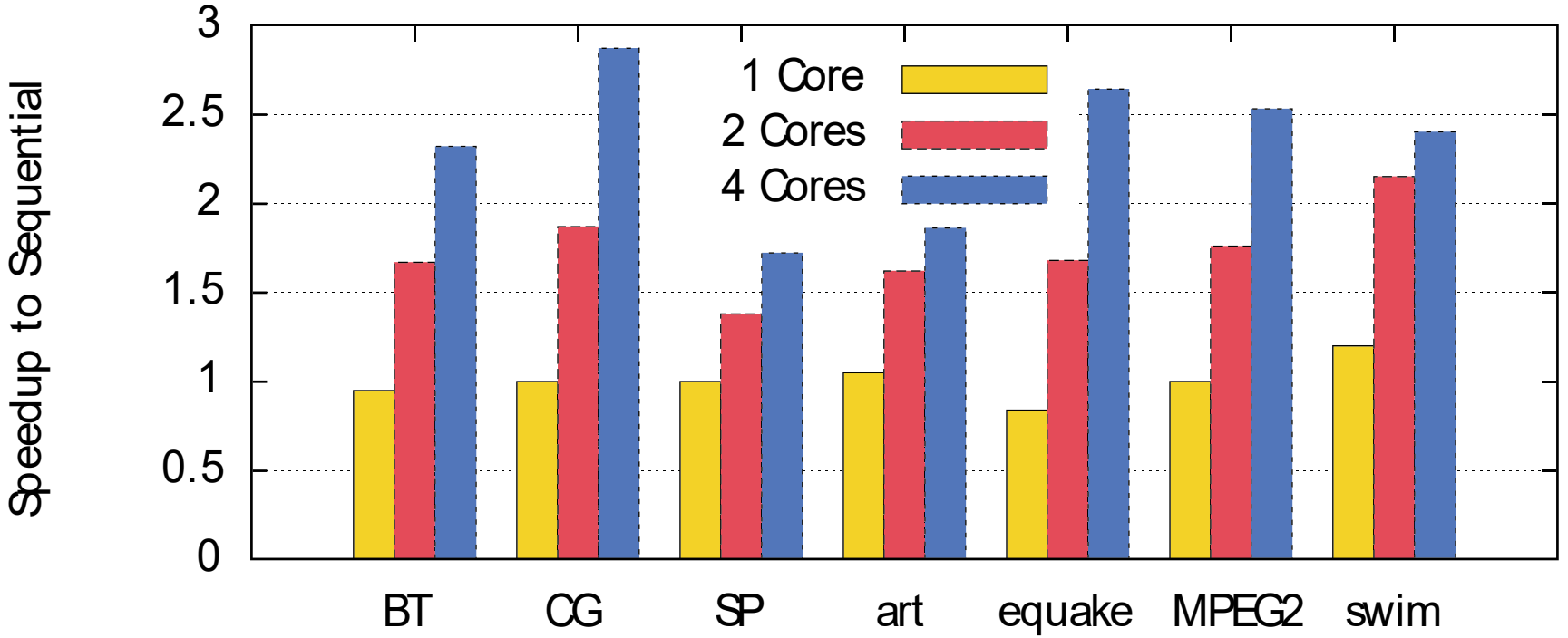


- swim shows superlinear speedup and 1 core speedup
  - seq.: 58.1 s, 1 core OSCAR: 33.2 s, 4 core OSCAR: 10.5 s

NVIDIA Carmel ARMv8.2 – Benchmark results on upto 4 cores

- Arm v8.2 based Architecture, 6 Cores, 1.4 GHz
- 4 MiB shared L3 cache,
- L3 Cache: Shared across all cores
- speedup to sequential versior
- gcc as backend

- NAS parallel benchmark suite
  - BT: Block Tri-diagonal solver
  - CG: Conjugate Gradient computation
  - SP: Scalar Penta-diagonal solver
- SPEC2000
  - art: Image recognition / Neural networks
  - equake: Seismic wave propagation simulation
  - swim: Shallow water modelling - Fortran 77
- MediaBench II : MPEG2 encoding



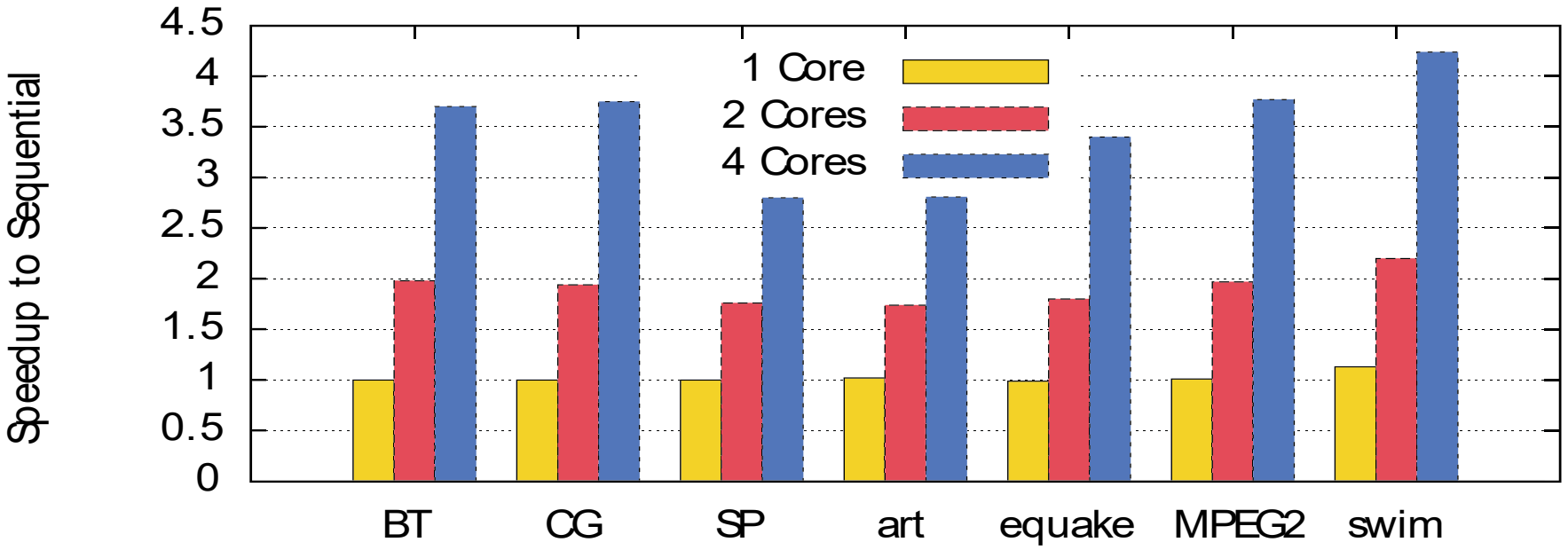
- overall good speedup is observed
  - equake: seq.: 19.0 s, 4 core OSCAR: 7.18 s

SiFive Freedom U740 – Benchmark results on upto 4 cores

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

- NAS parallel benchmark suite
  - BT: Block Tri-diagonal solver
  - CG: Conjugate Gradient computation
  - SP: Scalar Penta-diagonal solver
- SPEC2000
  - art: Image recognition / Neural networks
  - equake: Seismic wave propagation simulation
  - swim: Shallow water modelling - Fortran 77
- MediaBench II : MPEG2 encoding

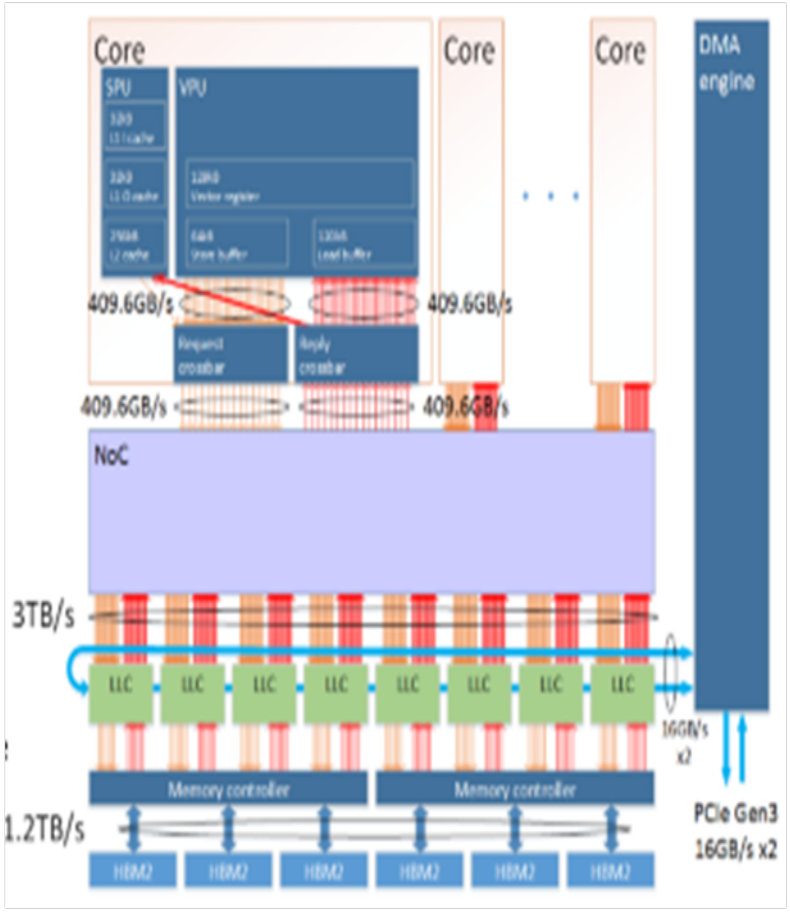
- speedup to sequential version
- gcc as backend

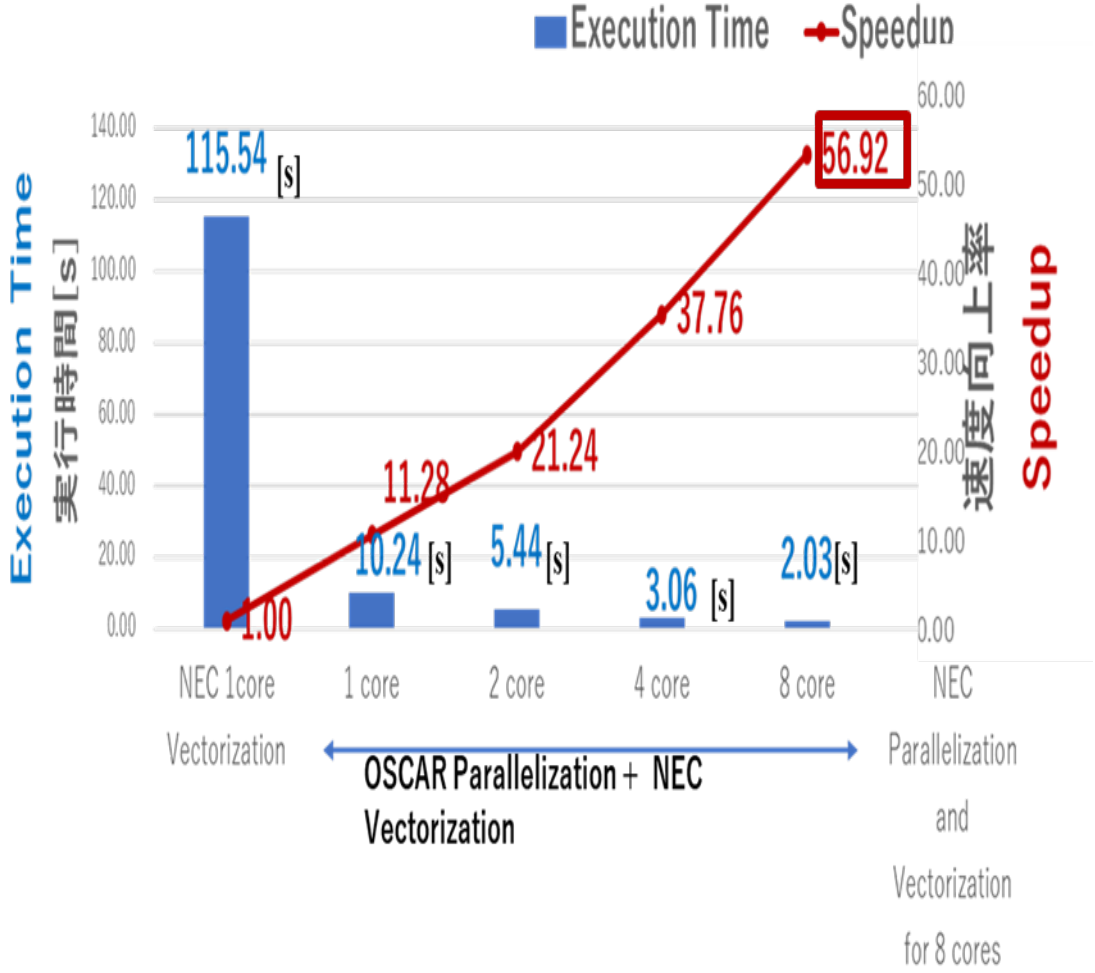


- overall good speedup is observed, swim superlinear
  - BT: seq.: 2041 s, 4 core OSCAR: 551 s

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



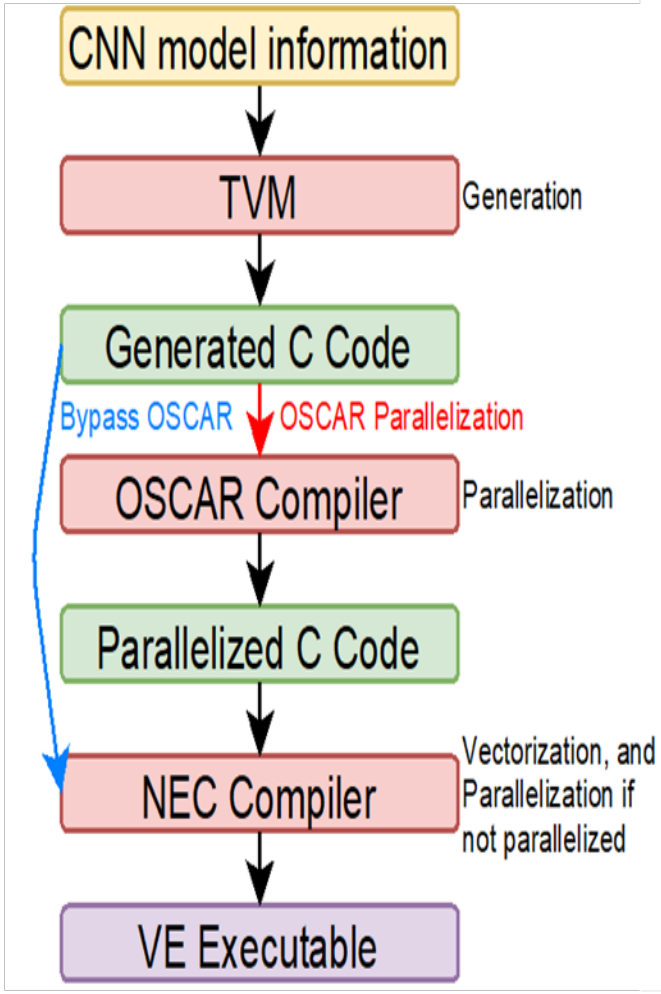


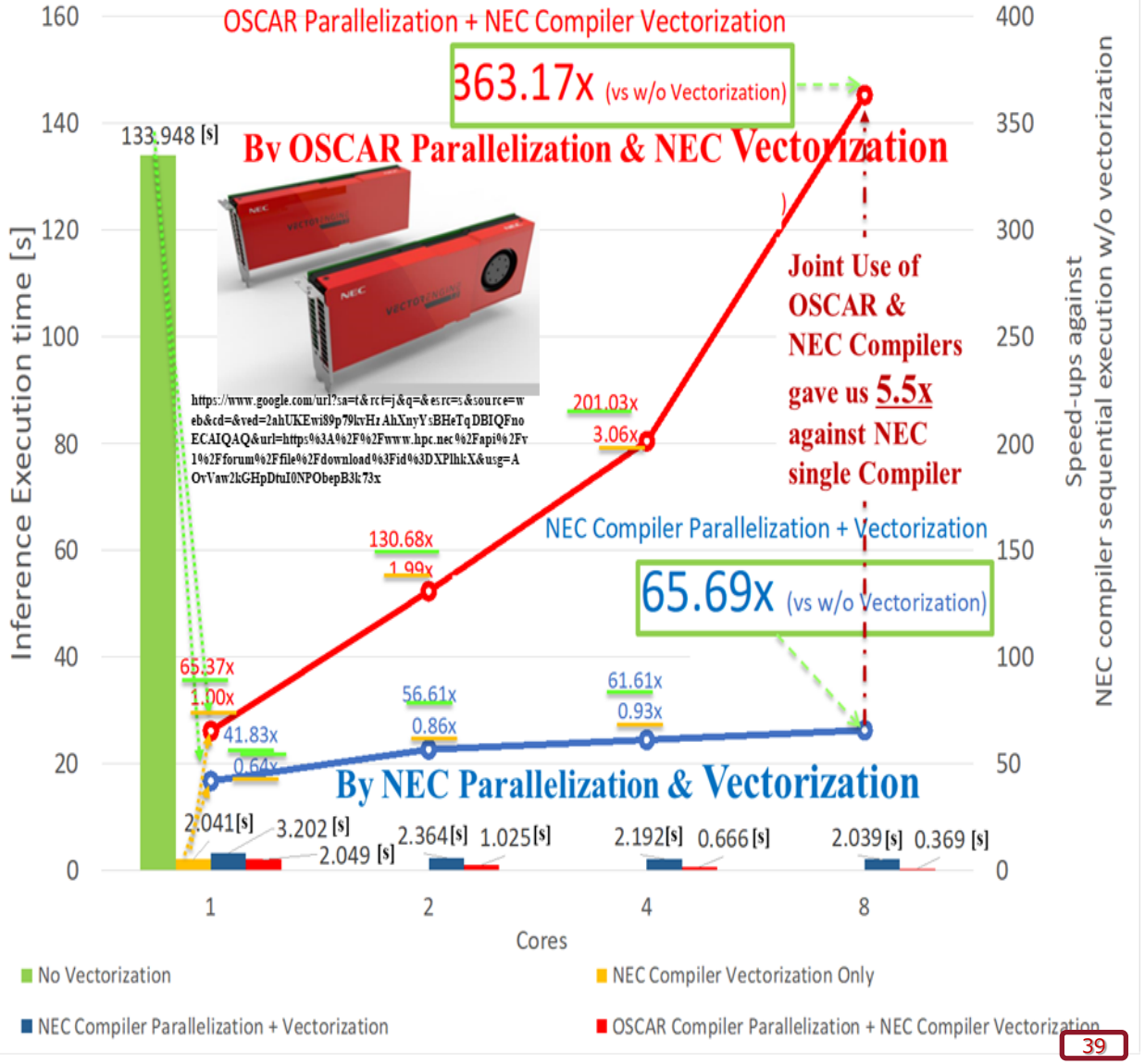
NEC Corporation. (2021) SX-Aurora TSUBASA Performance Tuning Guide. [Online]. Available: https://www.hpc.nec/documents/guide/pdfs/AuroraVE TuningGuide.pdf

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

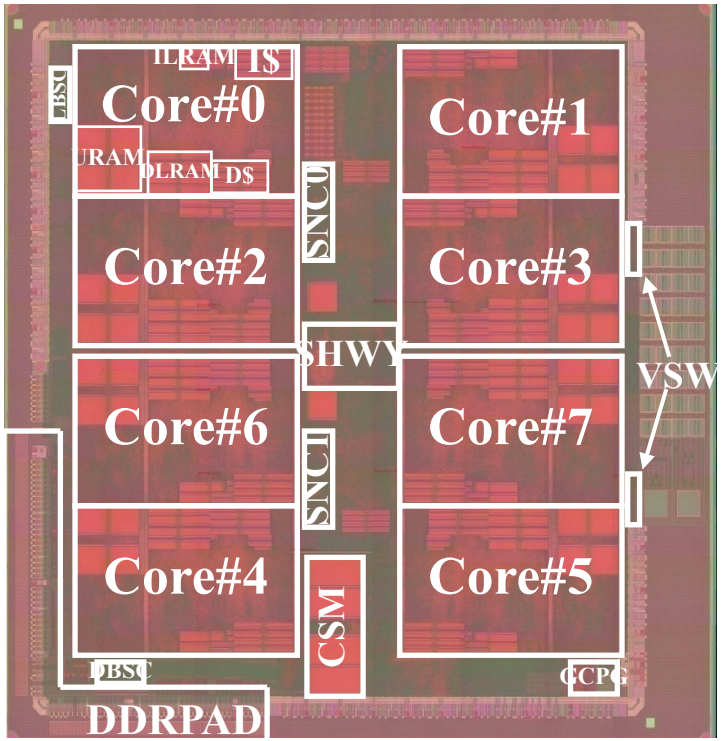




ムーアの法則の終焉 (End of Moore's Law)

ムーアの法則(Moore's law): インテル創業者の一人であるゴードン・ムーア(Gordon E. Moore: IEEE Computer Pioneer Award)が、1965年の論文で提唱した経験則:
"半導体の集積率は18か月で2倍になる" "Transistors on a chip 2× every 1.5 year"

コンピュータの高性能化と低消費電力化にはマルチコアが必須



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)

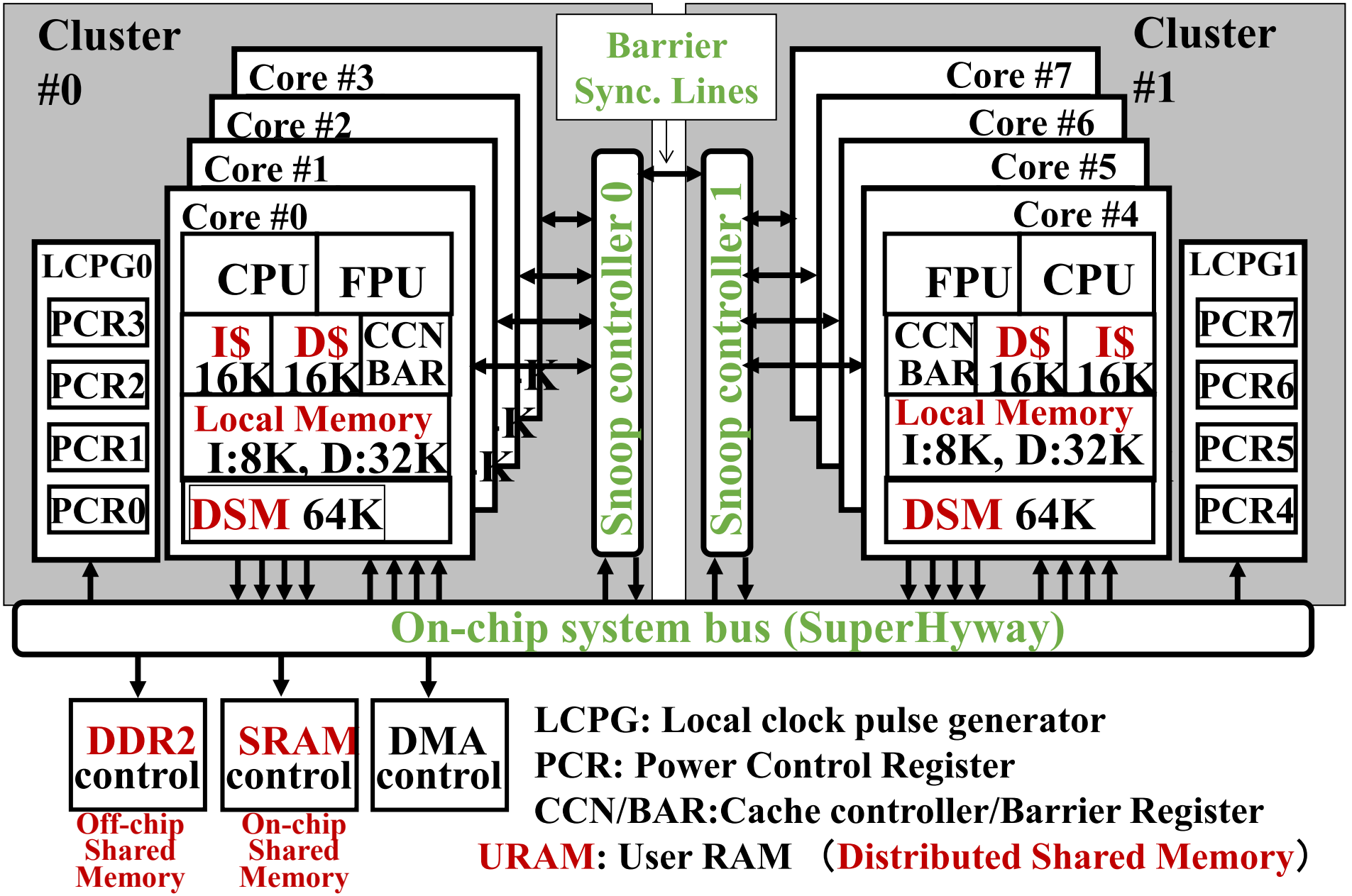
Power ∝ Frequency^3

周波数 Frequency を1/4にすると
(Ex. 4GHz→1GHz),
性能は 1/4、消費電力は 1/64
Performance 1/4, Power 1/64

<マルチコア(Multicore)>

8cores をチップに集積すると,
Performance 2 timesでPowerは 1/8

8 Core RP2 Chip Block Diagram

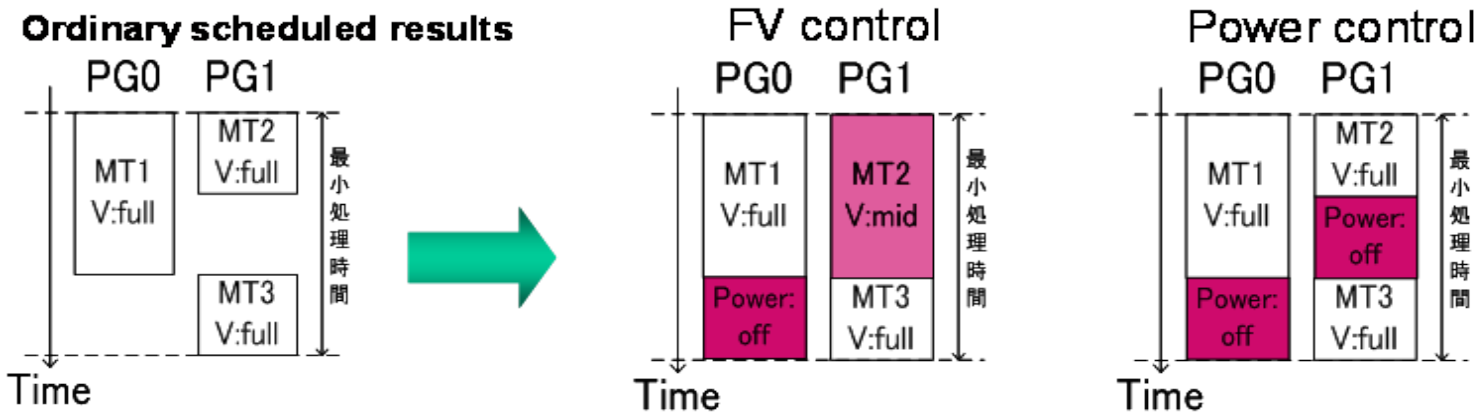


LCPG: Local clock pulse generator
PCR: Power Control Register
CCN/BAR:Cache controller/Barrier Register
URAM: User RAM （Distributed Shared Memory）

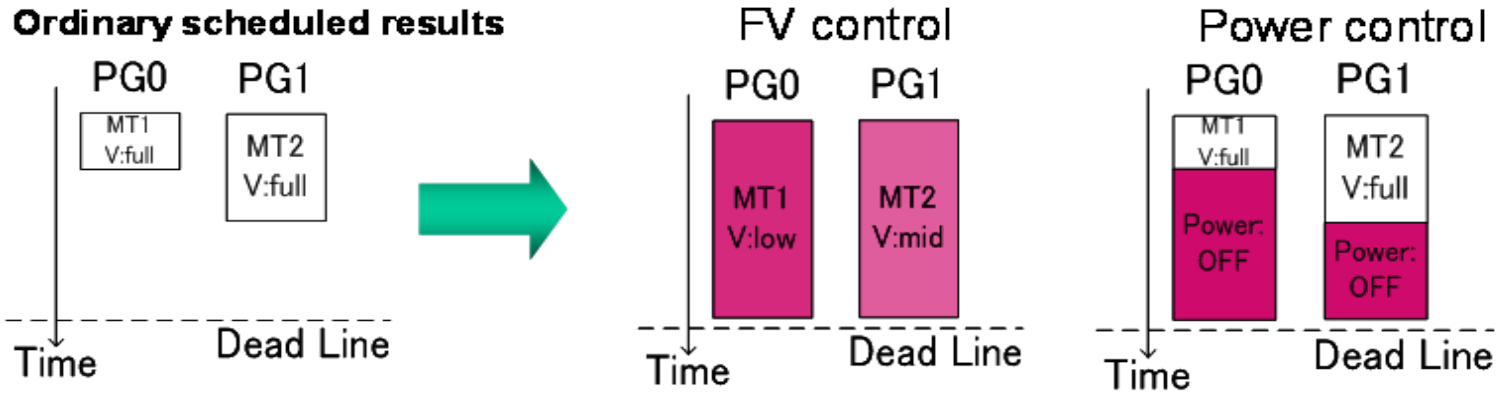
Power Reduction by Power Supply, Clock Frequency and Voltage Control by OSCAR Compiler

Frequency and Voltage (DVFS), Clock and Power gating of each cores are scheduled considering the task schedule since the dynamic power proportional to the cube of F (F^3) and the leakage power (the static power) can be reduced by the power gating (power off).

- **Shortest execution time mode**



- Realtime processing mode with dead line constraints



In this Fig. Frequency Full, Mid, Low

Power OFF: Power Gating

An Example of Machine Parameters for the Power Saving Scheme

- **Functions of the multiprocessor**
  - **Frequency of each proc. is changed to several levels**
  - **Voltage is changed together with frequency**
  - **Each proc. can be powered on/off**

state	FULL	MID	LOW	OFF
frequency	1	1 / 2	1 / 4	0
voltage	1	0.87	0.71	0
dynamic energy	1	3 / 4	1 / 2	0
static power	1	1	1	0

- **State transition overhead**

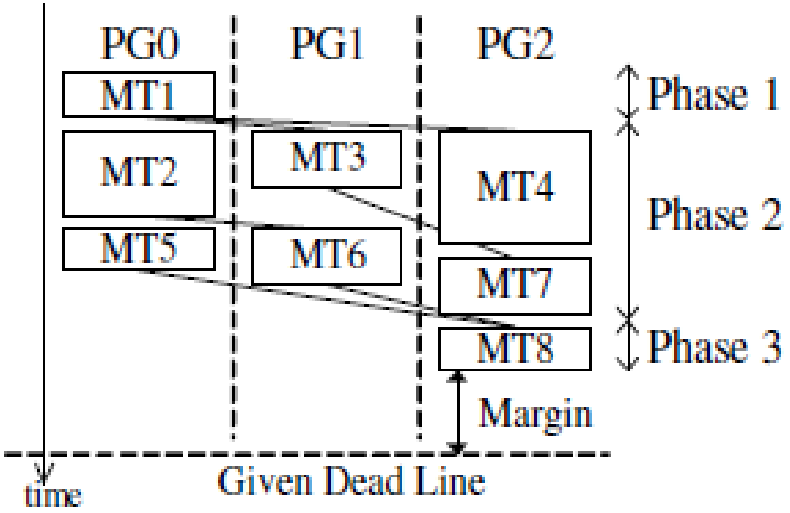
state	FULL	MID	LOW	OFF
FULL	0	40k	40k	80k
MID	40k	0	40k	80k
LOW	40k	40k	0	80k
OFF	80k	80k	80k	0

delay time [u.t.]

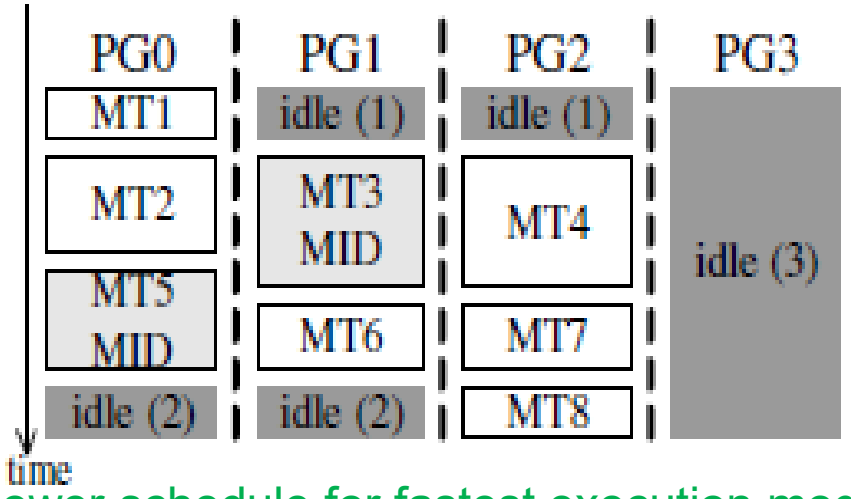
state	FULL	MID	LOW	OFF
FULL	0	20	20	40
MID	20	0	20	40
LOW	20	20	0	40
OFF	40	40	40	0

energy overhead [μJ]

Power Reduction Scheduling



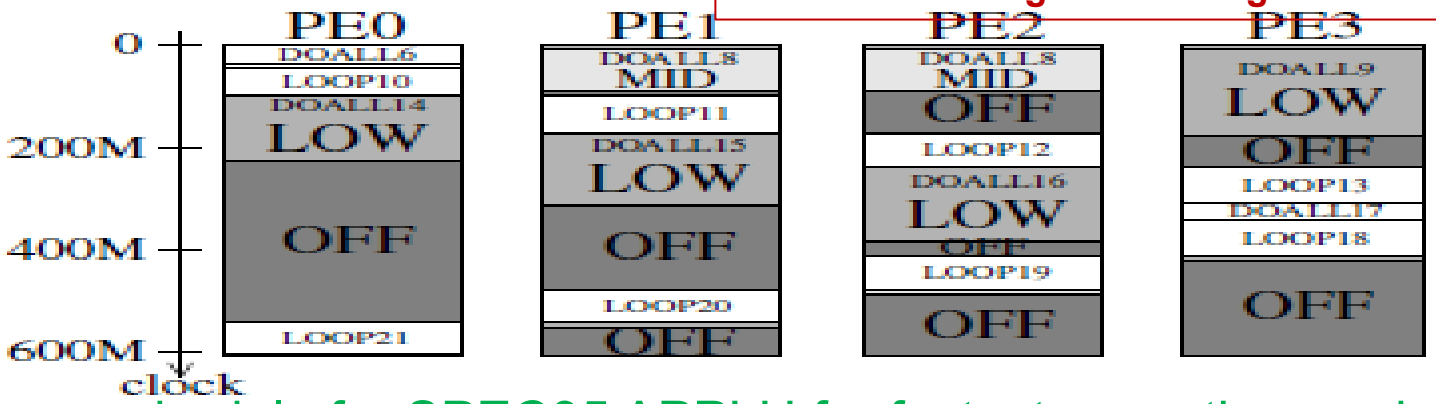
A macrotask graph assigned to 3 cores



A power schedule for fastest execution mode

Realtime scheduling mode
MTs 1,4,7,8 are on Critical Path (CP)

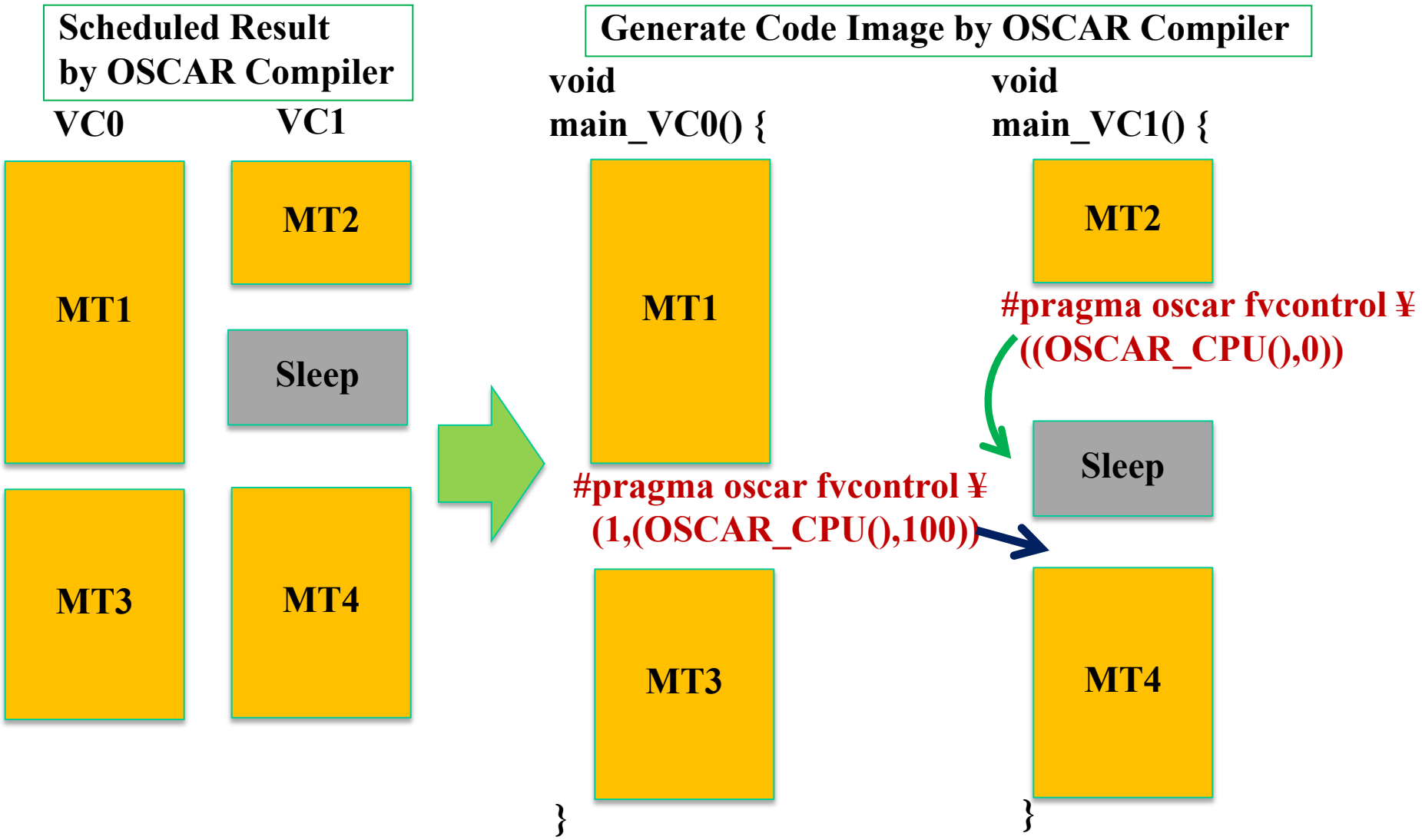
1) **Reduce frequencies (Fs) of MTs on CP considering dead line.**
2) **Reduce Fs of MTs not on CP. Idle: Clock or Power Gating considering overheads.**



A power schedule for SPEC95 APPLU for fastest execution mode

Doall6, Loop 10,11,12,13, Doall 17, Loop 18,19,20, 21 are on CP

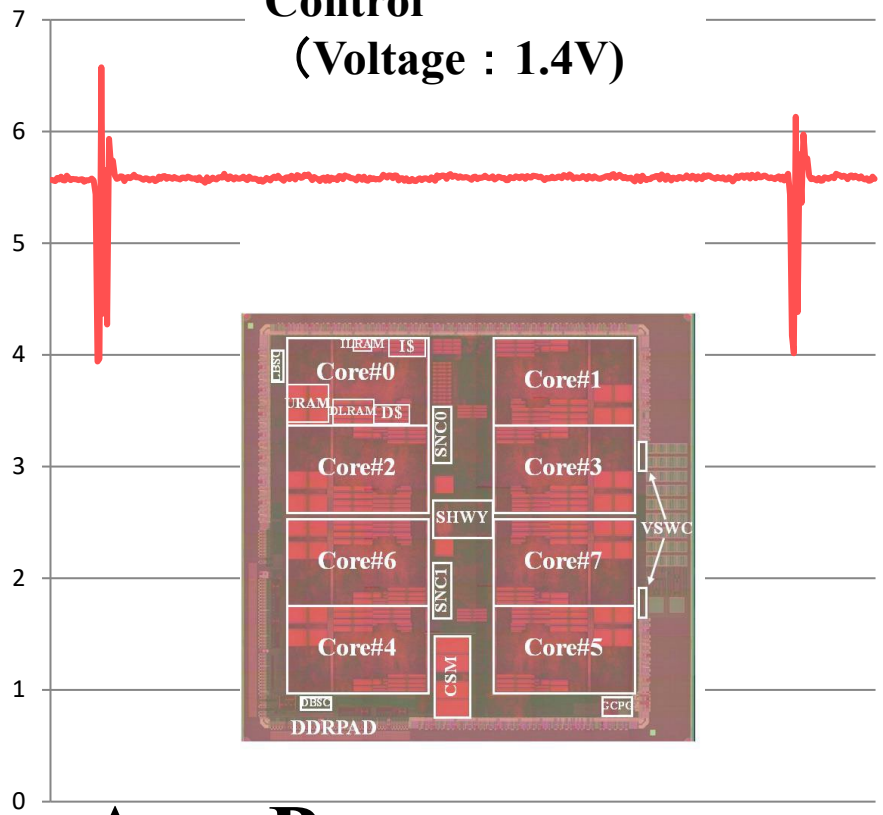
Low-Power Optimization with OSCAR API



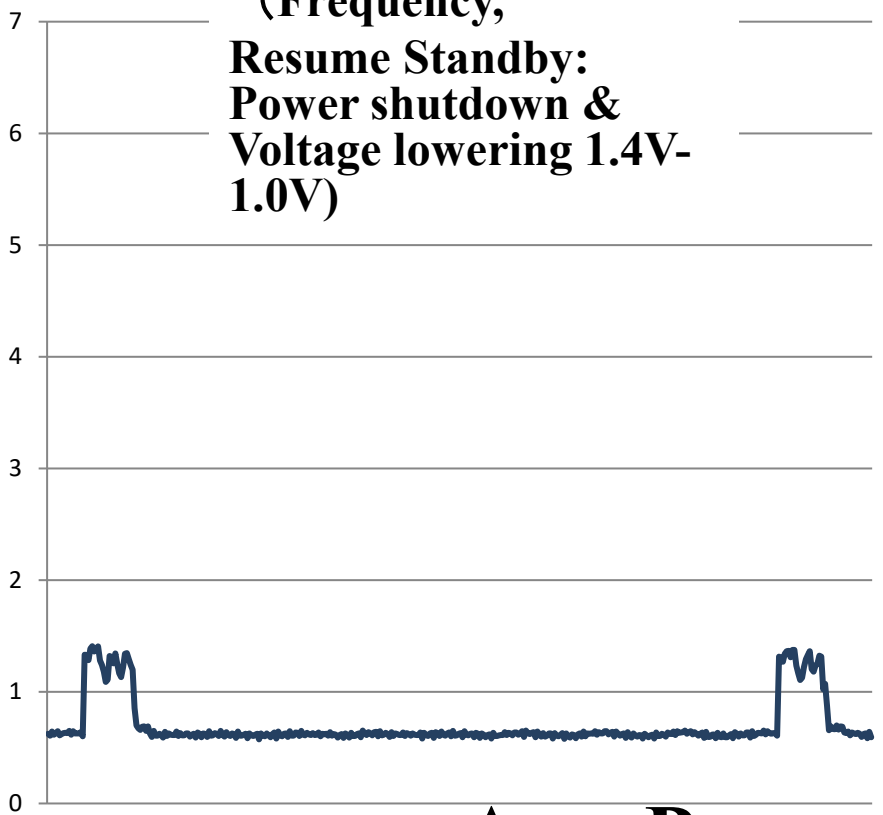
Power Reduction by OSCAR Parallelizing Compiler for Secure Audio Encoding

AAC Encoding + AES Encryption with 8 CPU cores

Without Power Control (Voltage : 1.4V)



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

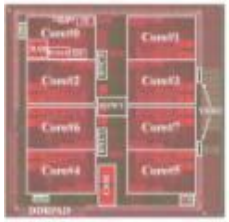


Avg. Power 5.68 [W]

88.3% Power Reduction

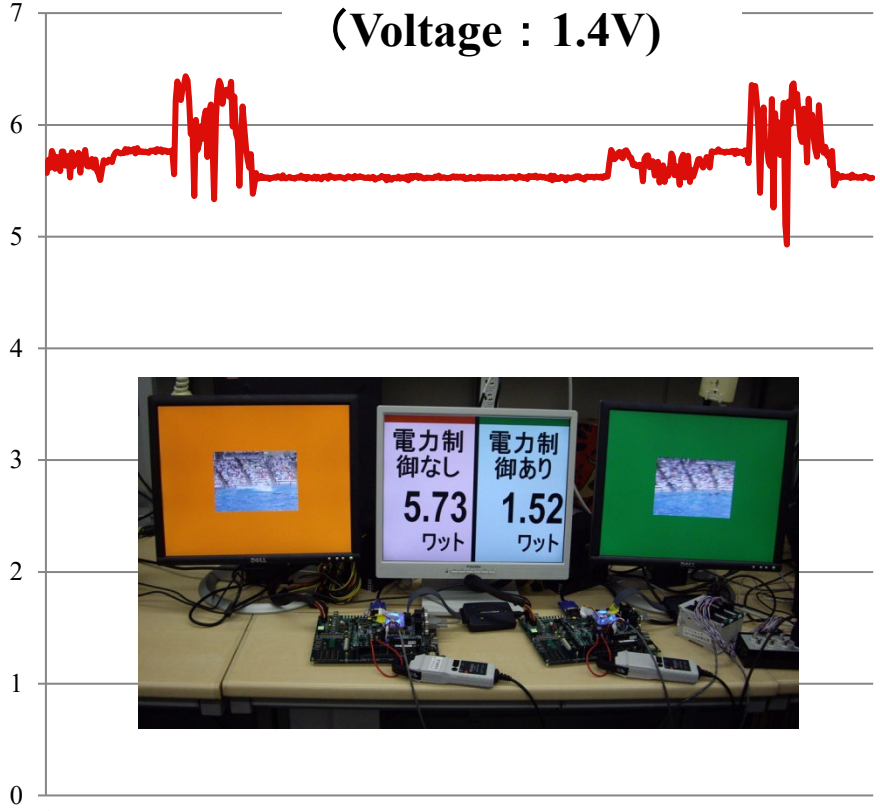
Avg. Power 0.67 [W]

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

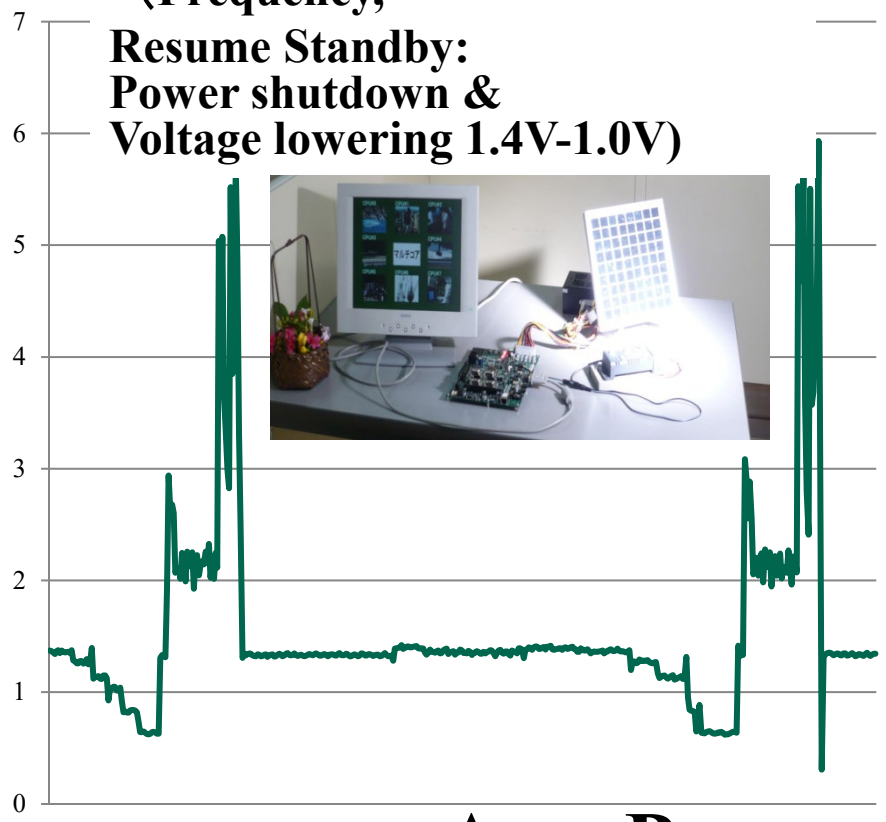


MPEG2 Decoding with 8 CPU cores

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]

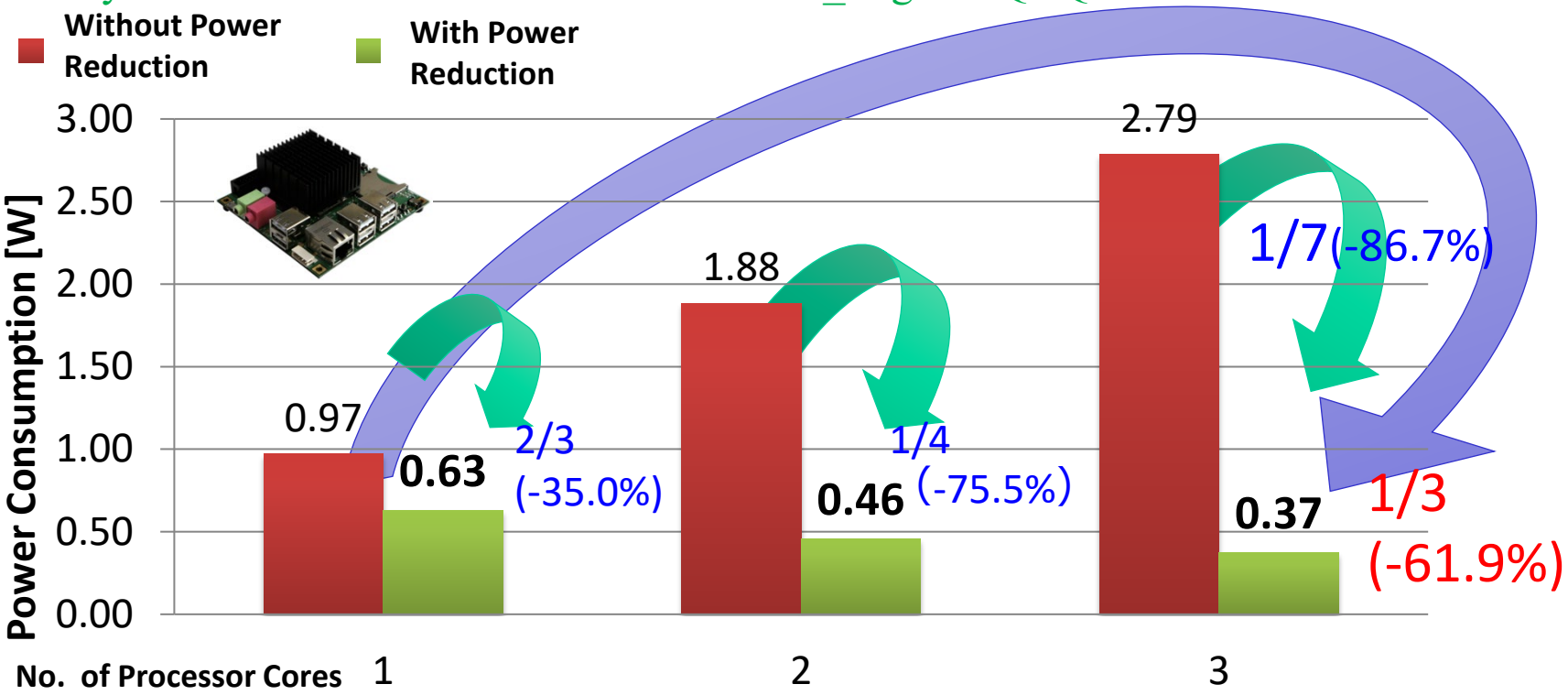
73.5% Power Reduction

Avg. Power 1.52 [W]

Automatic Power Reduction for MPEG2 Decode on Android Multicore

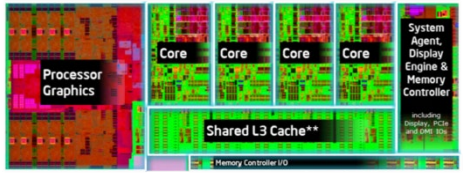
ODROID X2 ARM Cortex-A9 4 cores

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



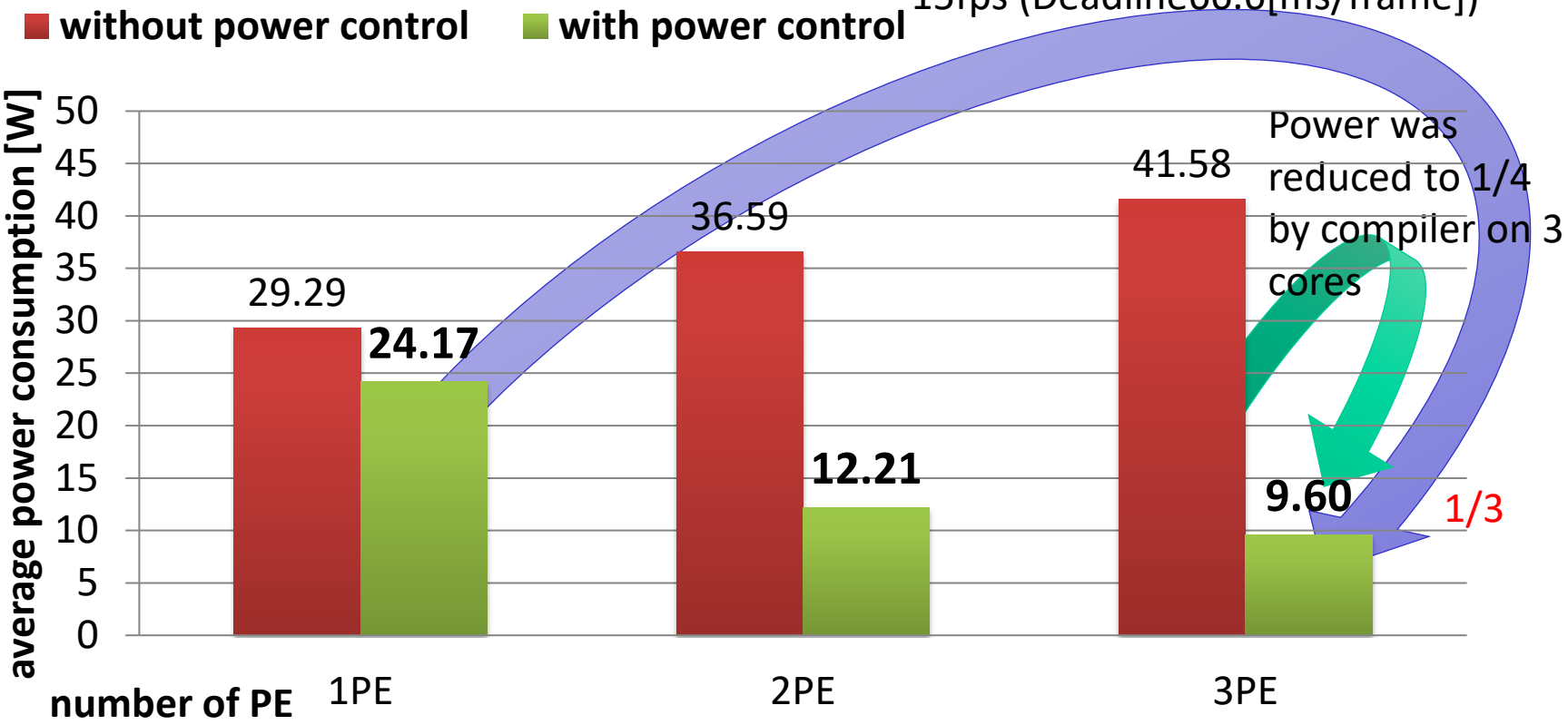
- On 3 cores, Automatic Power Reduction control successfully reduced power to 1/7 against without Power Reduction control.
- 3 cores with the compiler power reduction control reduced power to 1/3 against ordinary 1 core execution.

Power Reduction on Intel Haswell for Real-time Optical Flow



Intel CPU Core i7 4770K

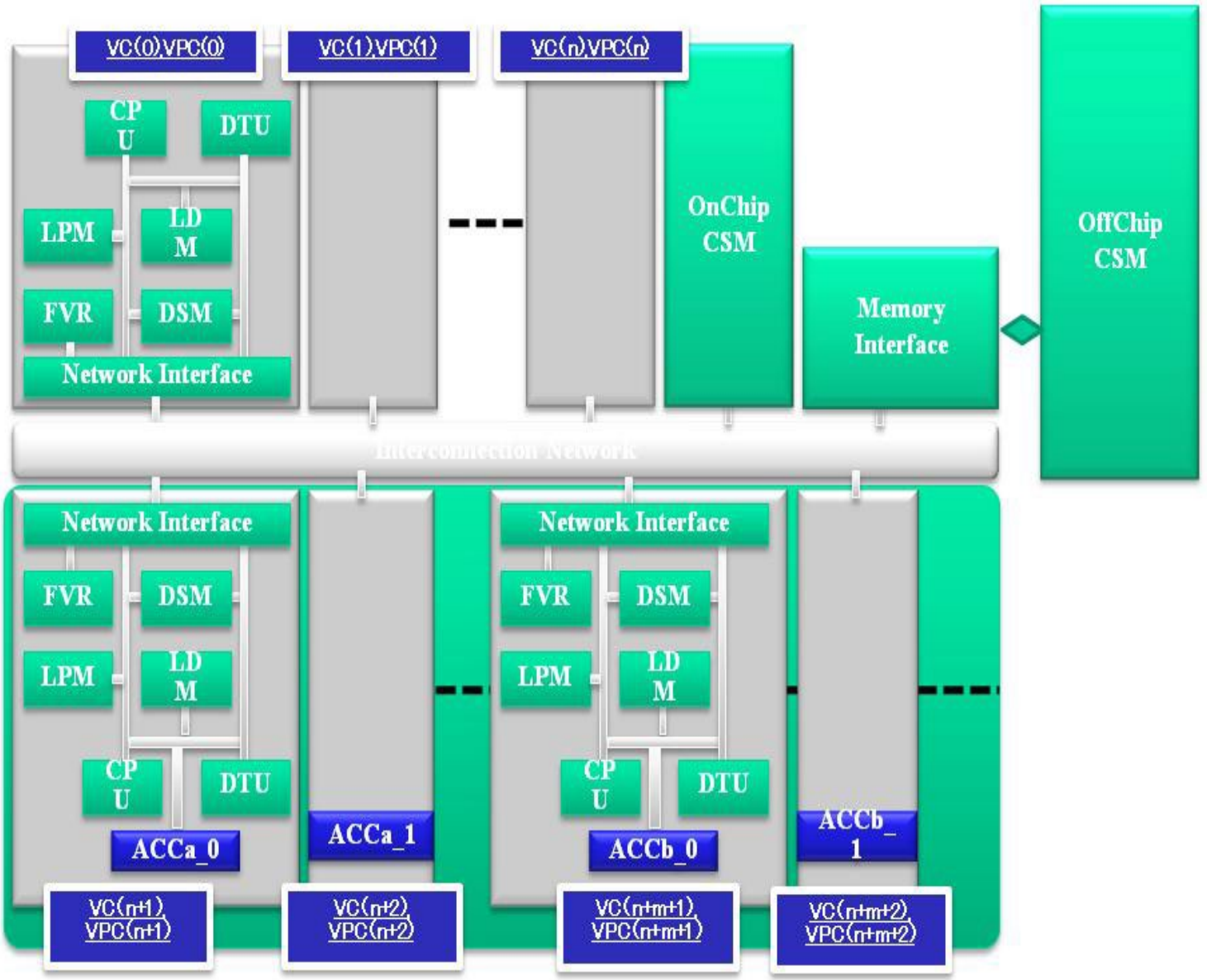
For HD 720p(1280x720) moving pictures
15fps (Deadline66.6[ms/frame])



Power was reduced to 1/4 (9.6W) by the compiler power optimization on the same 3 cores (41.6W).

Power with 3 core was reduced to 1/3 (9.6W) against 1 core (29.3W) .

OSCAR Heterogeneous Multicore



DTU
- Data Transfer Unit

LPM
- Local Program Memory

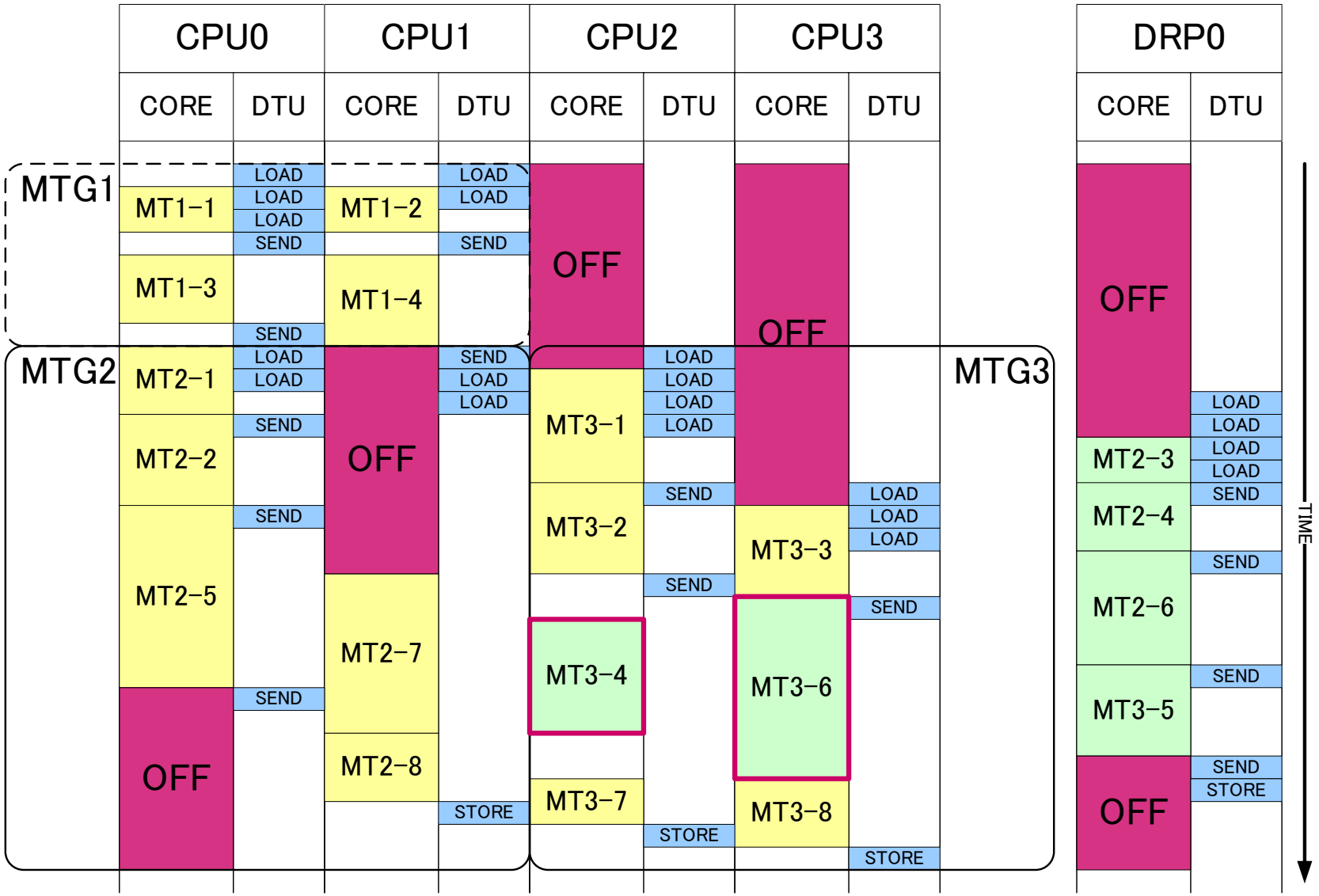
LDM
- Local Data Memory

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



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

(LCPC2009 Homogeneous, 2010 Heterogeneous)

Specification: http://www.kasahara.cs.waseda.ac.jp/api/regist.php?lang=en&ver=2.1

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

(* from OpenMP)

- 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

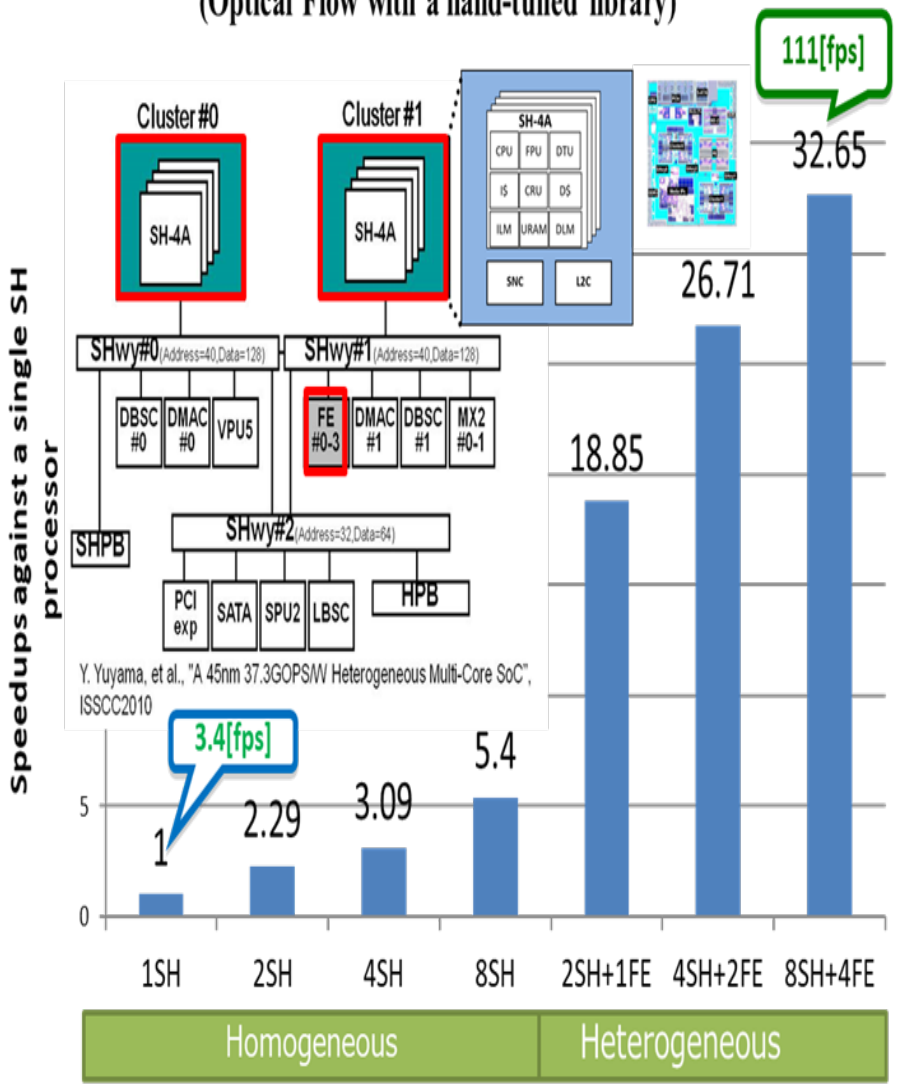
2 hint directives for OSCAR compiler
- accelerator_task
- oscar_comment

from V2.0

Speedups & 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)

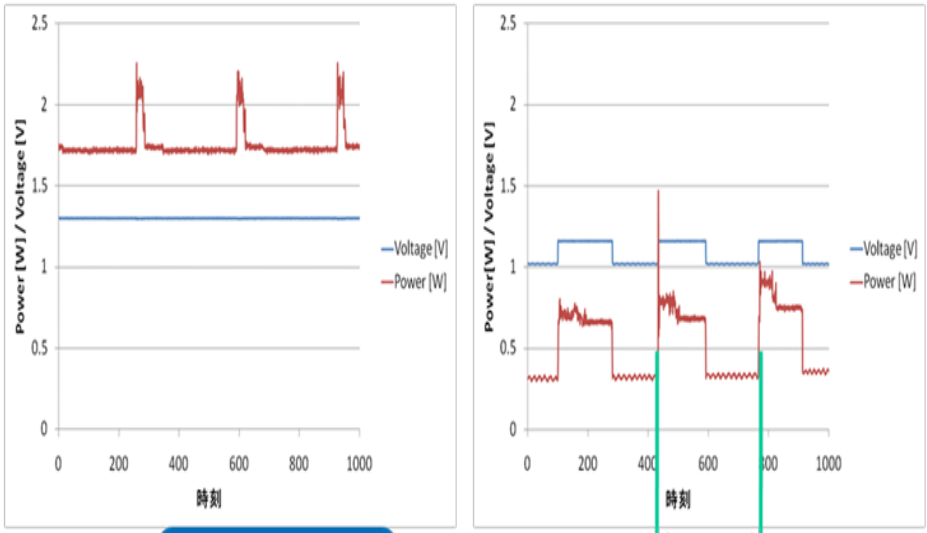
Without Power Reduction

With Power Reduction by OSCAR Compiler

70% of power reduction

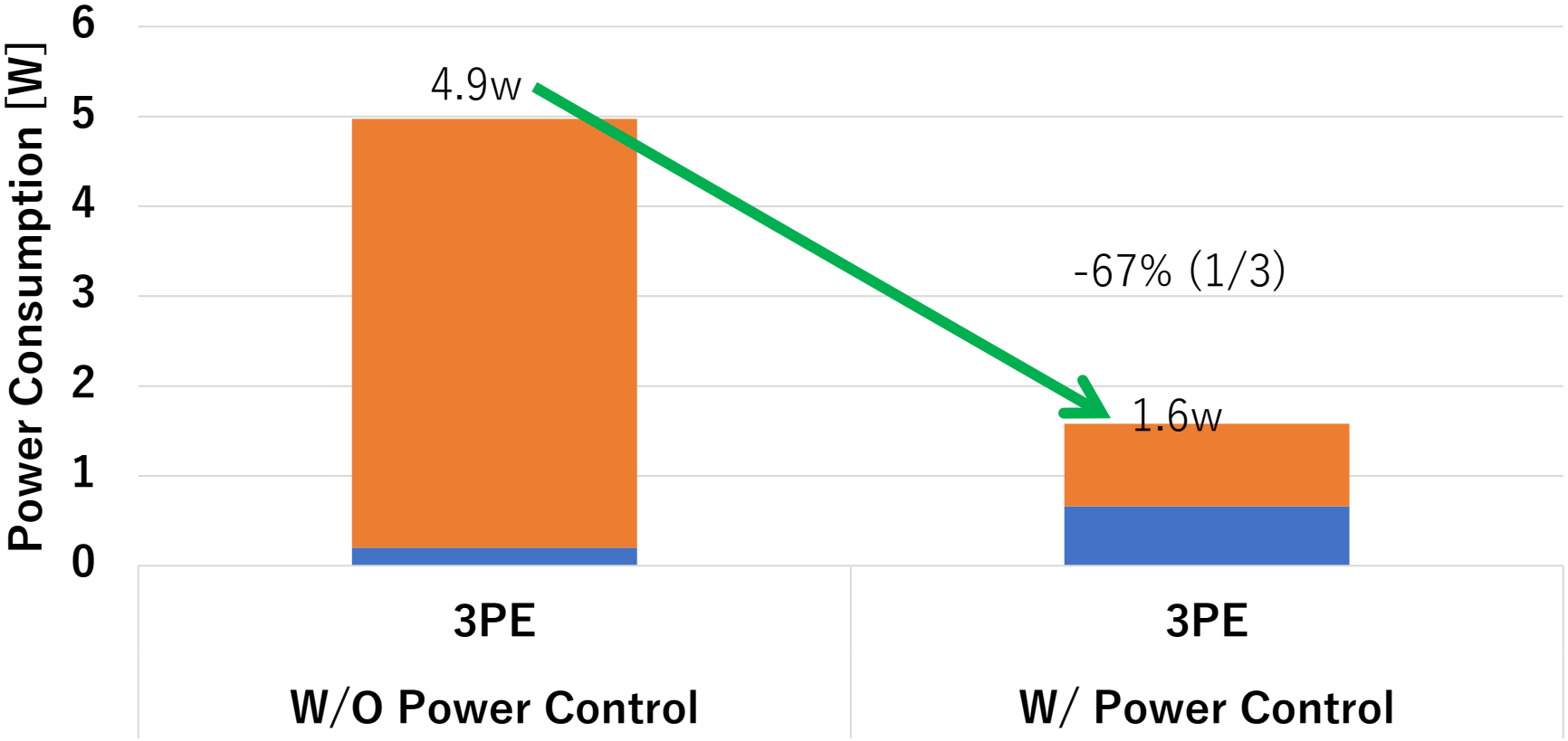
Average:1.76[W]

Average:0.54[W]



1cycle : 33[ms]
→30[fps]

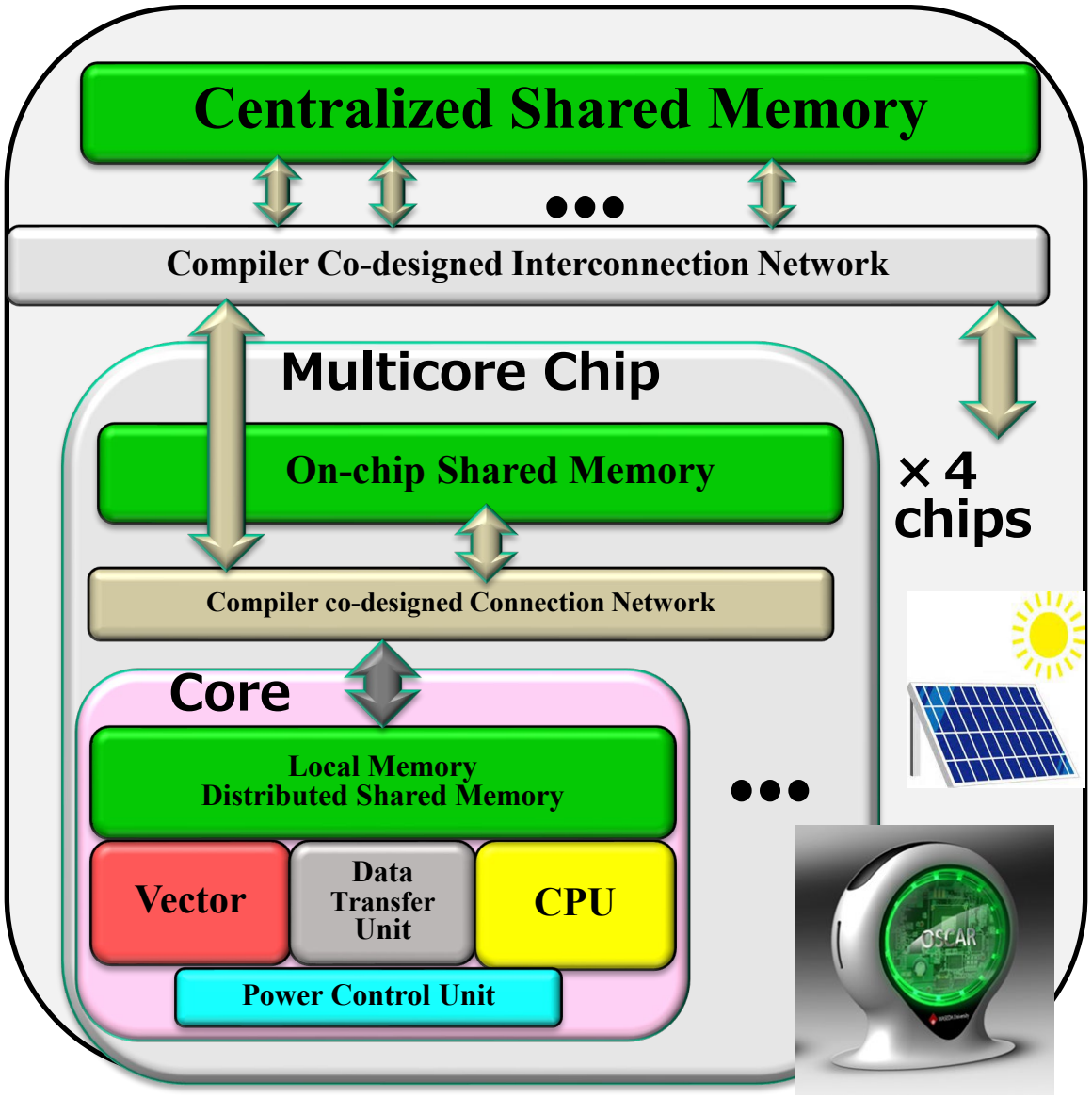
Automatic Power Reduction of OpenCV Face Detection on big.LITTLE ARM Processor



- 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

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



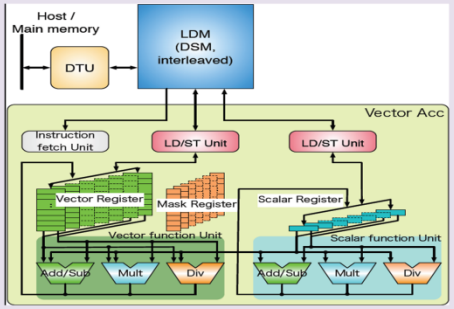
Target:

- **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

IEEE COMPUTER SOCIETY

https://www.computer.org/product/education/multi-core-video-lectures-bundle

Created by Multicore STC Chair Hironori Kasahara

Multi-Core Lecture Series consists of 11 one-hour lectures by some of the world's leading researchers in the field. This series is not a course and it consists of the presentation for those who are in the research field. This is more intended for research information sharing than educational training. Topics that are covered during these lectures are listed below. This series also includes an hour discussion of the lecturers.

Video Presentations:

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

Dependences and Dependence Analysis by Utpal Banerjee Utpal Banerjee's research interests in computer science are in the general area of parallel processing and he has published four books on loop transformations and dependence analysis, with a fifth one on instruction level parallelism on the way.

Multigrain Parallelization and Power Reduction by Hironori Kasahara. Professor Kasahara has been researching on OSCAR Automatic Parallelizing and Power Reducing Compiler and OSCAR Multicore architecture for more than 30 years, and led four Japanese national projects on parallelizing compilers, multicores, and green computing.

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

ISCA2025, June 21-25, 2025, Waseda University, Tokyo, Japan

ACM SIGARCH

General Co-Chairs: Jean-Luc Gaudiot (Prof. UCI)
Hironori Kasahara (Prof. Waseda)

IEEE COMPUTER SOCIETY TCCA
Technical Committee on Computer Architecture