高性能・低消費電力マルチコアプロセッサのためのコンパイラ・アーキテクチャ協調設計

早稲田大学　理工学術院　情報理工学科　教授　笠原博徳

グリーンコンピューティング機構　アドバンストマルチコアプロセッサ研究所長

IEEE Computer Society President 2018

1976 早稲田大学高等学院卒
1980 早大電気工学科卒、1982 同修士課程了
1985 早大大学院博士課程了 工学博士,学振第一回PD
カリフォルニア大学バークレー客員研究員
1986 早大理工専任講師、1988年 助教授
1989～1990 イリノイ大学Center for Super computing R&D客員研究員
1997 教授、現在 理工学術院情報理工学科
2004 アドバンストマルチコア研究所所長
2017 日本工学アカデミー会員（2020より理事）、日本学術会議連携会員
2018 IEEE Computer Society会長、早大副総長（-2022年9月）
2019 産業競争力懇談会(COCN) 理事

【受賞】
1987 IFAC World Congress Young Author Prize
1997 情報処理学会坂井記念特別賞
2005 半導体理工学研究センタ共同研究賞
2008 LSI・オブ・ザ・イヤー 2008 準グランプリ、Intel Asia Academic Forum Best Research Award
2010 IEEE CS Golden Core Member Award
2014 文部科学大臣表彰科学技術賞研究部門
2015 情報処理学会フェロー
2017 IEEE Fellow, IEEE Eta-Kappa-Nu
2019 IEEE CS Spirit of Computer Society Award
2020 情報処理学会功績賞、SCAT表彰 会長大賞
2023 IEEE Life Fellow

査読付き論文232件、招待講演234件、国際特許取得67件(米・英・中・日等)、新聞・Web記事・TV等メディア掲載 698件

【政府・学会委員等】 歴任数 295件
IEEE Computer Society President 2018、Executive Committee委員長、理事(2009-14)、戦略的計画委員長、Nomination Committee委員長、Multicore STC 委員長、IEEE CS Japan委員長、IEEE技術委員、IEEE Medal選定委員、ACM/IEEE SC'21基調講演選定委員等
【経済産業省・NEDO】情報家電用マルチコア・アドバンスト並列化コンパイラ・グリーンコンピューティング・プロジェクトリーダ、NEDOコンピュータ戦略委員長等
【内閣府】スーパーコンピュータ戦略委員、政府調達苦情検討委員、総合科学技術会議情報通信PT 研究開発基盤領域&セキュリティ・ソフト検討委員、日本国際賞選定委員
【文部科学省・海洋研】地球シミュレータ（ES）中間評価委員、情報科学技術委員、HPCI計画推進委員、次世代スパコン（京）中間評価委員・概念設計評価委員、地球シミュレータES2導入技術アドバイザイリー委員長等、JST: ムーンショットG3ロボット&AI Vice Chair, SBIRフェーズ1委員長等

AIチップ設計拠点フォーラム（第46回）, 2023年4月28日（金）, 13:35-14:35

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

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

Practical Multiprocessor Scheduling Algorithms for Efficient Parallel Processing

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

Courtesy of dexchao - Fotolia.com

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

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

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

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

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

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

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

PARALLEL PROCESSING OF ROBOT MOTION SIMULATION

H. Kasahara, H. Fujii and M. Iwata

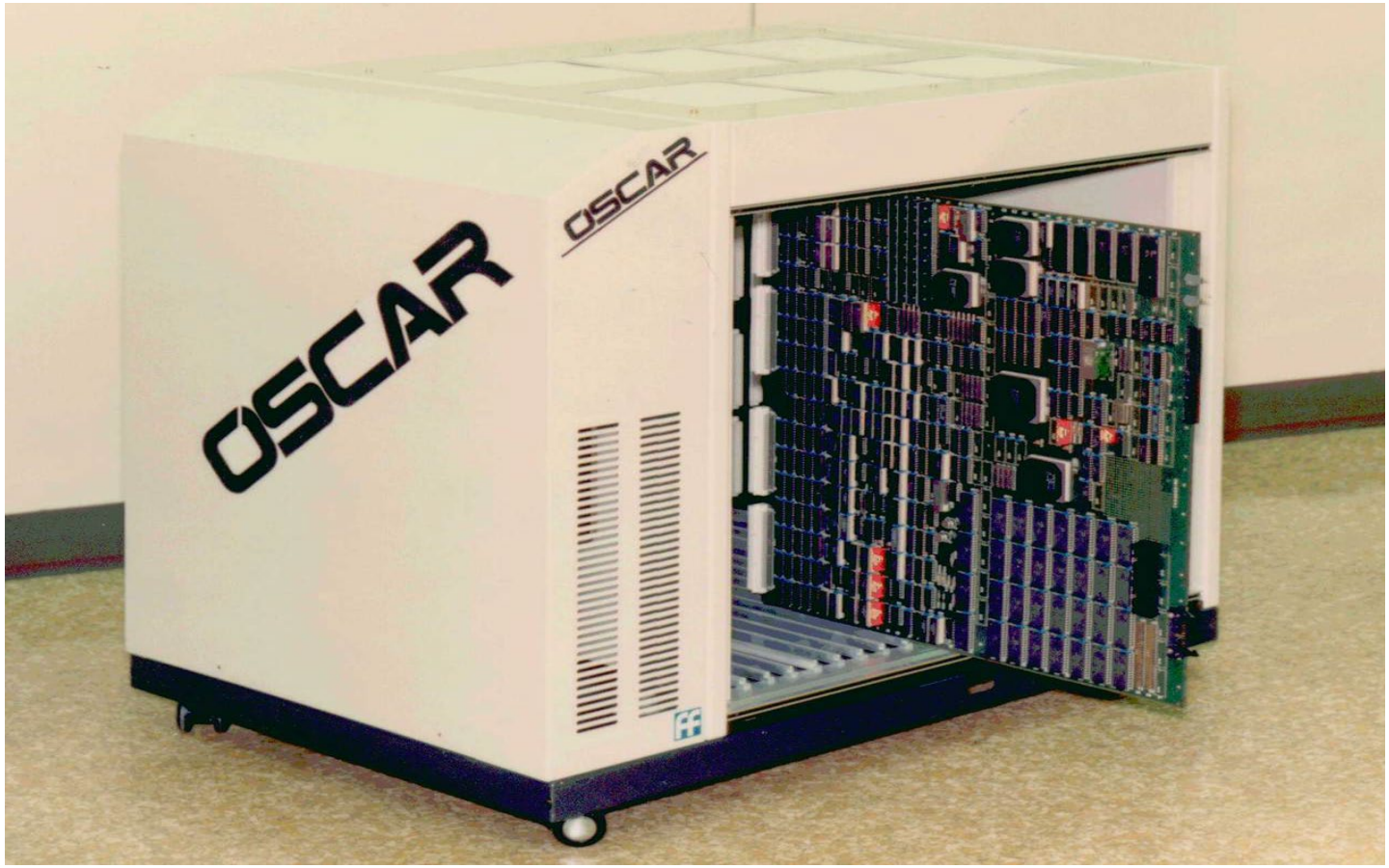
Department of Electrical Engineering, Waseda University, 3–4–1 Ohkubo Shinjuku-ku, Tokyo 160, Japan

1987 OSCAR(Optimally Scheduled Advanced Multiprocessor)

ソフトウェア（コンパイラ）とハードウェア（アーキテクチャ）の協調設計したマルチプロセッサ

Co-design of Compiler and Architecture

Looking at various applications, design a parallelizing compiler and design a multiprocessor/multicore-processor to support compiler optimization

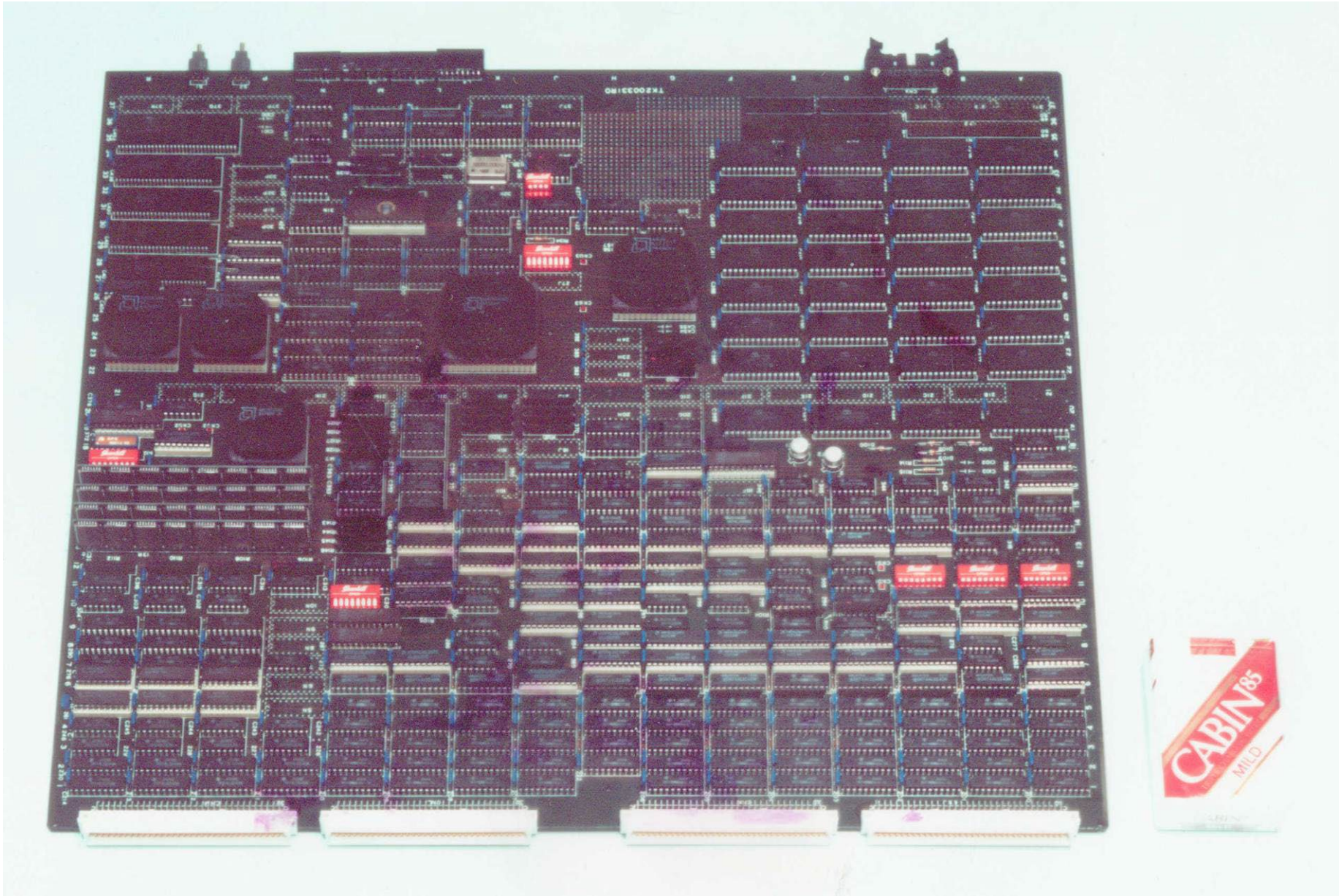


D. Lenoski, J. Laudon, K Gharachorloo, A. Gupta, J. Hennessy, "The directory-based cache coherence protocol for the DASH multiprocessor," ACM SIGARCH Computer Architecture News 18 (2SI), 148-159, 1990

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

1987 OSCAR PE Board

(32bit RISC: Reduced Instruction Set Computer)



毎年生産される32bit,64bitプロセッサの99%はRISC

https://www.acm.org/hennessy-patterson-turing-lecture

https://dl.acm.org/cms/attachment/164b78fa-de5b-4d7e-bb28-17084e83e06a/f6.jpg

RISCはスマホ,IoT,自動車制御からサーバ,スパコン等まで広く使用されている

例 : ARM,IBM Power,Renesas RH850, Infineon,SPARC,RISC V

64-bit
iPhone 13
2021

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

IBM Watson

http://watson2016.com/

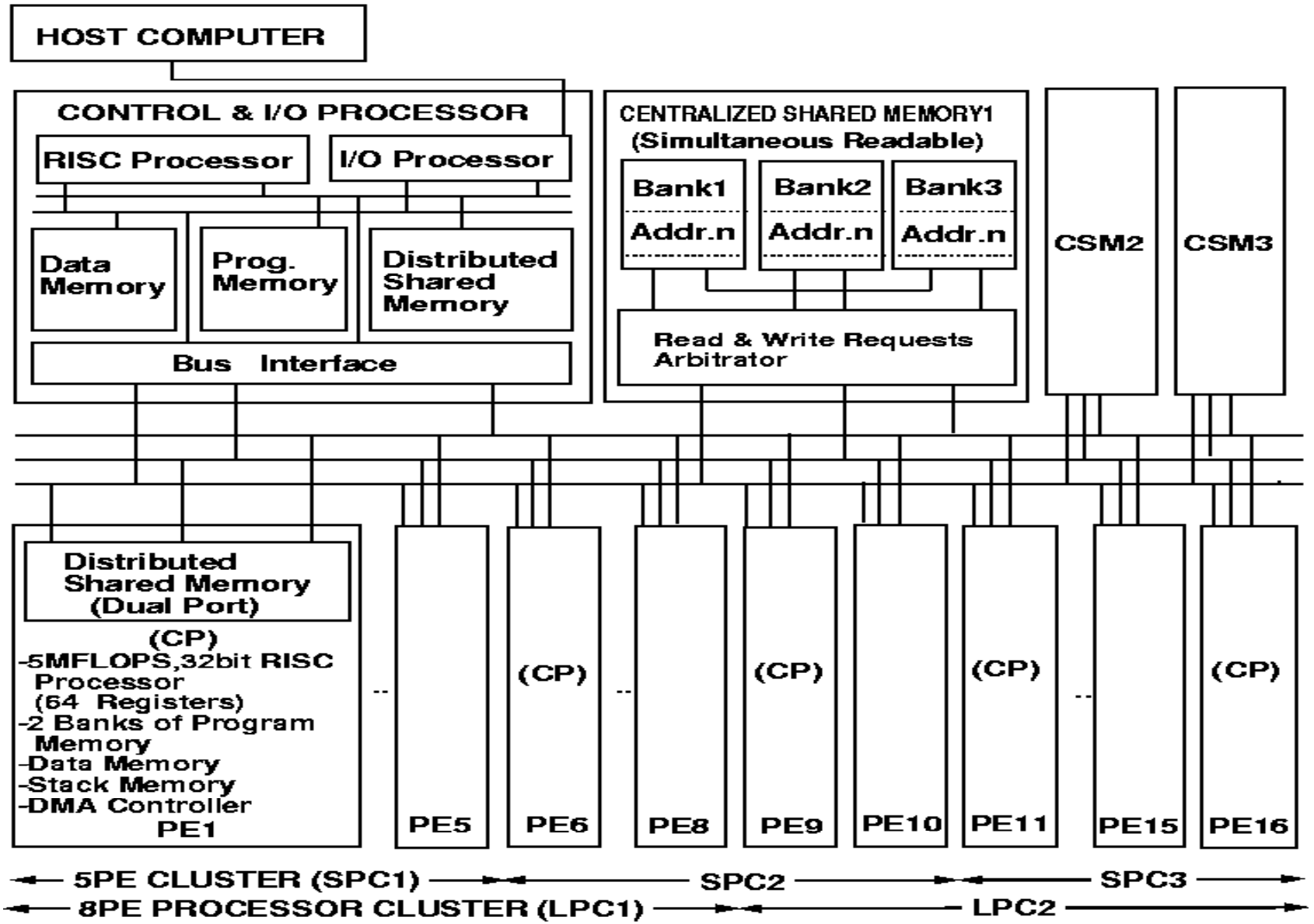
Renesas
自動車制御

https://www.renesas.com/

理研富岳

https://fugaku100kei.jp/fugaku/

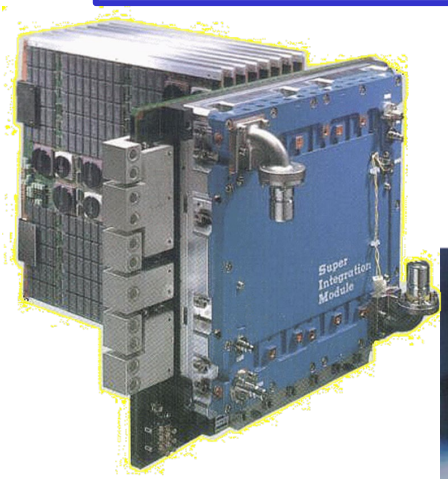
OSCAR(Optimally Scheduled Advanced Multiprocessor)

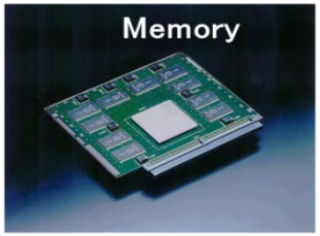


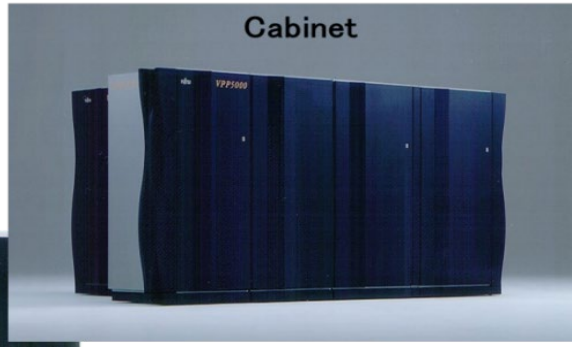
1993年 スーパーコンピュータVPP500、数値風洞(NWT)

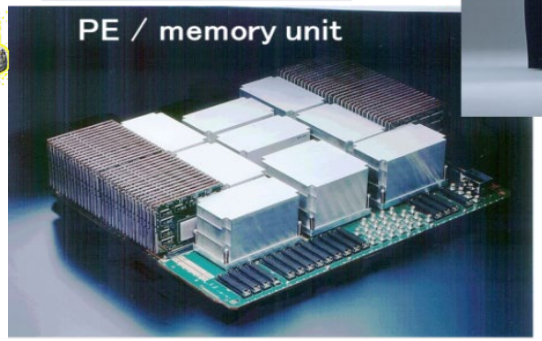
Mr. Hajime Miyoshi

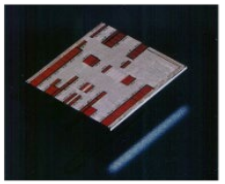
ACM/IEEE SC '94: Washington, D.C. November, 1994にて発表





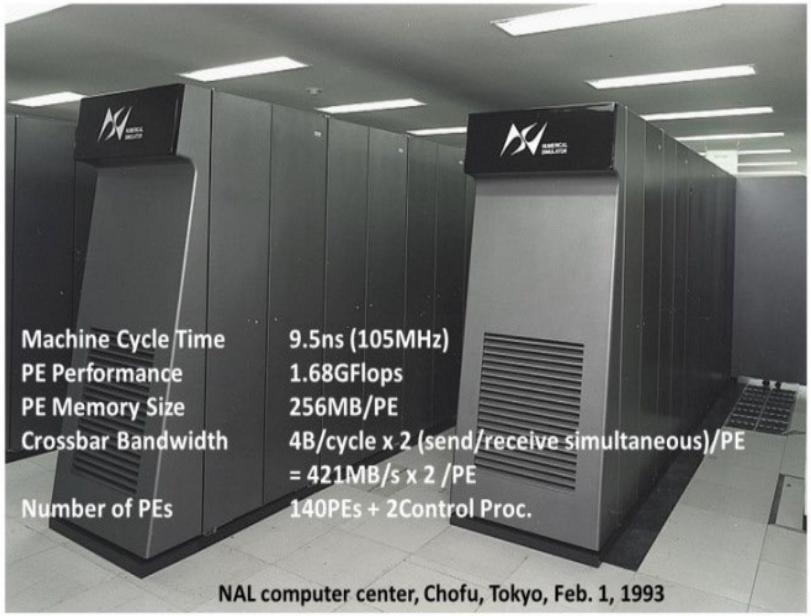




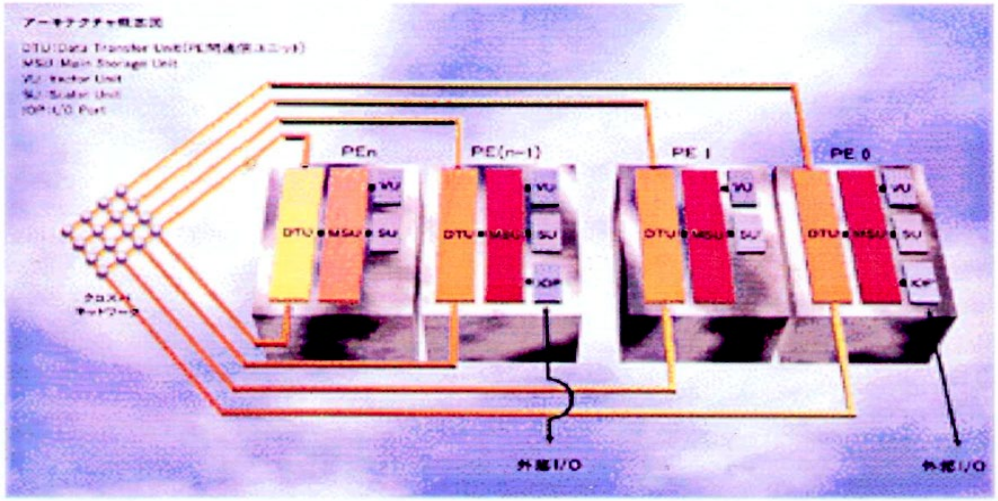


CMOS LSI

スパーコンピュータNWTの外観

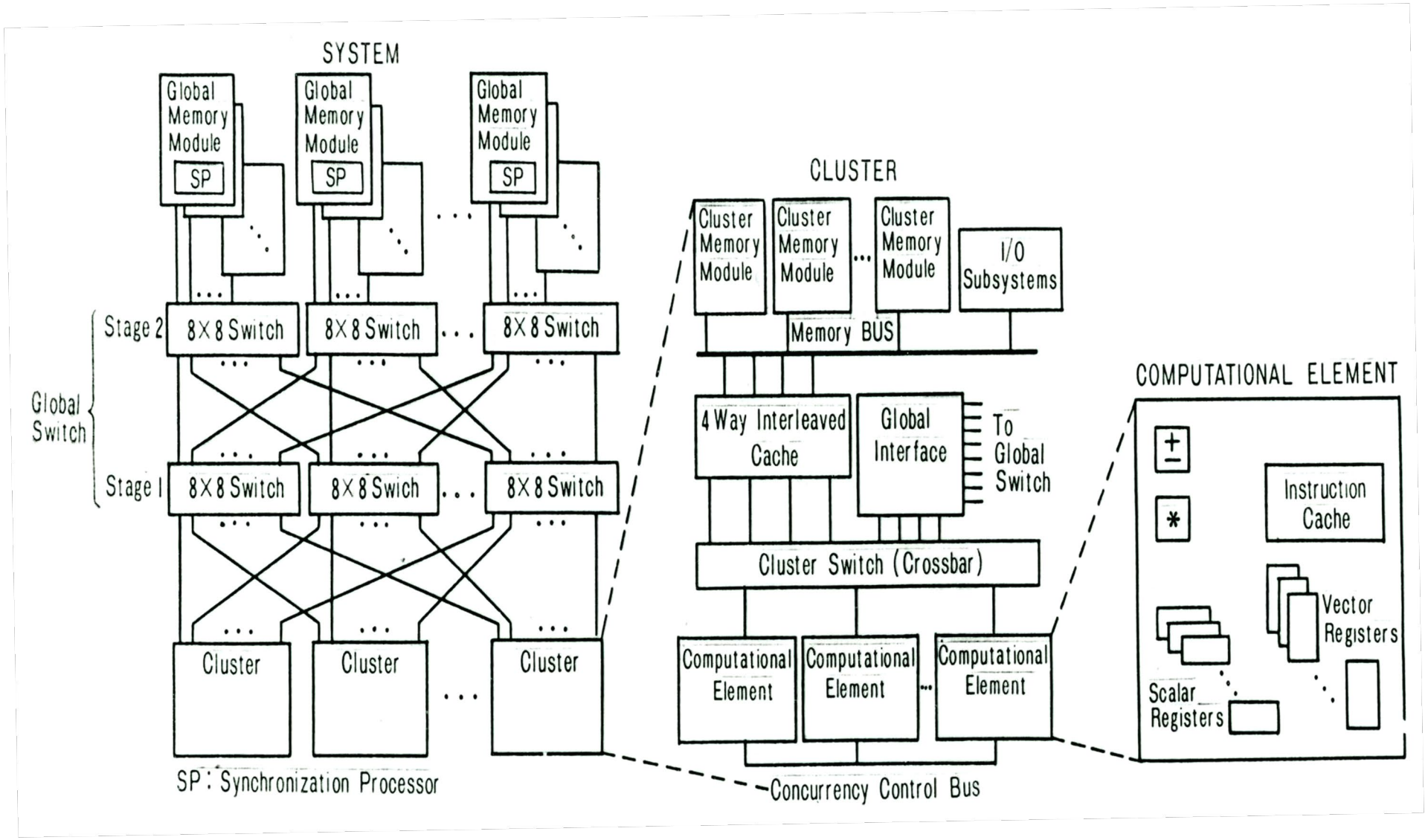


商用VPP5000 (仏気象庁他)



Cedar Supercomputer

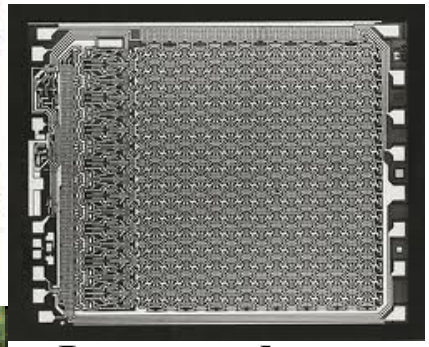
University of Illinois at Urbana-Champaign, CSRD (Center for Supercomputing R&D)



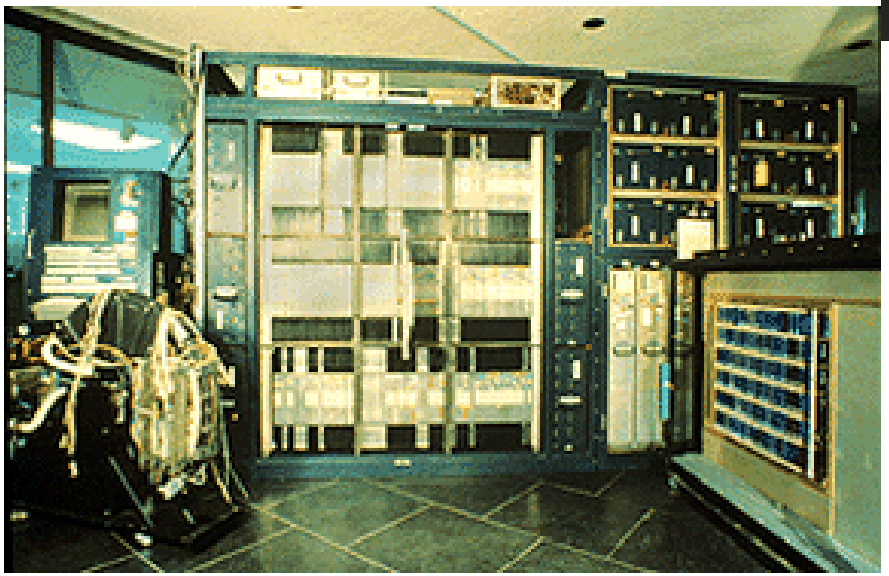
ILLIAC IV, Univ. Illinois at Urbana-Champaign & Burroughs

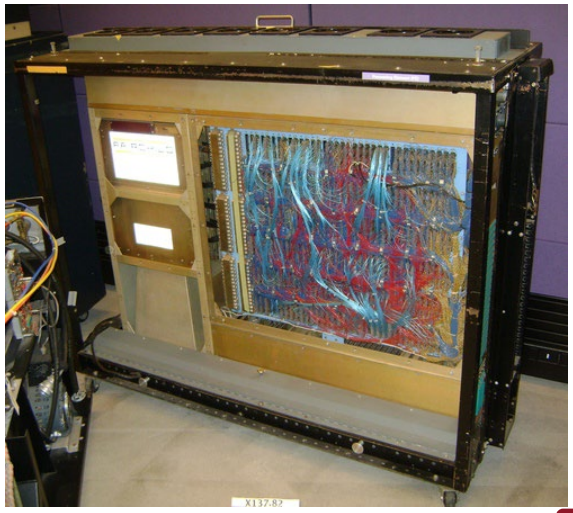
SIMD
64 Processor Element, Processor Array

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



Integrated circuits





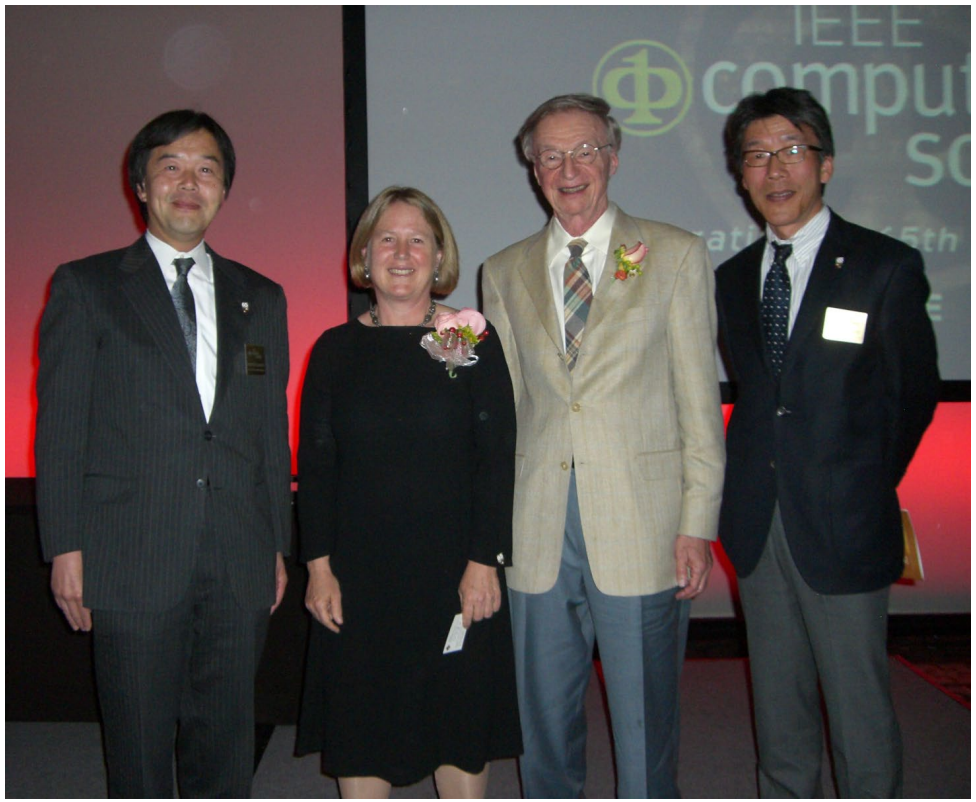
Processing Element

Prof. David J Kuck (Univ. Illinois, Intel)

IEEE Computer Pioneer Award 2011

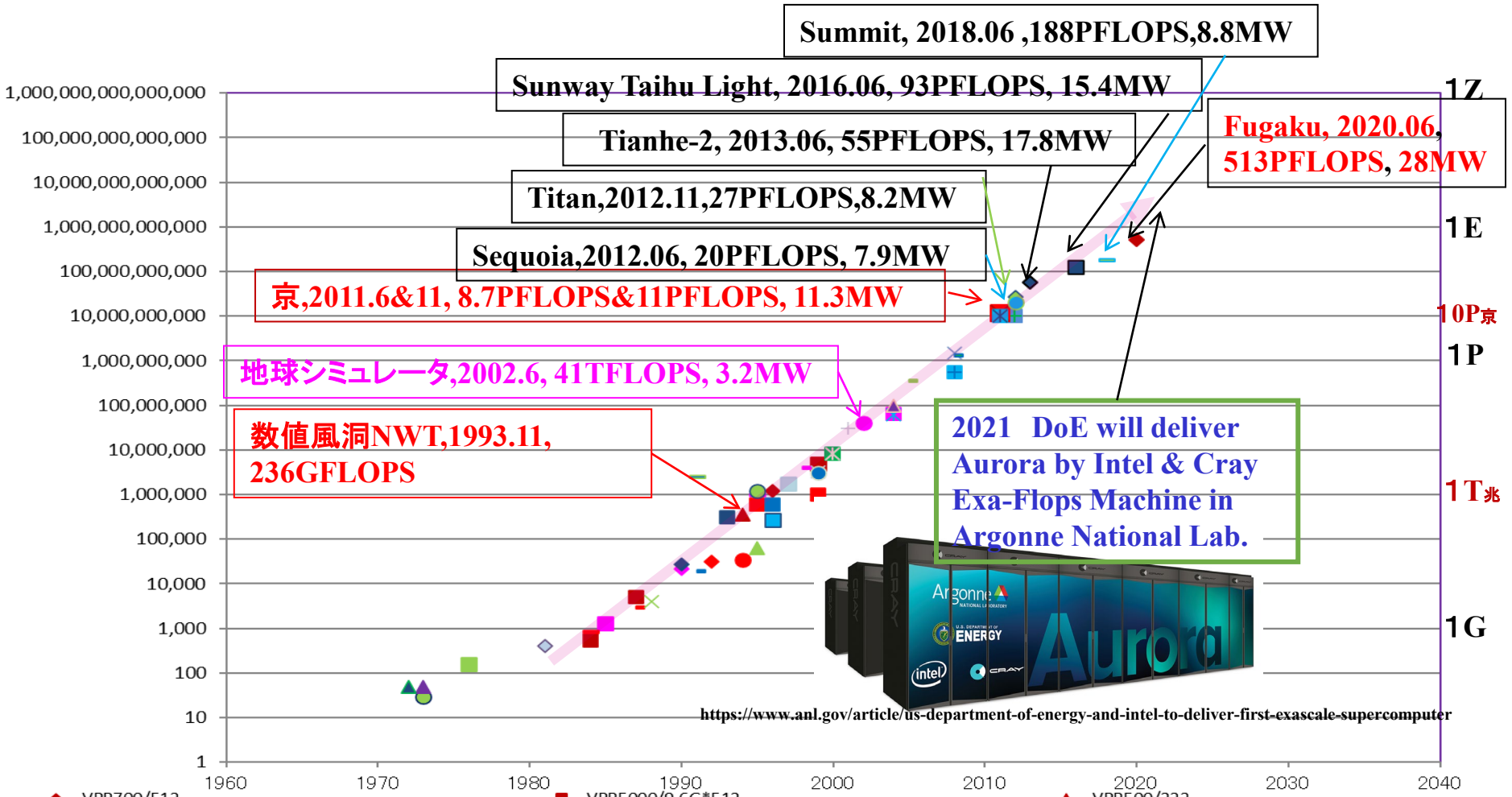
Ms. Diane B. Greene (VMware Cofounder & CEO)

IEEE Computer Society Computer Entrepreneur Award 2011



2018.03.08 WasedaUniv. Symposium on Future of High Performance Green Computing 2018 (HPGC2018)

Trend of Peak Performances of Supercomputers



Earth Simulator

2021年ノーベル物理学賞
プリンストン大 真鍋淑郎先生
大気・海洋大循環モデル

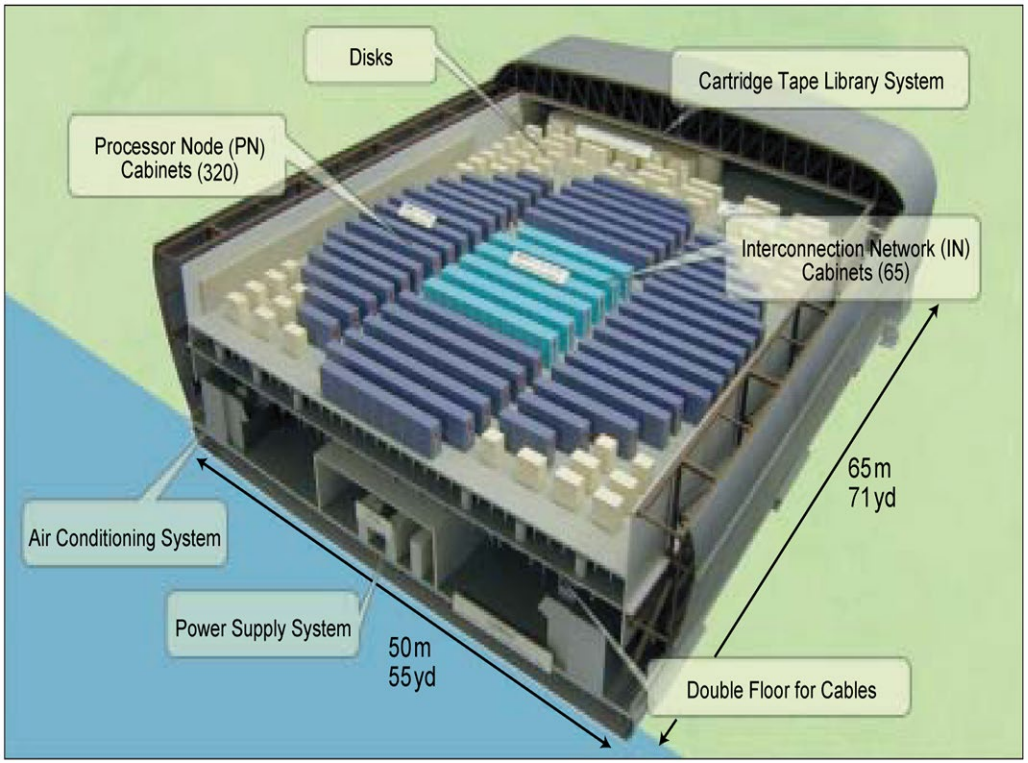
(http://www.es.jamstec.go.jp/)

- 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

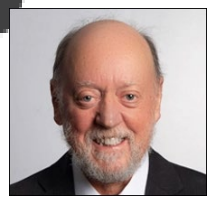
https://www.top500.org/

June 2022

PRESENTED BY FIND OUT MORE AT

TOP 500 The List.

2021 ACM A.M. Turing Award



DR. JACK DONGARRA

https://amturing.acm.org/award_winners/dongarra_3406337.cfm

For his pioneering contributions to numerical algorithms and libraries that enabled high performance computational software to keep pace with exponential hardware improvements for over four decades

top500.org

JUNE 2022	SYSTEM	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Frontier	HPE Cray EX235a, AMD Opt 3rd Gen EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-10	DOE/SC/ORNL	USA	8,730,112	1,102.0	21.3
2	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
3	LUMI	HPE Cray EX235a, AMD Opt 3rd Gen EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-10	EuroHPC/CSC	Finland	1,268,736	151.9	2.94
4	Summit	IBM POWER9 (22C, 3.07GHz), NVIDIA Volta GV100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/SC/ORNL	USA	2,414,592	148.6	10.1
5	Sierra	IBM POWER9 (22C, 3.1GHz), NVIDIA Tesla V100 (80C), Dual-Rail Mellanox EDR Infiniband	DOE/NNSA/LLNL	USA	1,572,480	94.6	7.44

No. 1 June 2022 **873万プロセッサ, 21MW**

Frontier - HPE Cray EX235a, 8,730,112 total cores, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X accelerators, HPE Slingshot-11 interconnect **168.5京回演算/秒**

Oak Ridge National Laboratory (ORNL) , USA

Rmax: 1.102 (ExaFlop/s), Rpeak 1.685 (ExaFlop/s)

Power:21.1 MW, Efficiency: 52.23 gigaflops/W

ACM チューリング賞

A.M. TURING CENTENARY CELEBRATION WEBCAST

MORE ACM AWARDS

コンピュータ分野のノーベル賞

Turing Award › Winners

毎年生産される200億個以上のプロセッサの99%がRISC

スタンフォード大前学長・Alphabet(Google親会社)会長

Jeffrey Ullman 2020 アルゴリズムとプログラミング言語

Geoffrey Hinton 2018 AI ディープ・ラーニング

John L. Hennessy 2017 コンピュータ構成法 RISC:スマホ-スパコン

Alfred Aho 2020 アルゴリズムとプログラミング言語

Yoshua Bengio 2018 AI ディープ・ラーニング

Tim Berners-Lee 2016 World Wide Web

ディズニー・アニメーション・スタジオ&ピクサーの元社長

Edwin Catmull 2019 アニメーションと3Dグラフィックス

トイ・ストーリー,モンスターズ・インク

Yann LeCun 2018 AI ディープ・ラーニング

Whitfield Diffie 2015 公開鍵暗号

Pat Hanrahan 2019 アニメーションと3Dグラフィックス

カリフォルニア大バークレー名誉教授, ACM元会長

David A Patterson 2017 コンピュータ構成法 RISC:スマホ-スパコン

Martin Hellman 2015 公開鍵暗号

チューリング賞受賞記念講演が開催されるコンピュータ・アーキテクチャの世界最高峰国際会議　ACM/IEEE ISCAの2025年6月21日−25日 早稲田開催が決定

Co-Chairs:Jean-Luc Gaudiot (Prof. UCI, IEEE CS President 2017)
Hironori Kasahara (SEVP Waseda, IEEE CS President2018)

ACM/IEEE International Symposium on Computer Architecture にてコンピュータ分野のノーベル賞と言われるチューリング賞記念講演会を早稲田大学大隈講堂で実施予定

前回の日本開催：　ISCA 1986: Tokyo, Japan

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

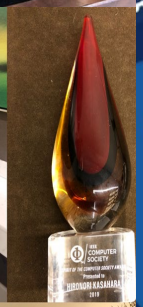
215
International Conferences

12 Magazines

35 Journals

47 Total Publications

847,000+
Articles in CSDL



6
New Standards

230
Active Standards

IEEE754, 802

12,000+
Volunteers

615
Committees/ Boards

2,352+
Meetings/ Teleconferences

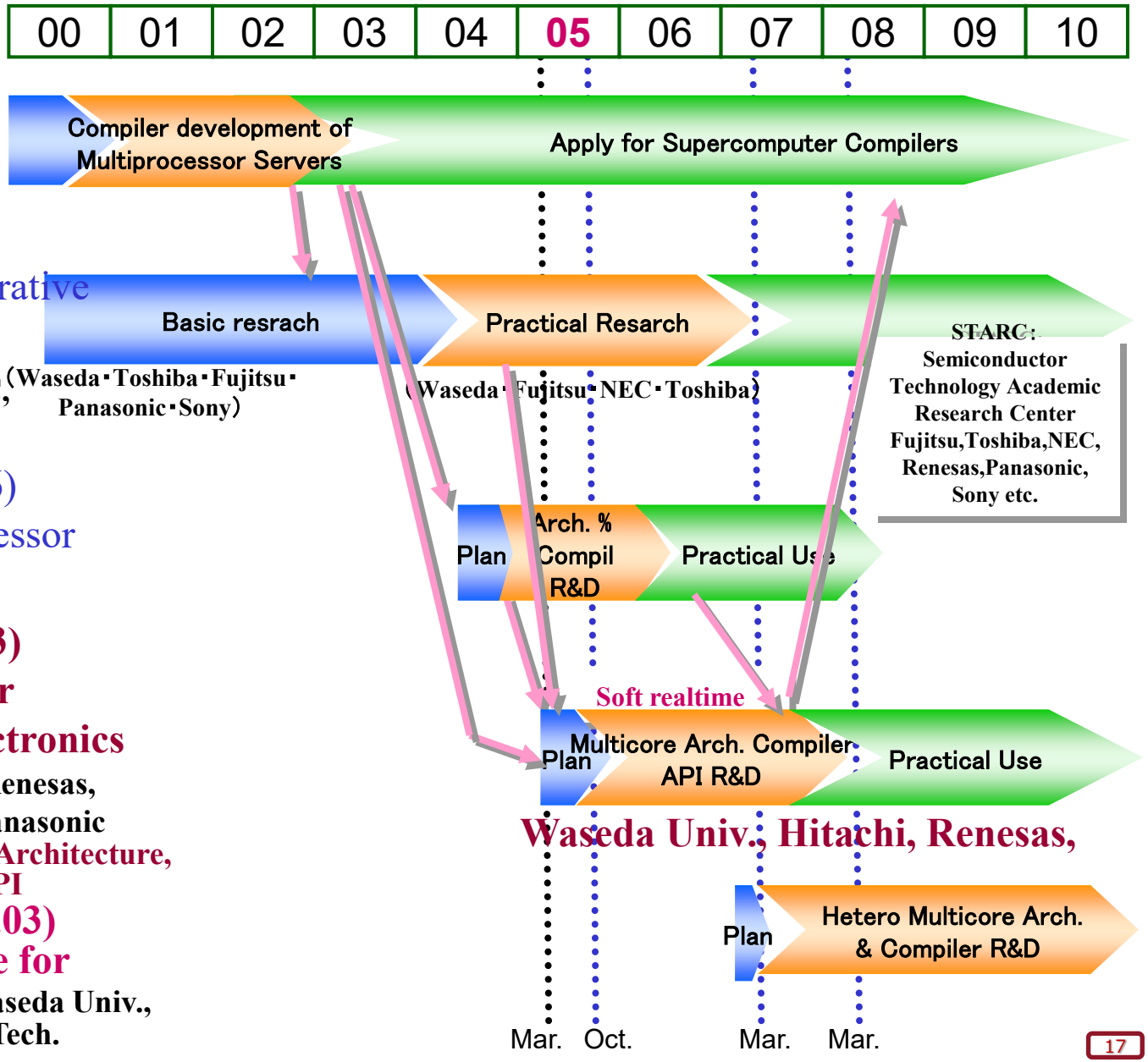
373,100+
Community Members

168
Countries with CS Members

634
Chapters

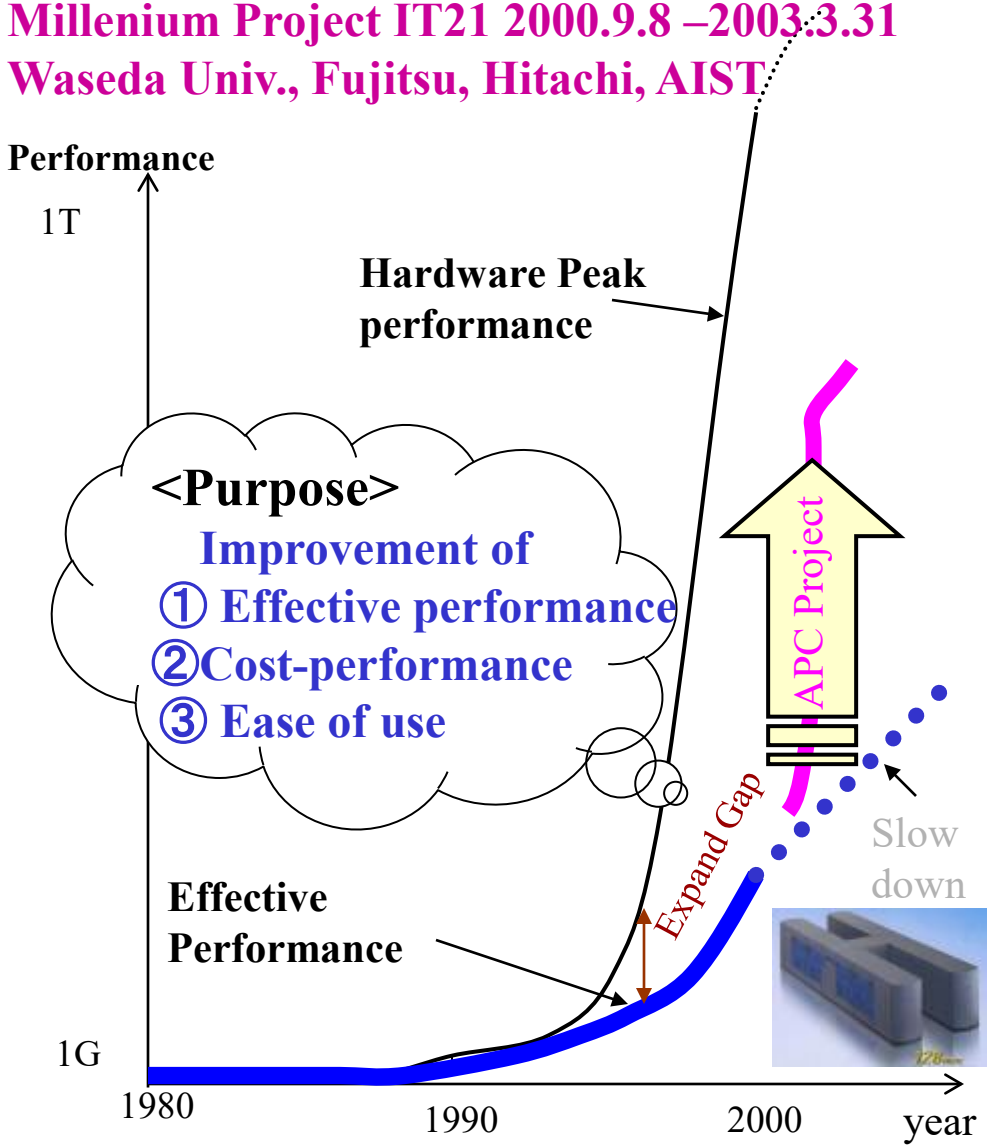
Roadmap of compiler cooperative multicore project

- Millennium Project IT21 NEDO Advanced Parallelizing Compiler
  (Waseda Univ. Fujitsu,Hitachi, JIPDEC,AIST)
- STARC Compiler Cooperative Chip Multiprocessor
  (Waseda Univ., Fujitsu, NEC, Toshiba, Panasonic,Sony)
- NEDO (2004.07-2007.06) Heterogeneous Multiprocessor
  (Waseda Univ., Hitachi)
- **NEDO (2005.06-2008.03) Multicore Technology for Realtime Consumer Electronics**
  - Waseda Univ., Hitachi, Renesas, Fujitsu, NEC, Toshiba, Panasonic
    - ➢Power Saving Multicore Architecture, Parallelizing Compiler,API
- **NEDO (2007.02- 2010.03) Heterogeneous Multicore for Consumer Electronics** Waseda Univ., Hitachi, Renesas, Tokyo Inst, of Tech.



METI/NEDO Advanced Parallelizing Compiler Technology Project

Millenium Project IT21 2000.9.8 –2003.3.31
Waseda Univ., Fujitsu, Hitachi, AIST



Theoretical maximum performance vs. Effective performance of HPC

Background and Problems

①Adoption of parallel processing as a core technology on PC to HPC
② Increase of importance of software on IT
③ Need for improvement of cost-performance and usability

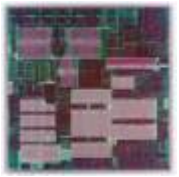
Contents of Research and Development

① R & D of advanced parallelizing compiler
Multigrain, Data localization, Overhead hiding
② R & D of Performance evaluation technology for parallelizing compilers

Goal: Double the effective performance

Ripple Effect

① Development of competitive next generation PC and HPC
② Putting the innovative automatic parallelizing compiler technology to practical use
③ Development and market acquisition of future single-chip multiprocessors
④ Boosting R&D in the following many fields:
IT, Bio-tech., Device, Earth environment, Next-generation VLSI design, Financial engineering, Weather forecast, New clean energy, Space development, Automobile, Electric Commerce, etc



世界をリードするマルチコア用コンパイラ技術

OSCARコンパイラの世界唯一技術

1.マルチグレイン並列化（全ての並列性を利用）

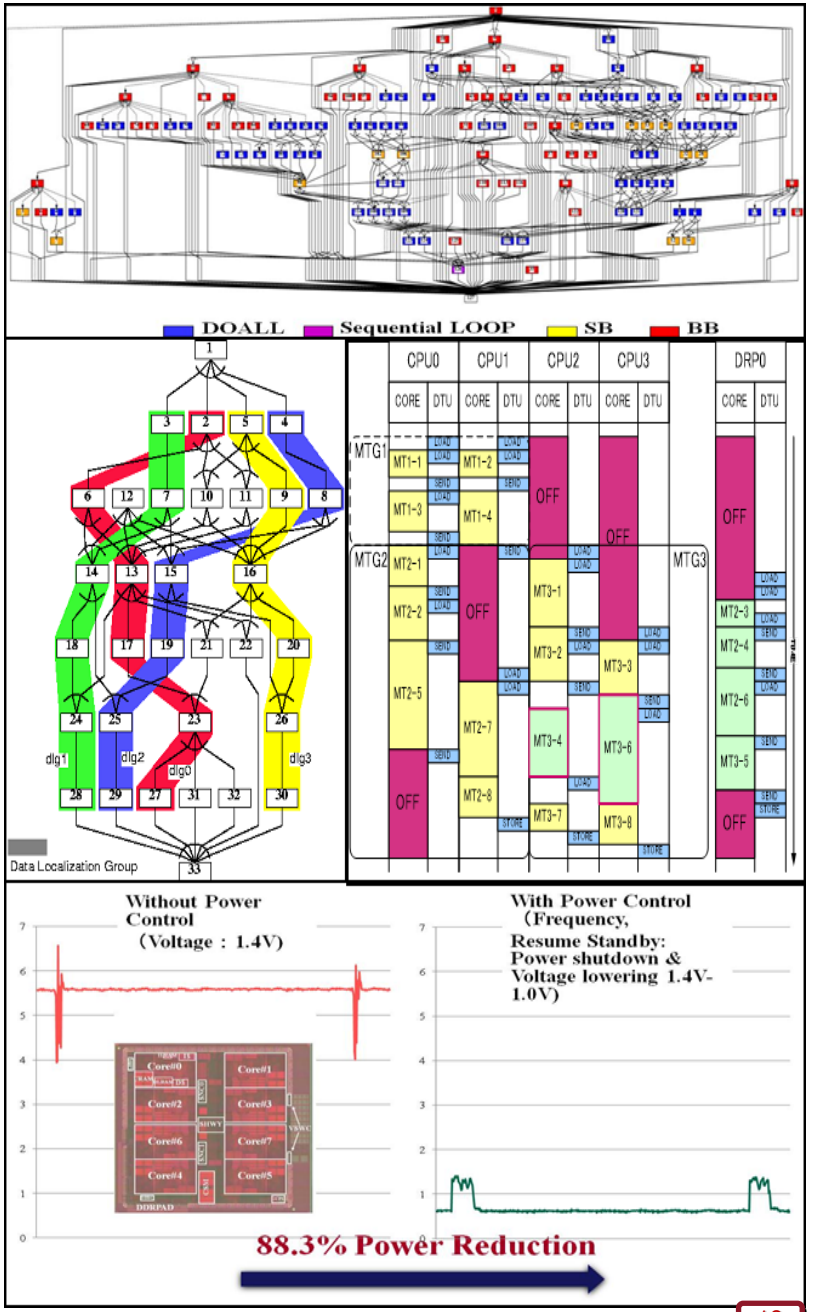
- 粗粒度タスク並列化、ループ並列化、近細粒度並列化によりプログラム全域の並列性を利用するマルチグレイン並列化機能により、従来の命令レベル並列性より大きな並列性を抽出し、複数マルチコアで速度向上

2.プログラム全域にわたるメモリ利用最適化

- コンパイラによるローカルメモリへのデータ分割配置、DMAコントローラによるタスク実行とオーバーラップしたデータ転送によりメモリアクセス・データ転送オーバーヘッド最小化

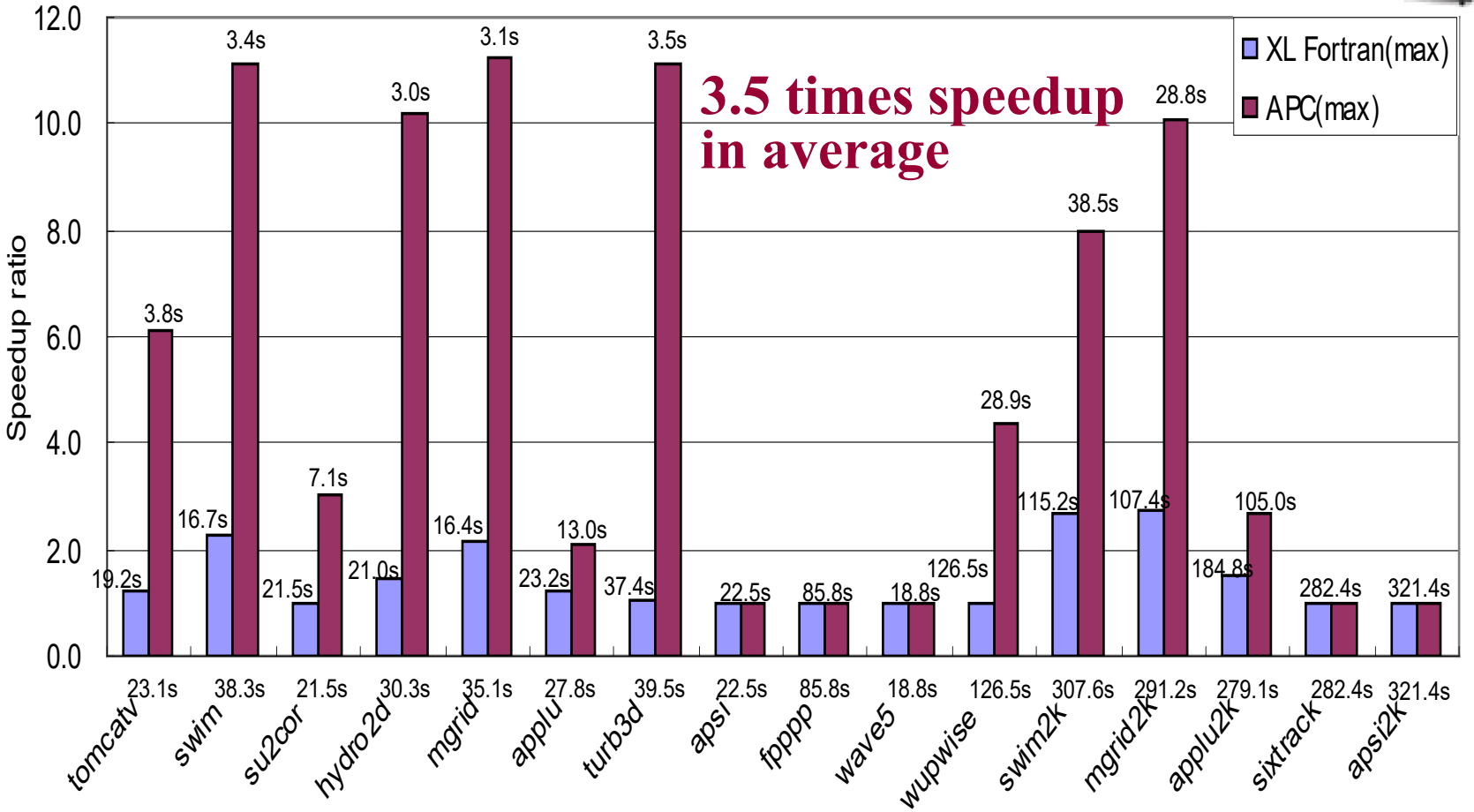
3.プロセッサ・メモリ・ネットワーク等の停止・動作速度制御による自動省エネ

- コンパイラによる低消費電力制御機能を用いたアプリケーション内でのきめ細かい周波数・電圧制御・電源遮断により消費電力低減

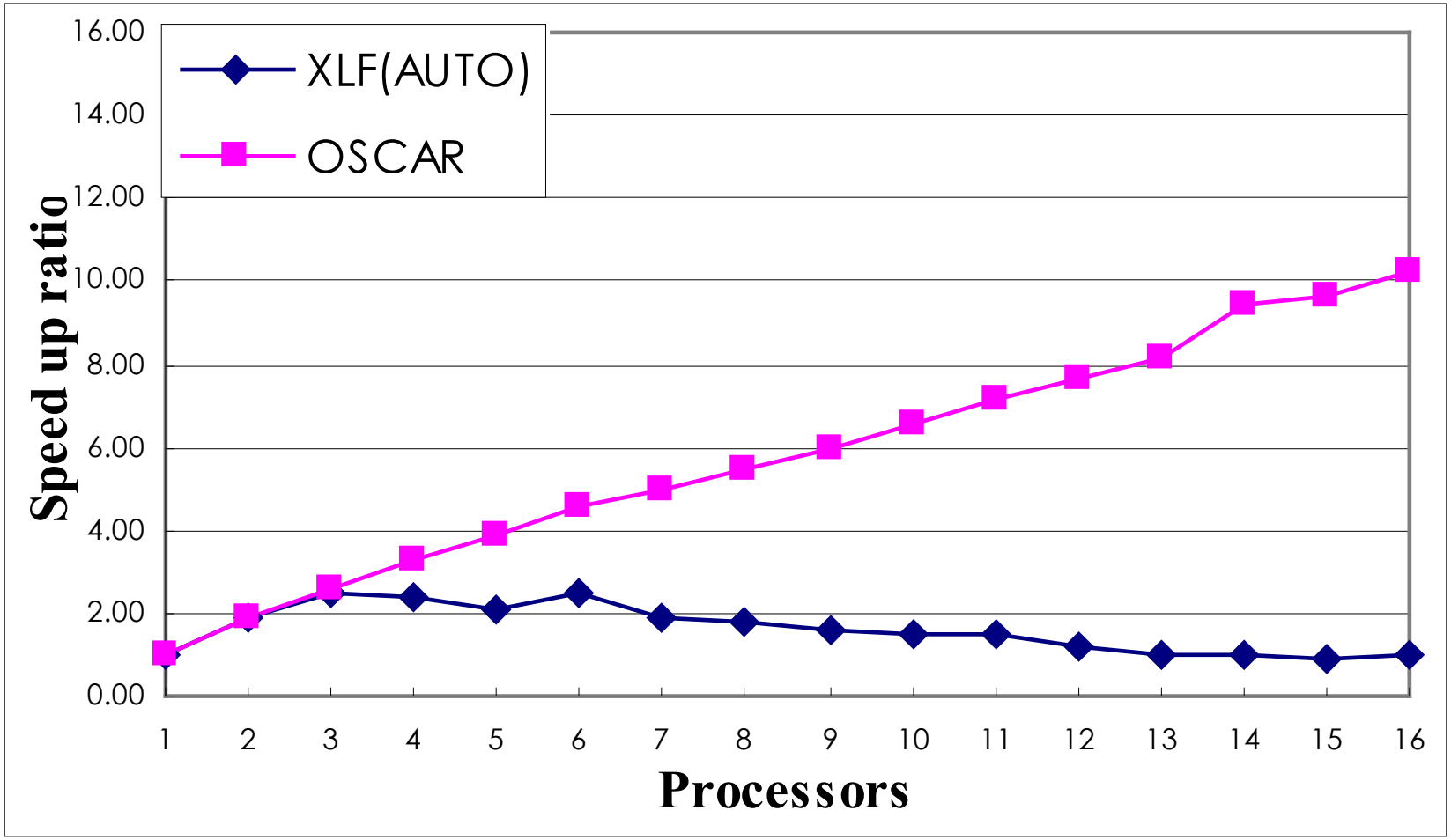


Performance of APC Compiler on IBM pSeries690 16 Processors High-end Server

- 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)

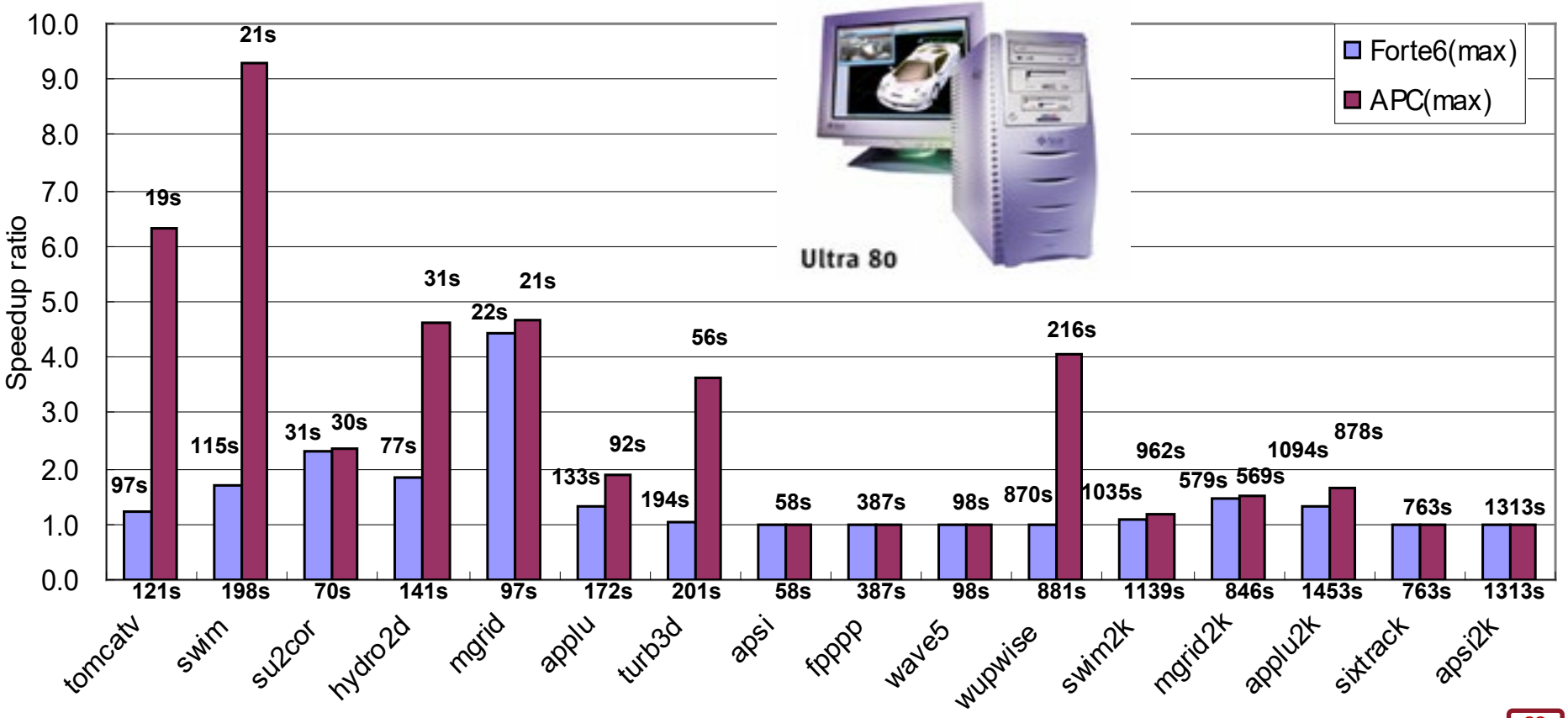


Performance of Multigrain Parallel Processing for 102.swim on IBM pSeries690

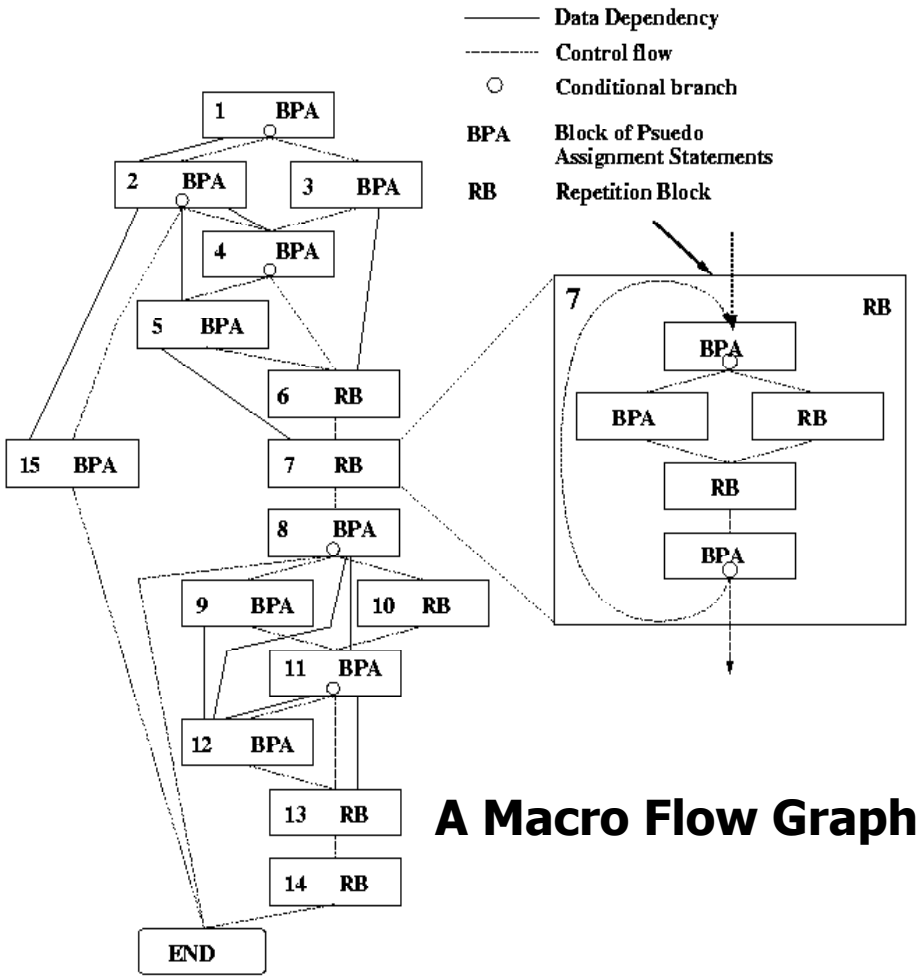


Performance of APC Compiler on Sun Ultra80 4 Processor Workstation

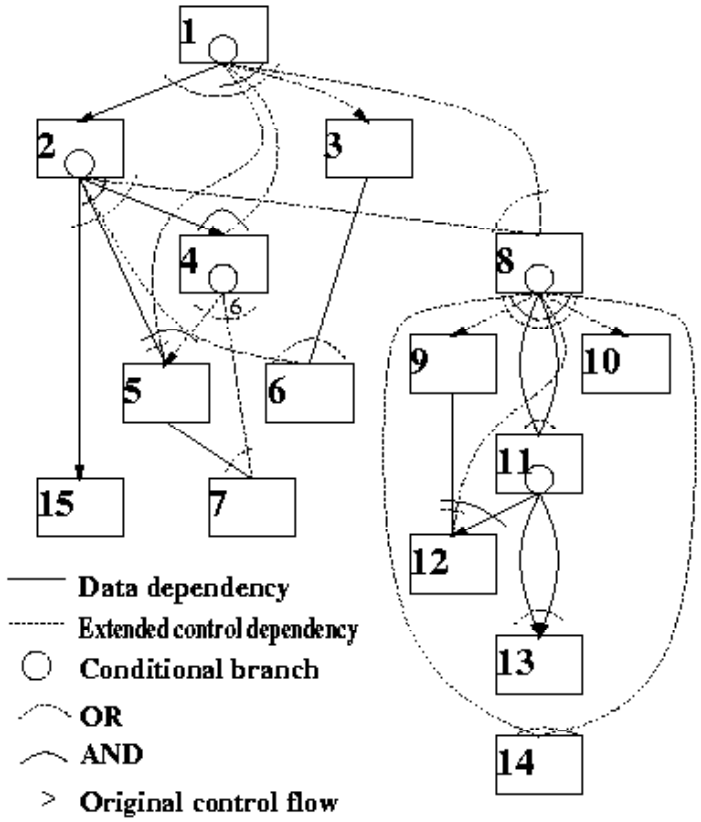
- Sun Forte Developer 6 Update 2
  - Sequential execution : -fast
  - Automatic loop parallelization : -fast -autopar -reduction -stackvar
  - OSCAR compiler : -fast -explicitpar -mp=openmp -stackvar



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



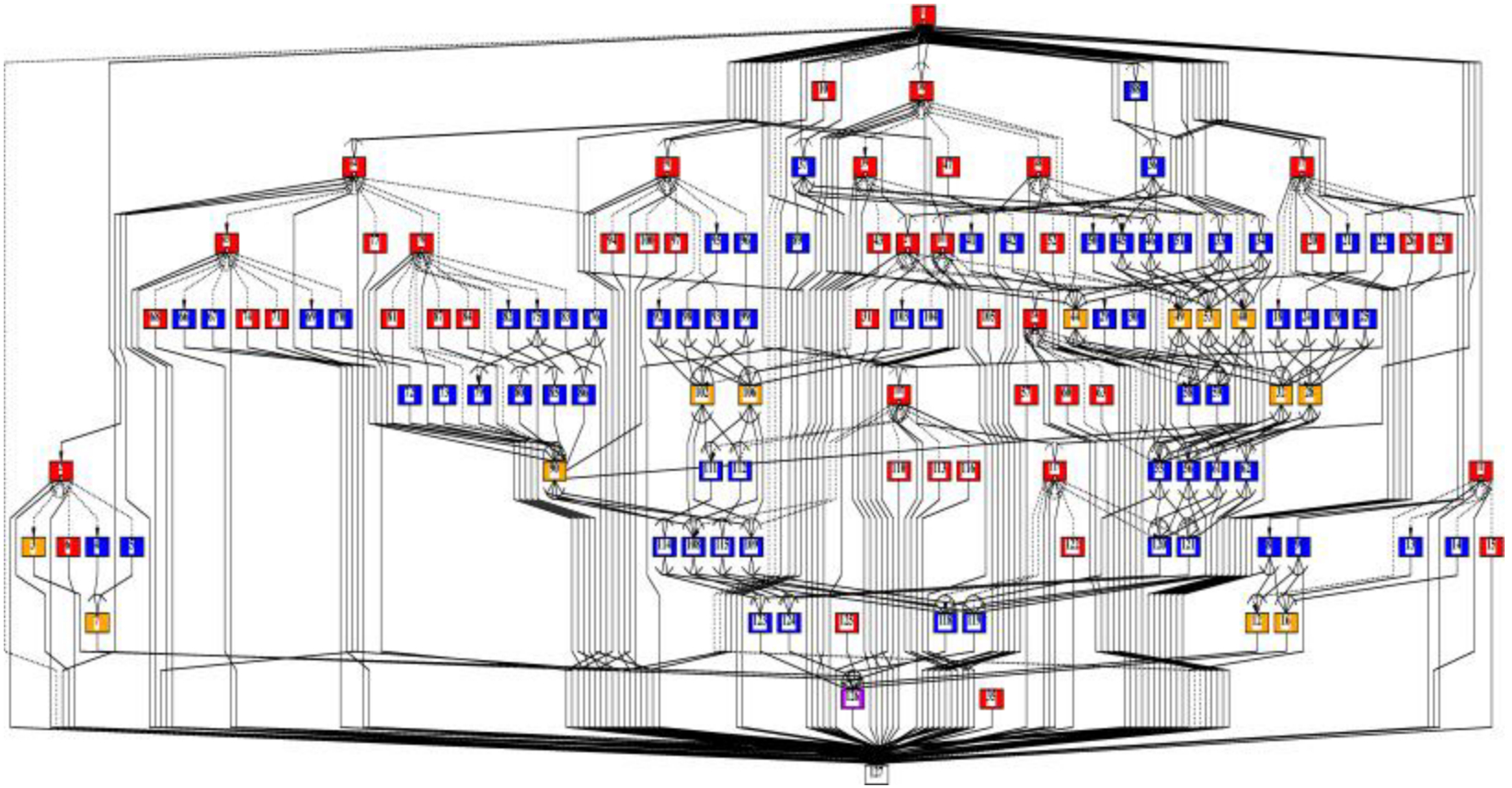
A Macro Flow Graph



A Macro Task Graph

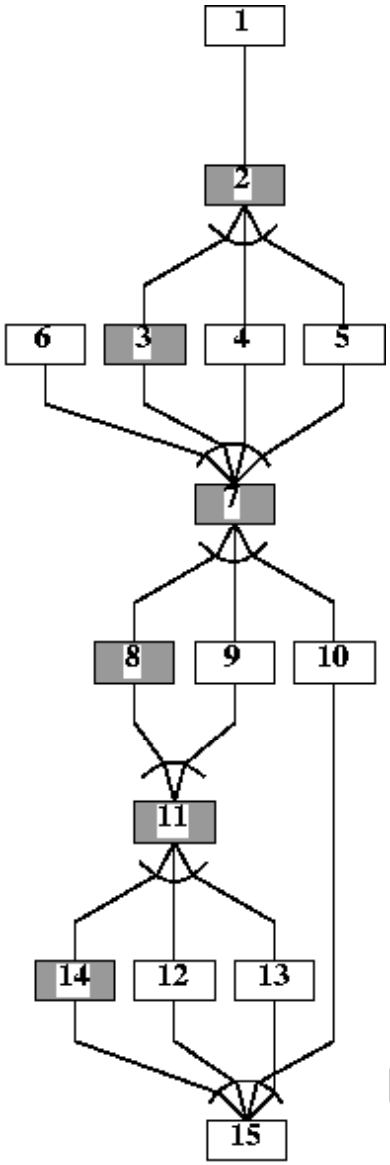
MTG of Su2cor-LOOPS-DO400

- Coarse grain parallelism PARA_ALD = 4.3

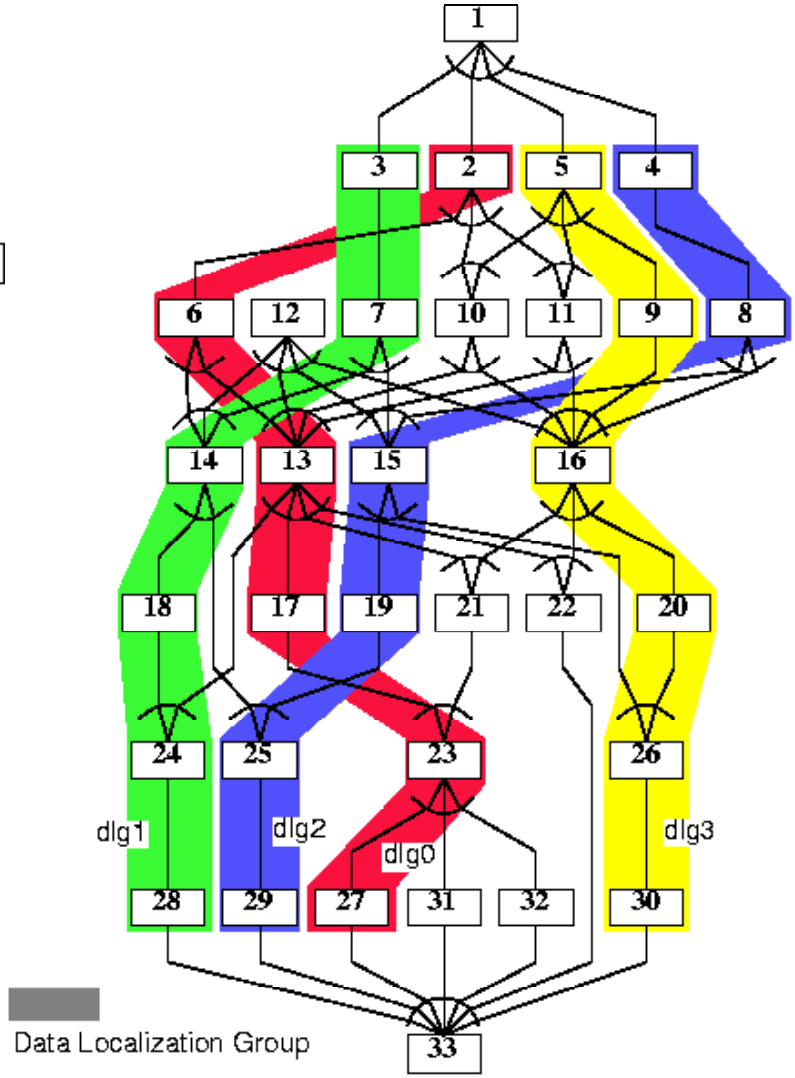


DOALL Sequential LOOP SB BB

Data Localization



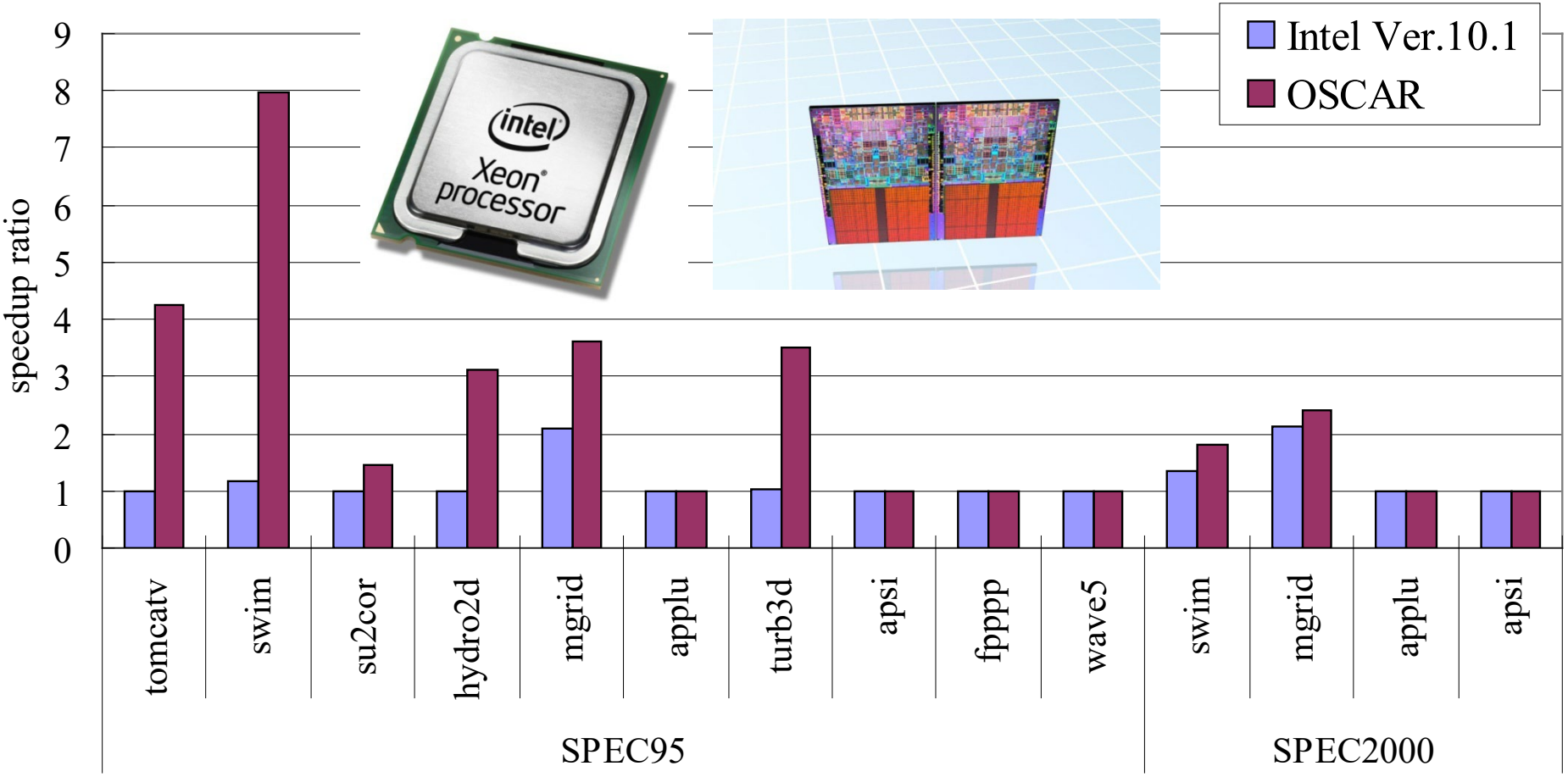
MTG



MTG after Division

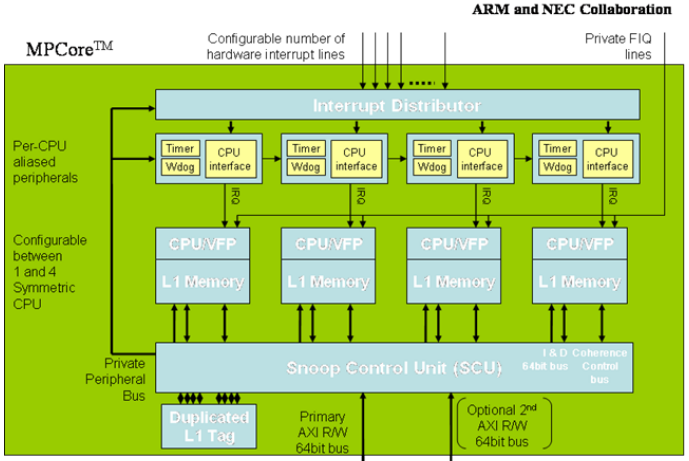
A schedule for two processors

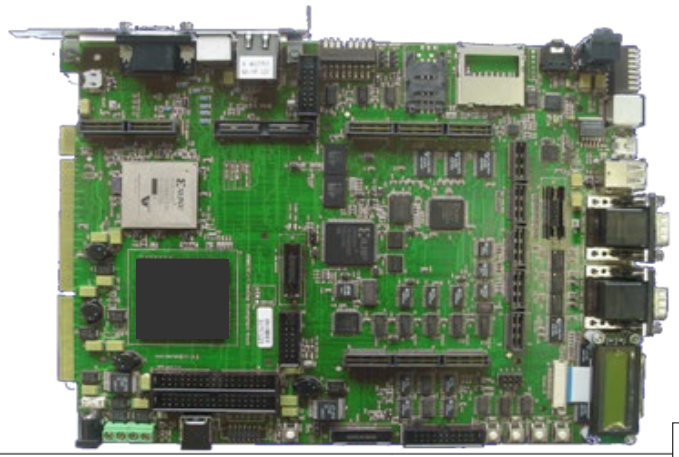
Performance of OSCAR Compiler Using the Multicore API on Intel Quad-core Xeon

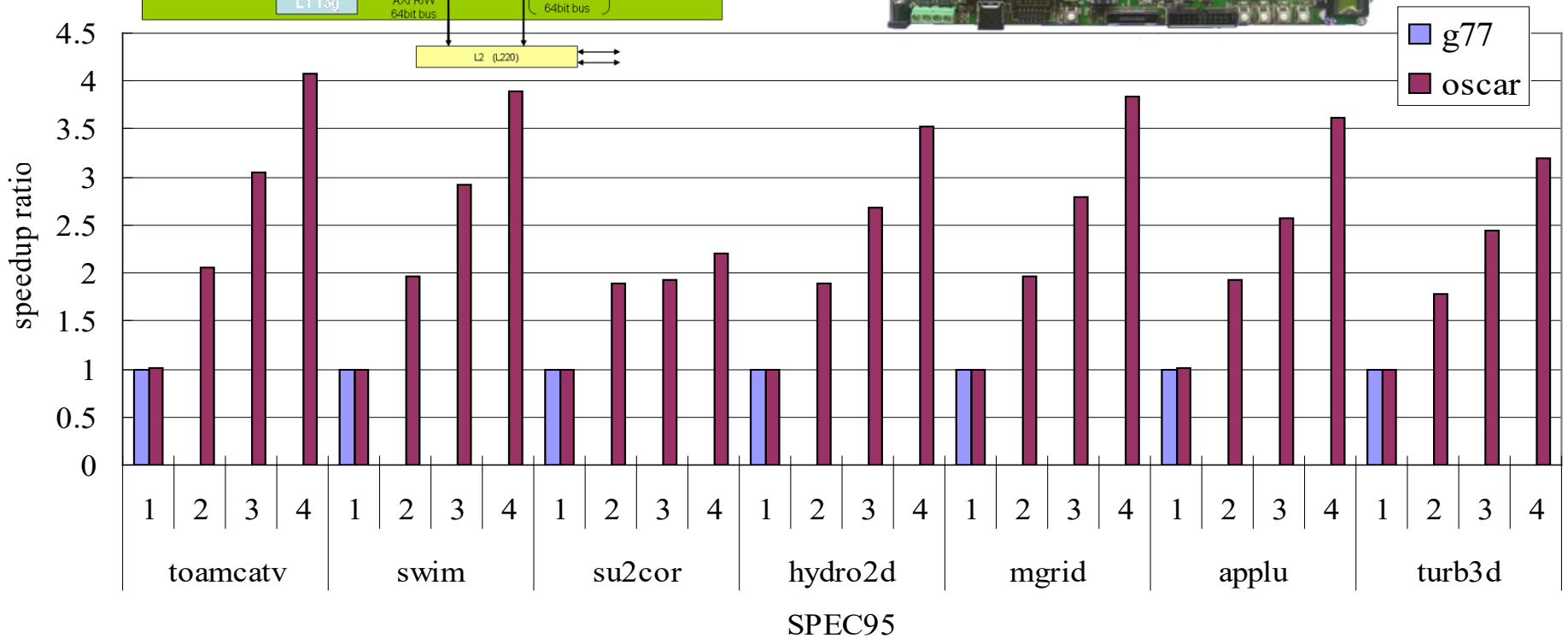


- **OSCAR Compiler gives us 2.09 times speedup on the average against Intel Compiler ver.10.1**

NEC/ARM MPCore Embedded 4 core SMP





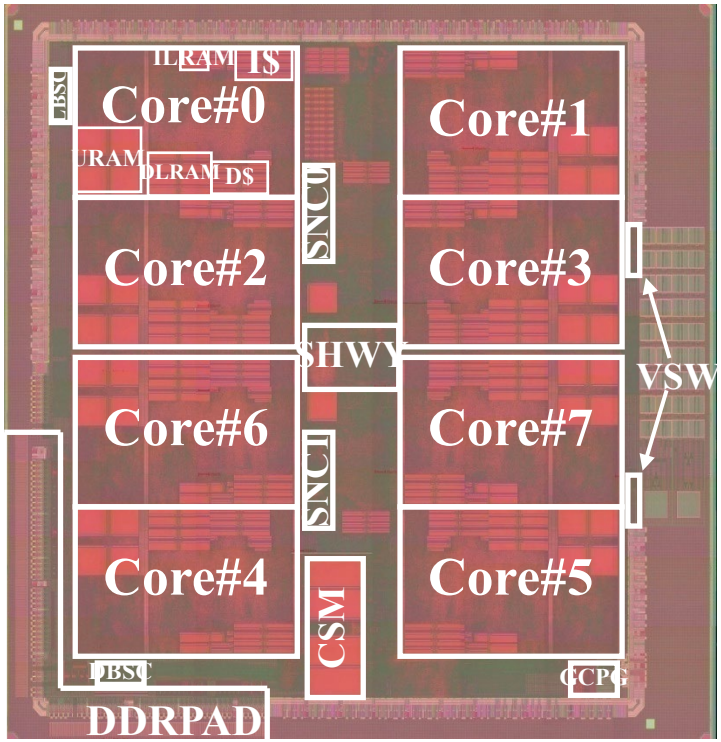


3.48 times speedup by OSCAR compiler against sequential processing

ムーアの法則の終焉

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「半導体の集積率は18か月で2倍になる」という経験則。

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"

$$\text{Power} \propto \text{Frequency} * \text{Voltage}^2$$

(Voltage ∝ Frequency)

$$\text{Power} \propto \text{Frequency}^3$$

周波数 Frequency を1/4にすると

(Ex. 4GHz→1GHz),

消費電力は 1/64に削減

性能は 1/4に低下 .

<マルチコア>

8cores をチップに集積すると,

電力は 依然1/8で性能 は2 倍向上

METI/NEDO National Project

Multi-core for Real-time Consumer Electronics

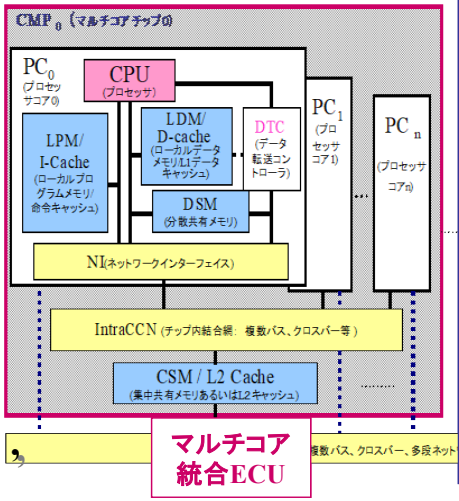
(2005.7～2008.3)**

<Goal> R&D of compiler cooperative multi-core processor technology for consumer electronics like Mobile phones, Games, DVD, Digital TV, Car navigation systems.

<Period> From July 2005 to March 2008

<Features>
- Good cost performance
- Short hardware and software development periods
- Low power consumption
- Scalable performance improvement with the advancement of semiconductor
- Use of the same parallelizing compiler for multi-cores from different vendors using newly developed API

API : Application Programming Interface

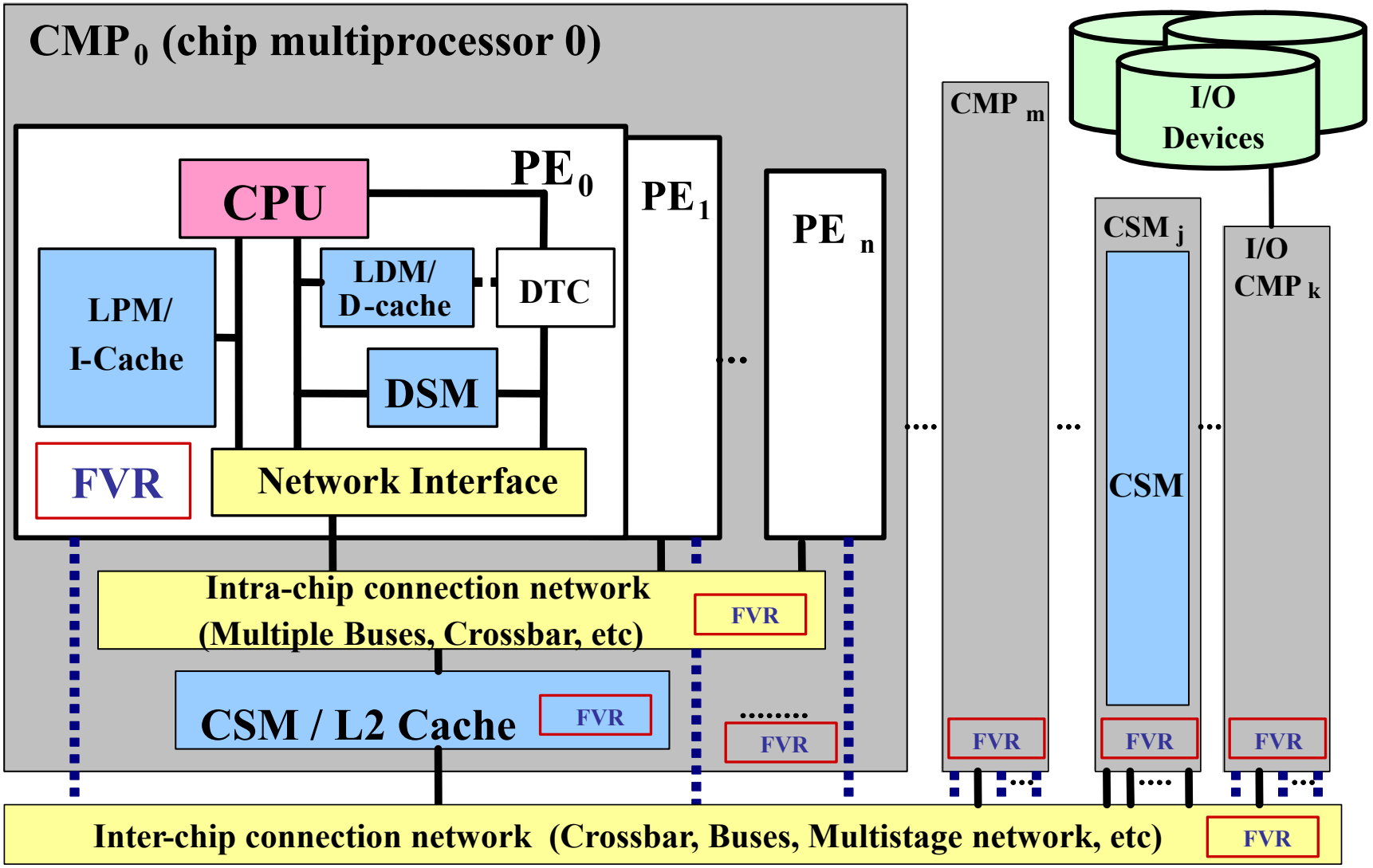


新マルチコアプロセッサ
- 高性能
- 低消費電力
- 短HW/SW開発期間
- 各チップ間でアプリケーション共用可
- 高信頼性
- 半導体集積度と共に性能向上

開発マルチコアチップは情報家電へ

**Hitachi, Renesas, Fujitsu, Toshiba, Panasonic, NEC

OSCAR Multi-Core Architecture



CSM: central shared mem.

DSM: distributed shared mem.

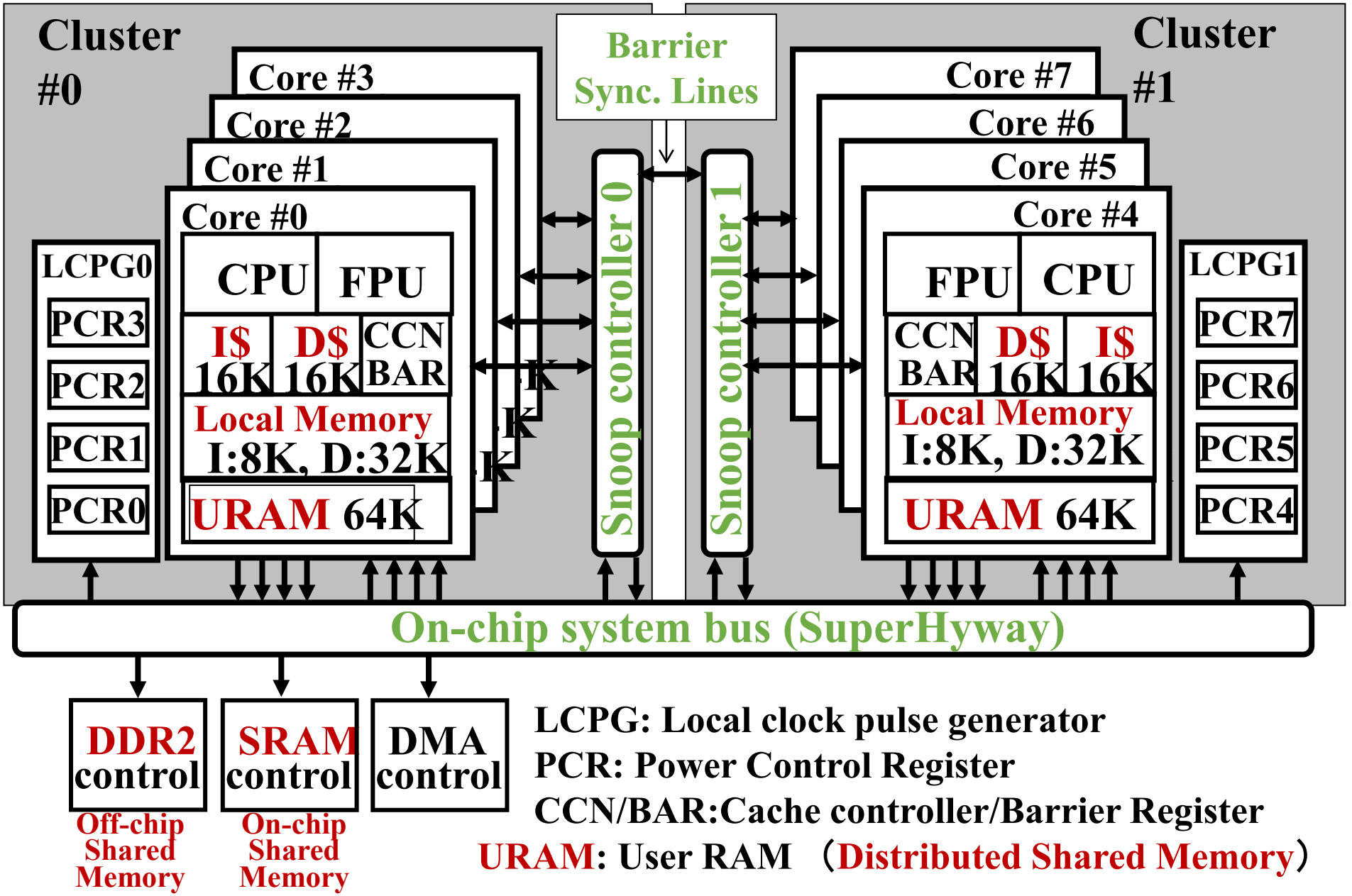
DTC: Data Transfer Controller

LDM : local data mem.

LPM : local program mem.

FVR: frequency / voltage control register

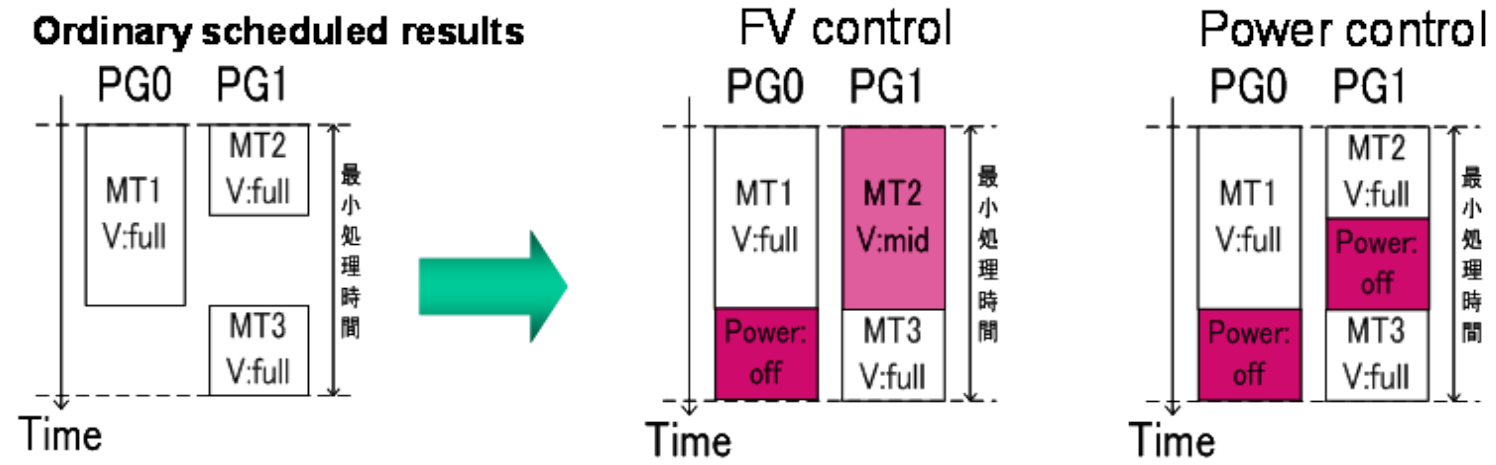
8 Core RP2 Chip Block Diagram



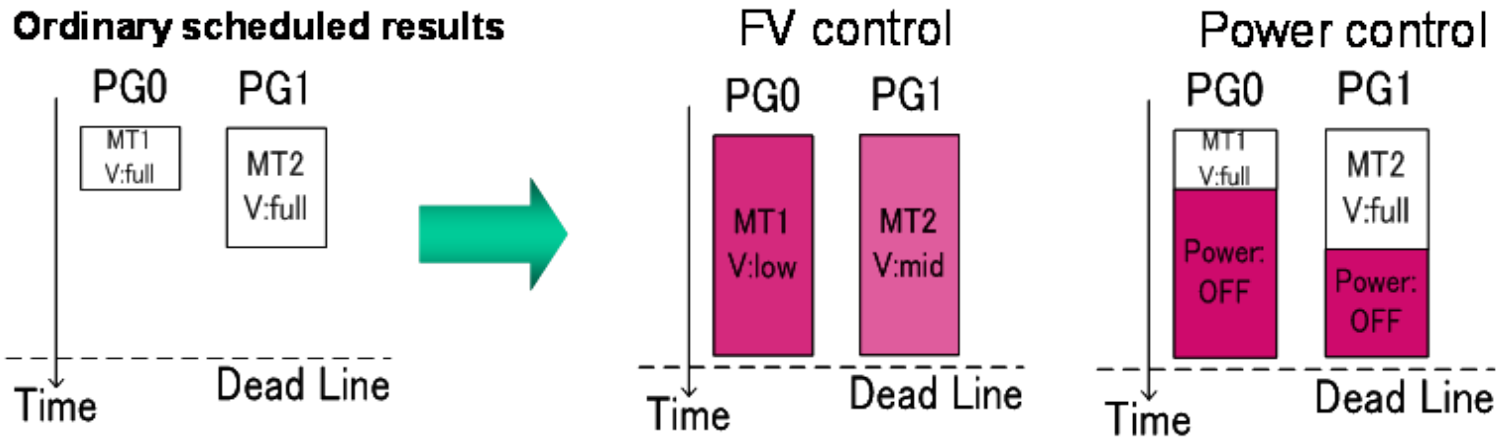
LCPG: Local clock pulse generator
PCR: Power Control Register
CCN/BAR:Cache controller/Barrier Register
URAM: User RAM （Distributed Shared Memory）

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

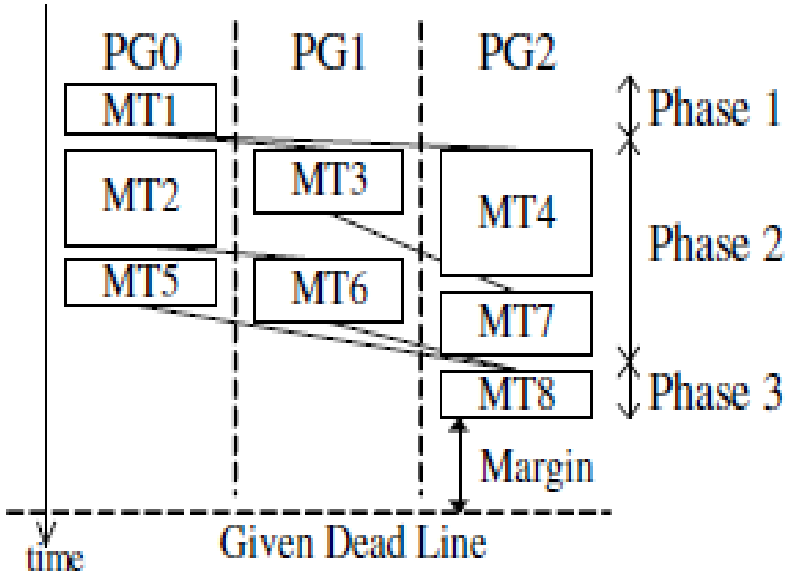
- Shortest execution time mode

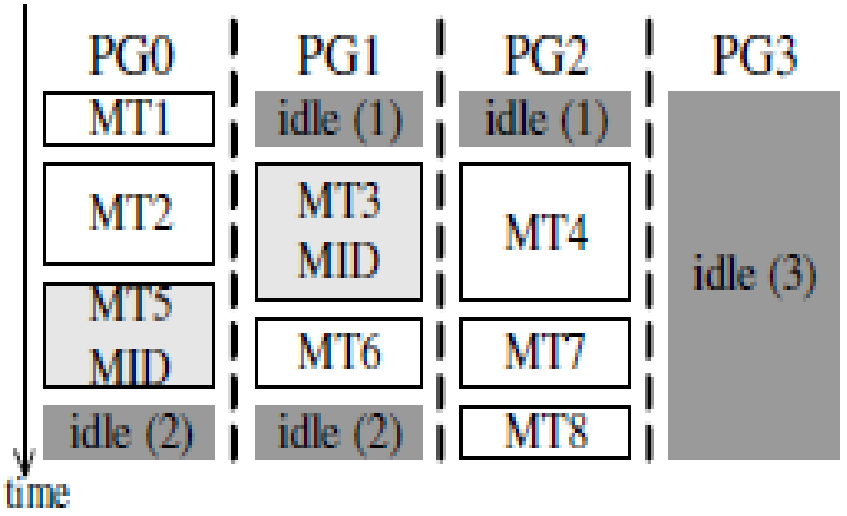


- Realtime processing mode with dead line constraints



Power Reduction Scheduling





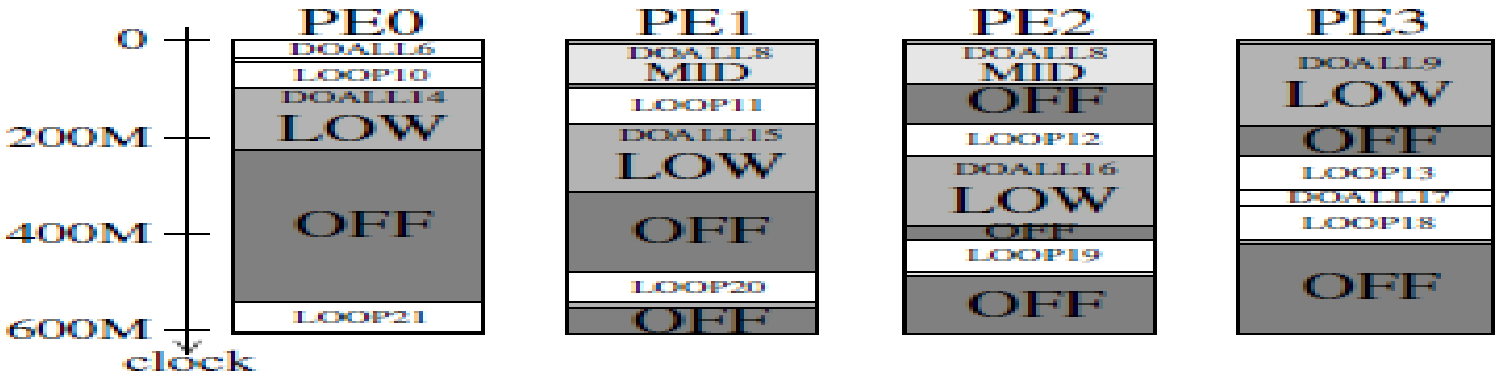


Fig. 6. V/F control of applu(4proc.)

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]

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

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

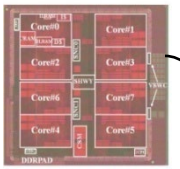
Accelerator 2
Code

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

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

Server Code Generation
OpenMP Compiler

Generation of parallel machine codes using sequential compilers



Homegeneous Multicore s from Vendor A (SMP servers)

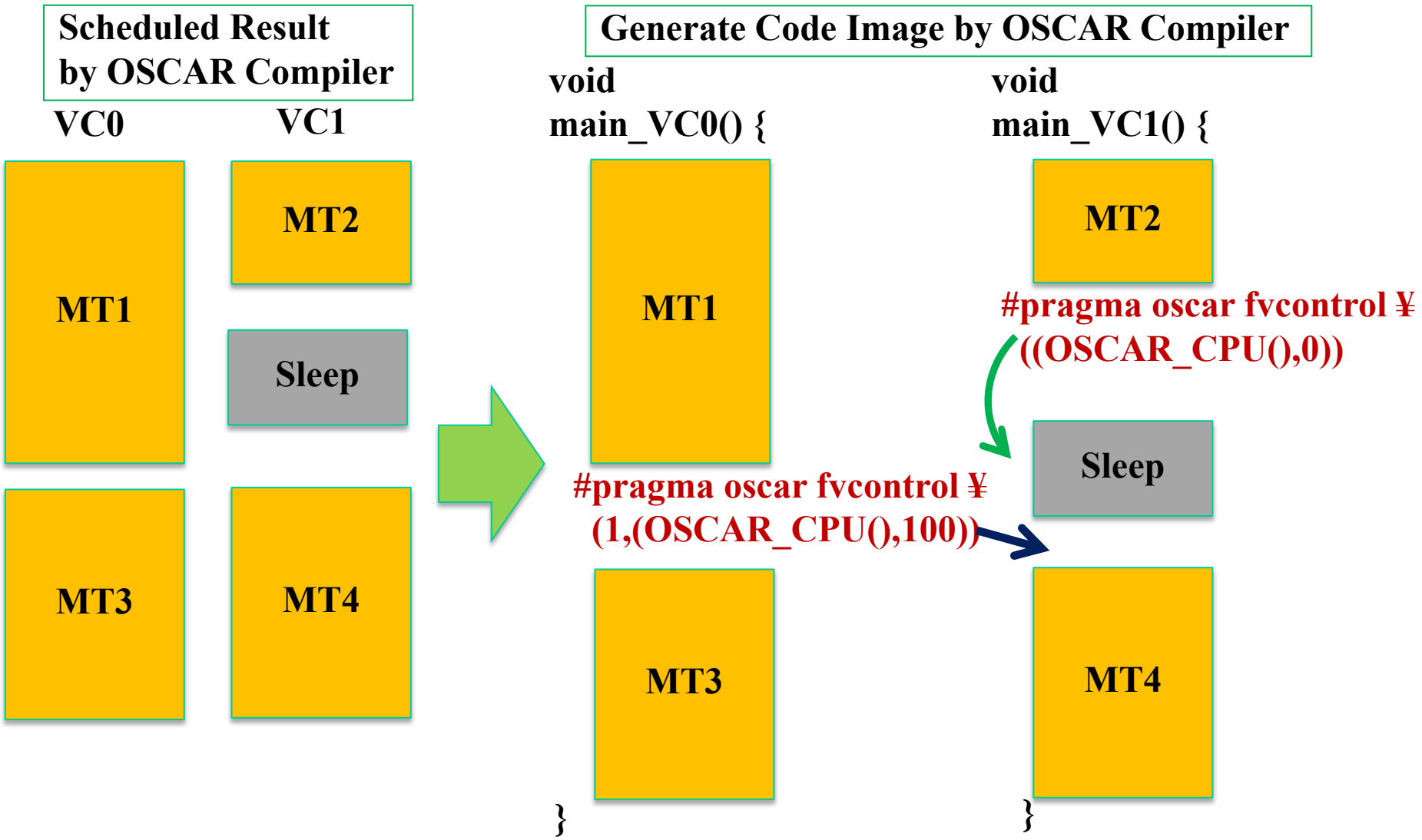
Heterogeneous Multicores from Vendor B

Shred memory servers

Executable on various multicores

OSCAR: Optimally Scheduled Advanced Multiprocessor
API : Application Program Interface

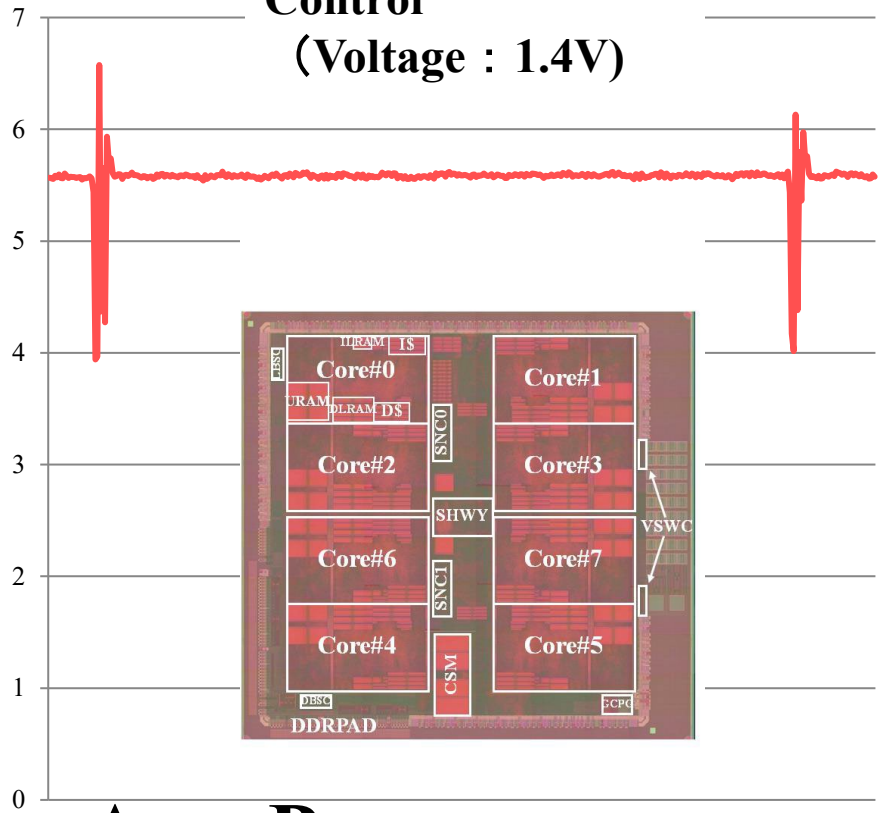
Low-Power Optimization with OSCAR API



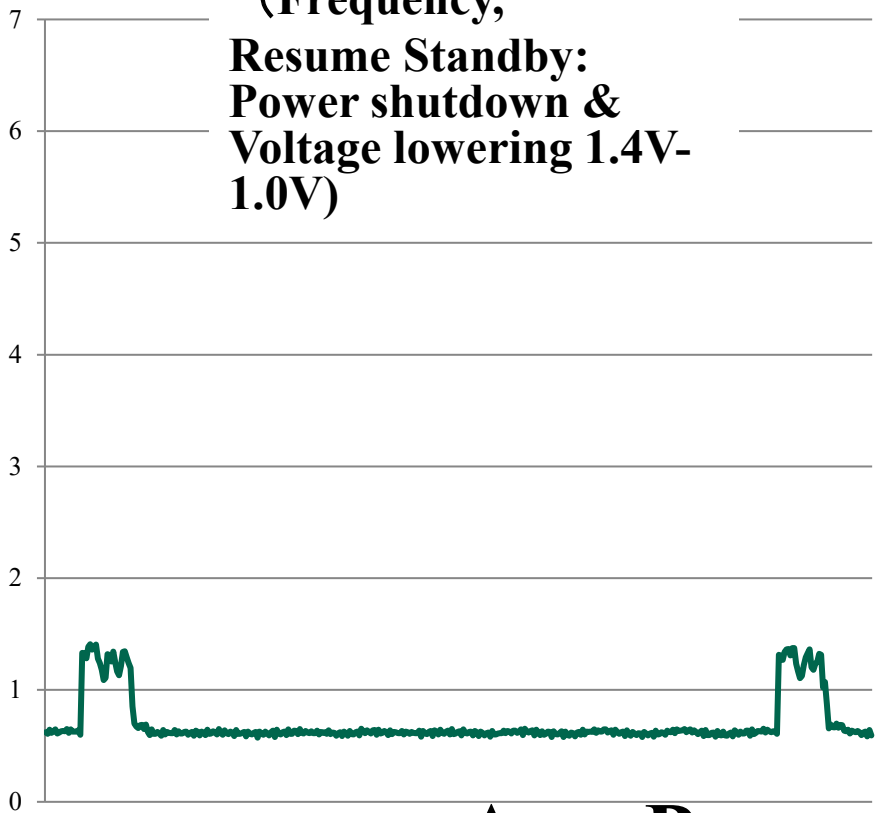
Power Reduction by OSCAR Parallelizing Compiler for Secure Audio Encoding

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

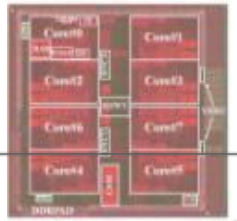
Avg. Power 0.67 [W]

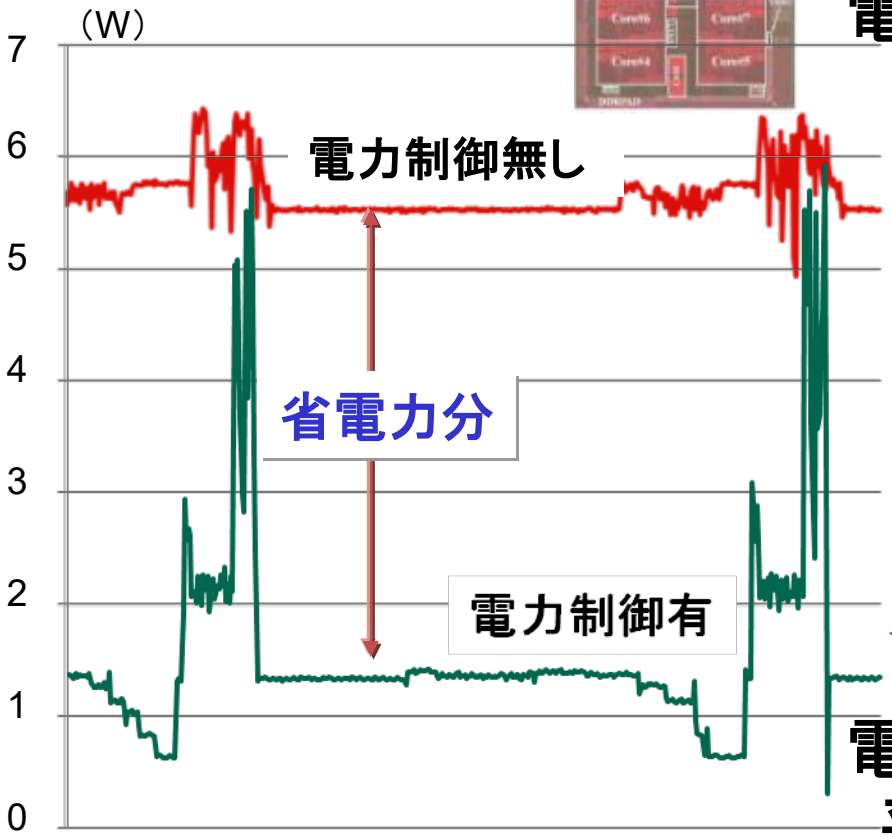
太陽光電力で動作する情報機器

コンピュータの消費電力をHW&SW協調で低減。電源喪失時でも動作することが可能。

リアルタイムMPEG2デコードを、8コアホモジニアス
マルチコアRP2上で、消費電力1/4に削減

世界唯一の差別化技術



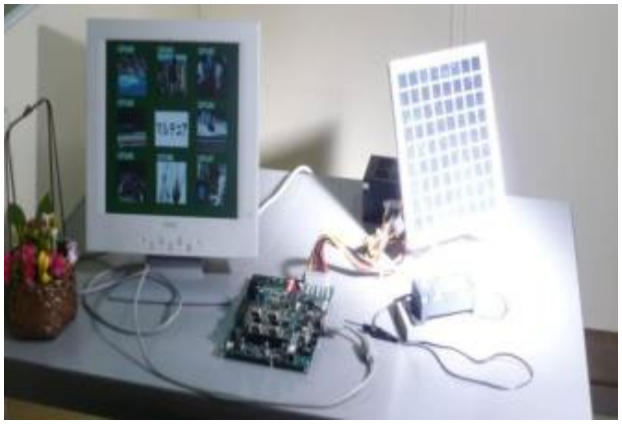


電力制御無し
平均電力
5.73 [W]

電力を
ソフトで
1/4に
削減

電力制御有
平均電力
1.52 [W]

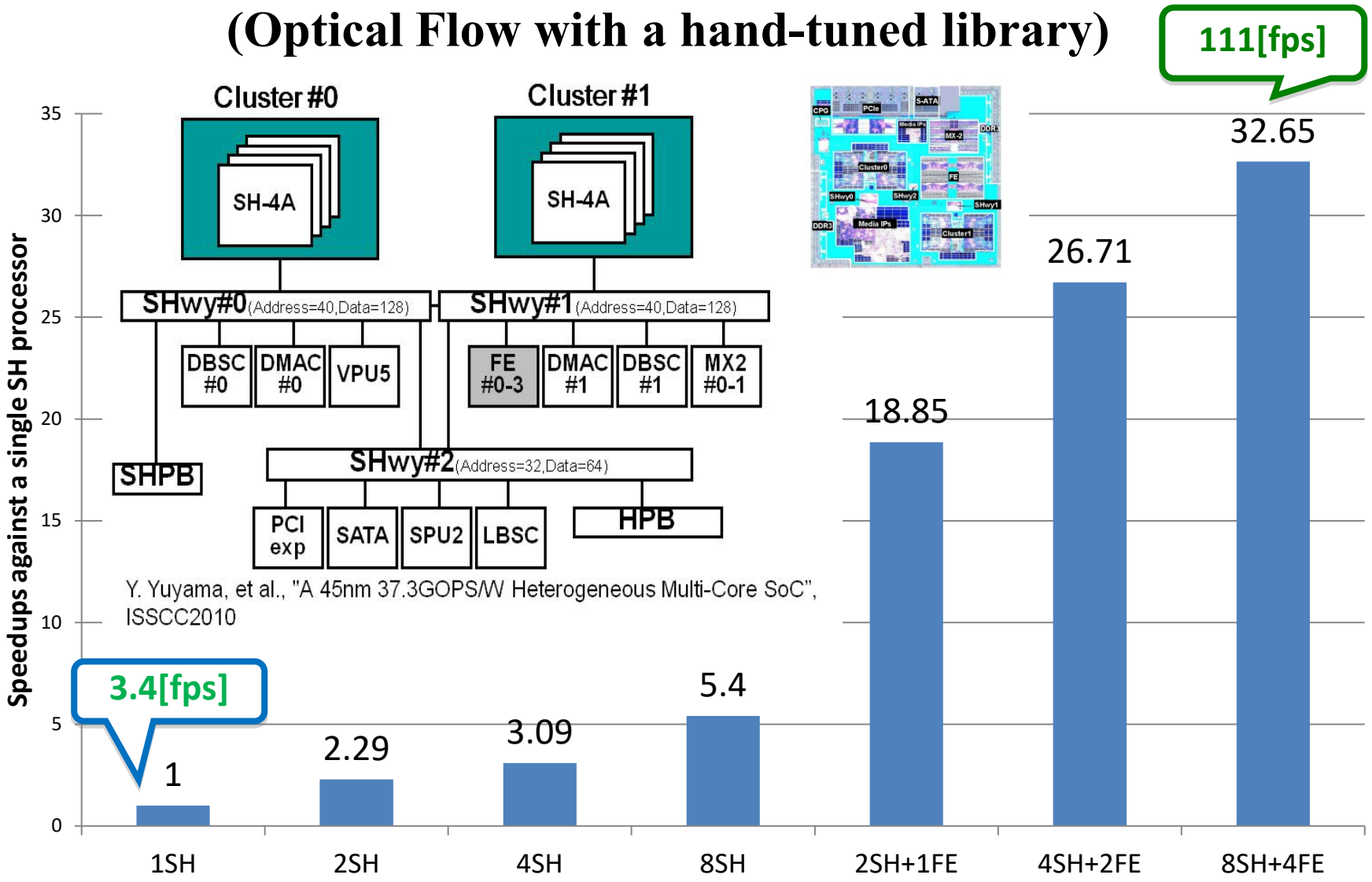
周波数/電圧・電源遮断制御



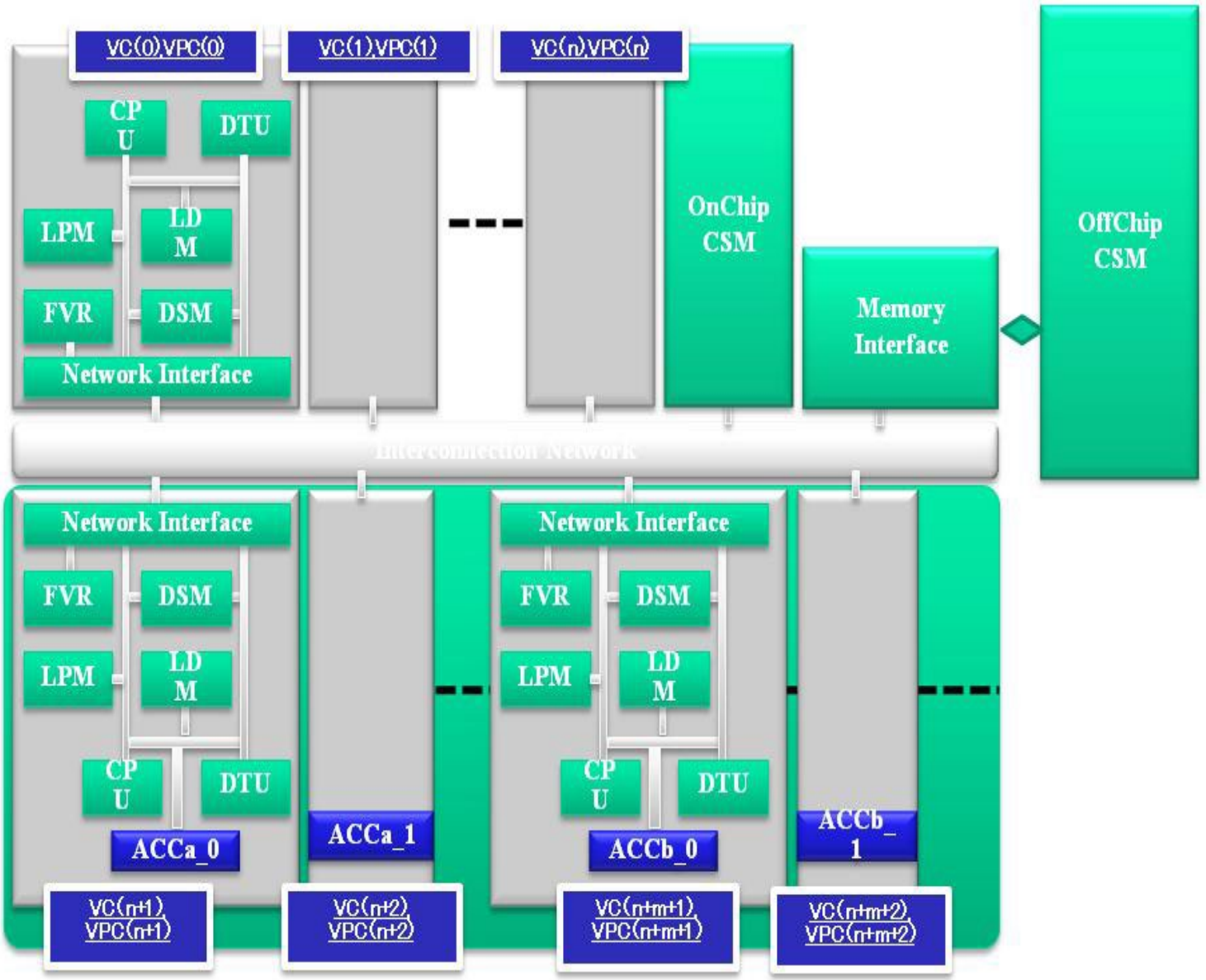
太陽電池で駆動可

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

(Optical Flow with a hand-tuned library)



OSCAR Heterogeneous Multicore



DTU
- Data Transfer Unit

LPM
- Local Program Memory

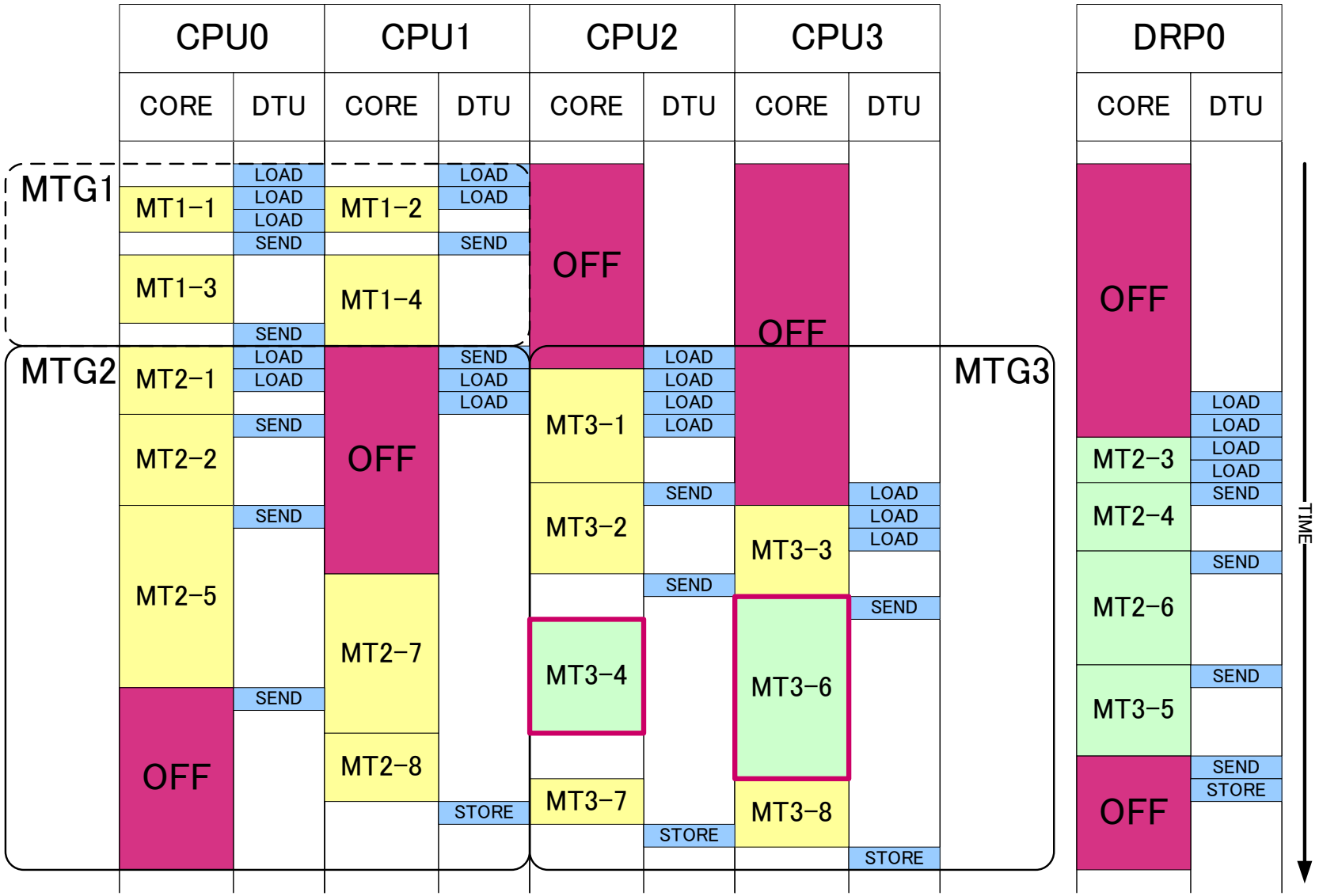
LDM
- Local Data Memory

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



総合科学技術会議（平成20年4月10日）での NEDOリアルタイム情報家電用マルチコアチップ（笠原リーダー）・デモの様子

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)

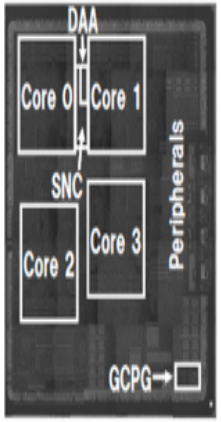
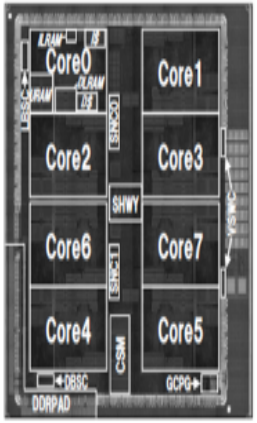
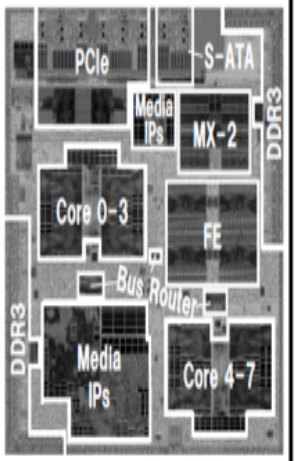
第74回総合科学技術会議の様子(2)

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)

1985年よりコンパイラ（ソフト）・アーキテクチャ（ハード）協調設計マルチプロセッサの研究

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)

実施場所：グリーン・コンピューティング・システム研究開発センター

2011年4月13日竣工，2011年5月13日開所

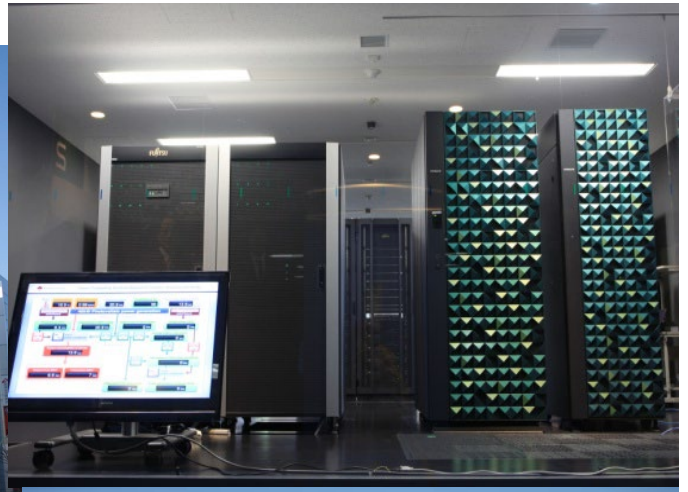
経済産業省「2009年度産業技術研究開発施設整備費補助金」
先端イノベーション拠点整備事業

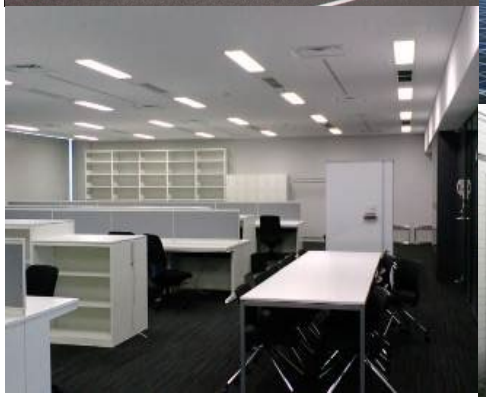
＜目標＞
太陽電池で駆動可能で
冷却ファンが不要な
超低消費電力・高性能マルチコア/
メニーコアプロセッサ*のハードウェア、
ソフトウェア、応用技術の研究開発

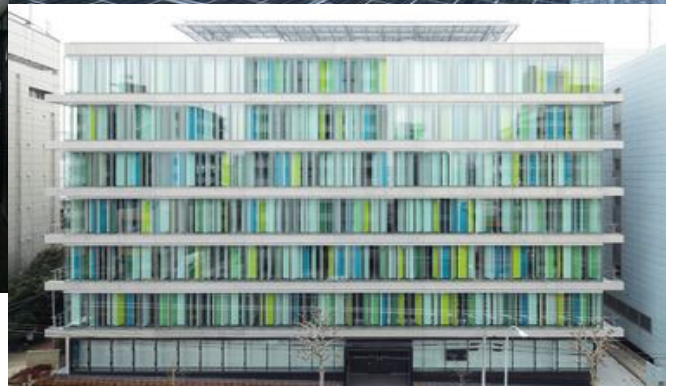
*1チップ上に多数のプロセッサコアを
集積する次世代マルチコアプロセッサ

＜産学連携＞
日立,富士通, ルネサス,NEC, トヨタ,
デンソー，オリンパス,NSITEX、三菱電機,
オスカーテクノロジ等

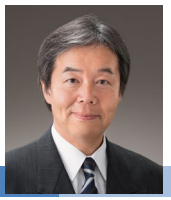
＜波及効果＞
超低消費電力メニーコア
- CO_2排出量削減
- サーバ国際競争力強化
- 我が国の産業利益を支える
  情報家電,自動車等の高付加価値化







グリーン・コンピューティング： 環境に優しい低消費電力・高性能計算



笠原博徳

制御

交通シミュレーション・信号制御 NTTデータ・日立

環境への貢献
カーボンニュートラル

生命・SDGsへの貢献

木村啓二

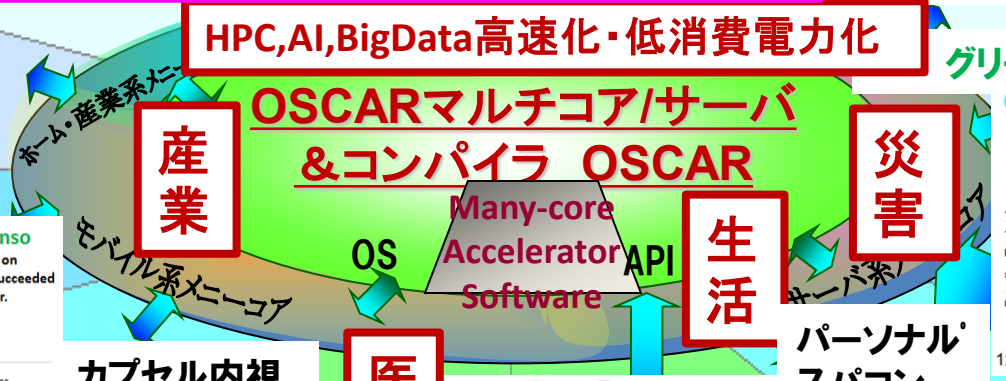
データセンター: 100WM(火力発電所必要)
→ 1000MW=1GW（原子力発電所必要）

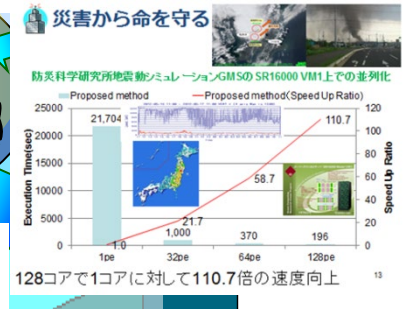
グリーンスパコン

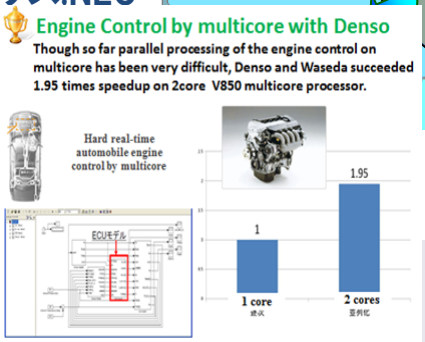
車載(エンジン制御・自動運転Deep Learning・ADAS・MATLAB/Simulink自動並列化) デンソー, ルネサス,NEC

HPC,AI,BigData高速化・低消費電力化

グリーンデータ・クラウドサーバ





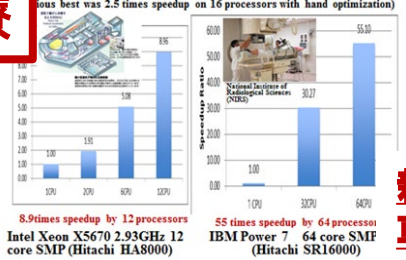


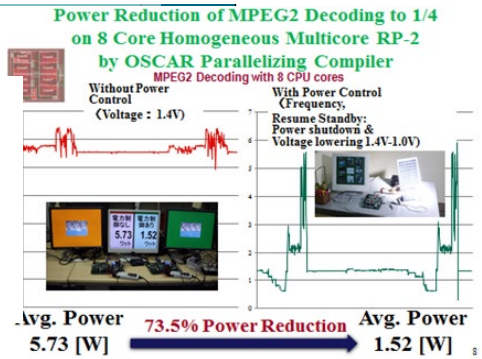
カプセル内視鏡オリンパス

パーソナルスパコン

首都圏直下型地震火災延焼、住民避難指示

スマホ





高信頼・低コスト・ソフト開発

FA 三菱

カメラ

太陽光駆動

重粒子ガン治療日立

新幹線車体設計・ディープラーニング・日立

世界の人々への貢献
安全安心便利な製品・サービス
（産官学連携・ベンチャー）

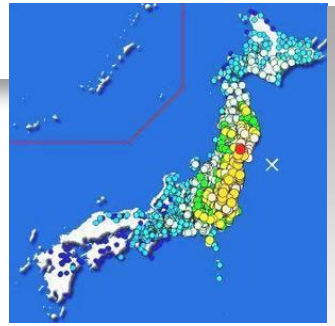
高速化

低消費電力化

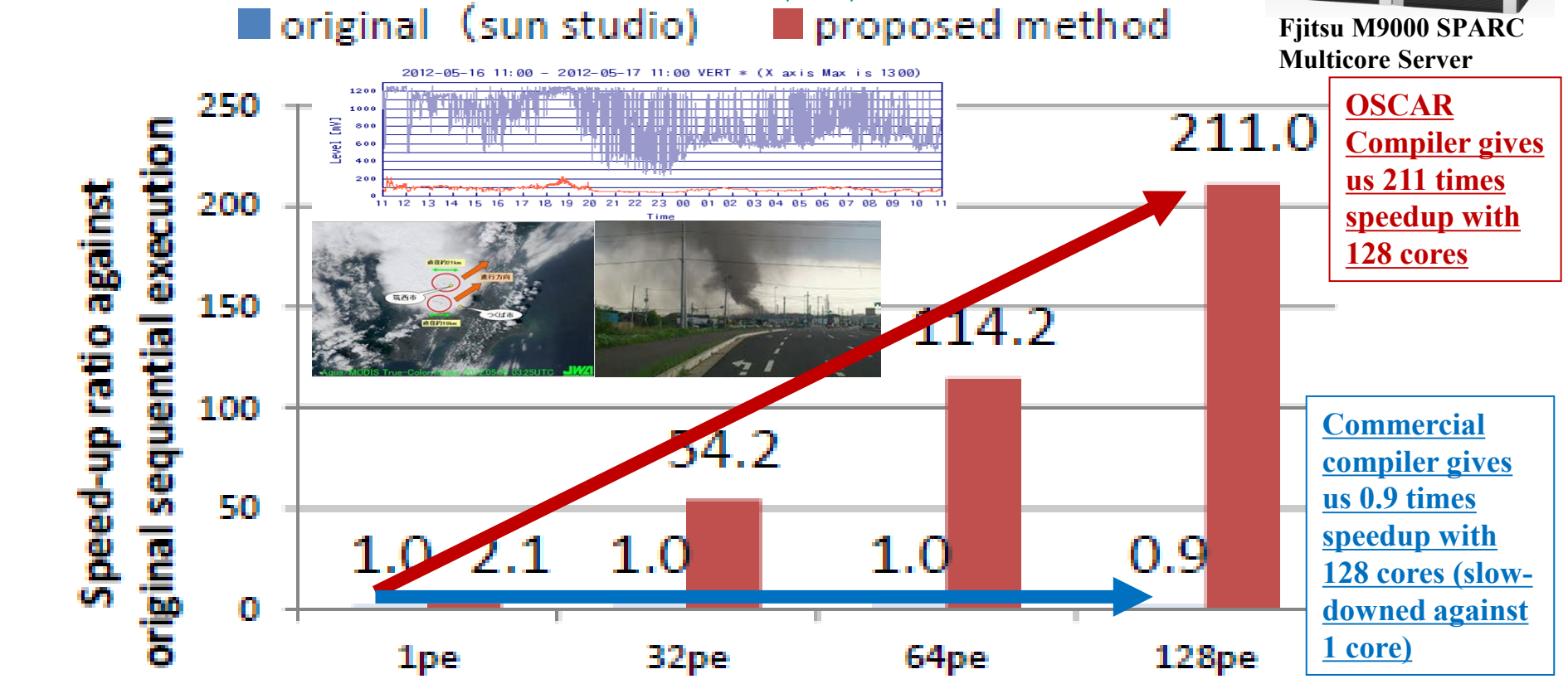
地震波伝搬シミュレーションの高速化：富士通スパコン128プロセッサコア上で、従来1コアより211倍の高速化に成功

- 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



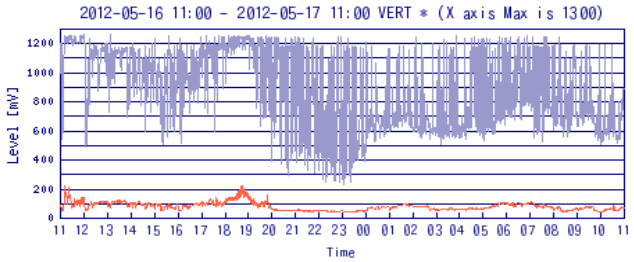
OSCAR Compiler gives us 211 times speedup with 128 cores

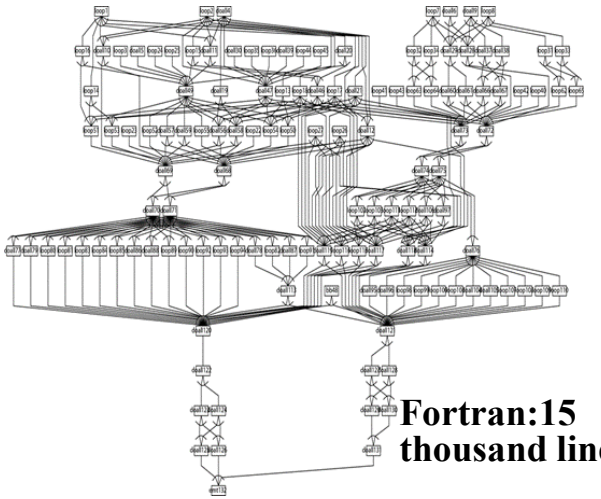
Commercial compiler gives us 0.9 times speedup with 128 cores (slow-downed against 1 core)

- 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

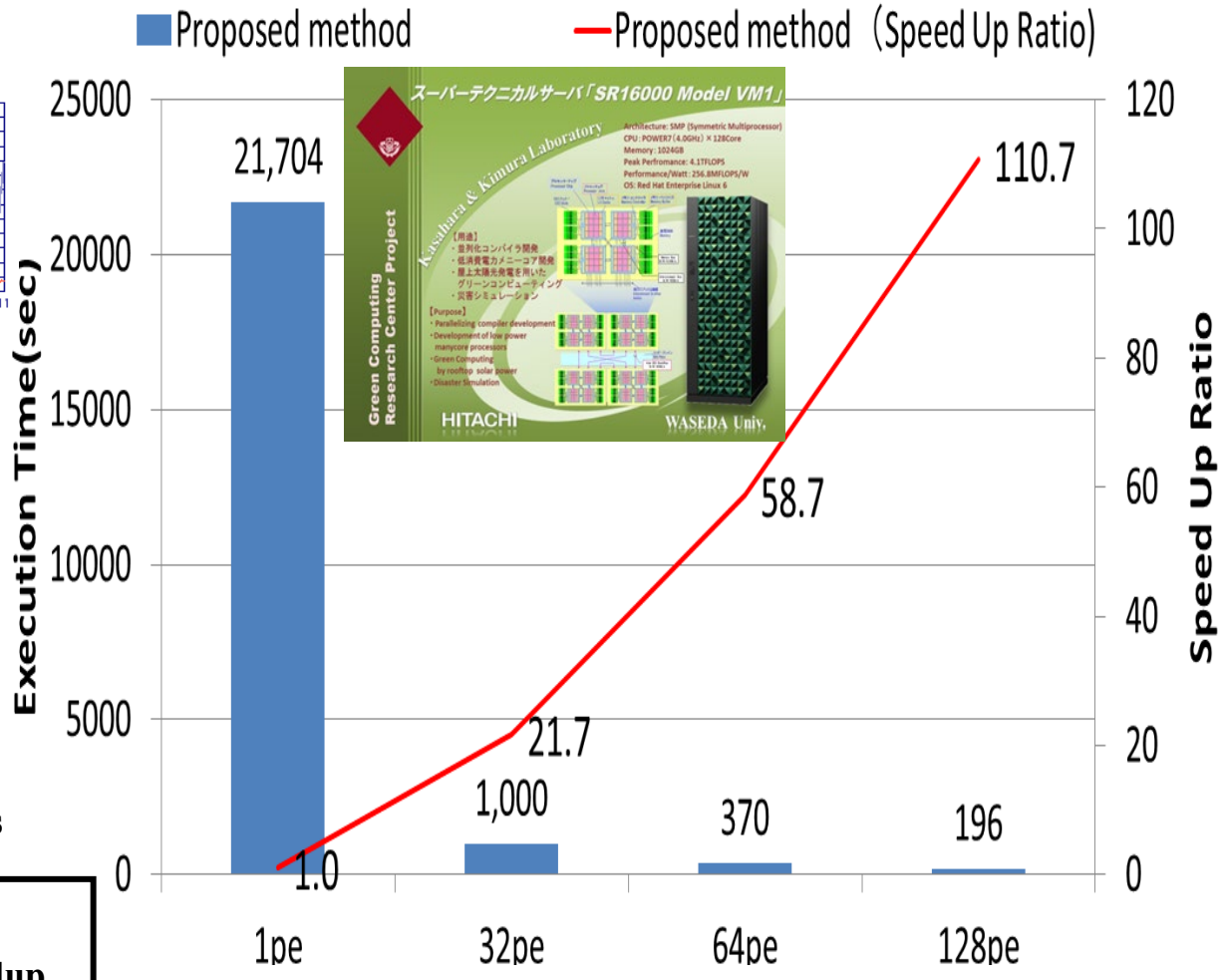
110 Times Speedup against the Sequential Processing for GMS Earthquake Wave Propagation Simulation on Hitachi SR16000

（Power7 Based 128 Core Linux SMP） (LCPC2015)





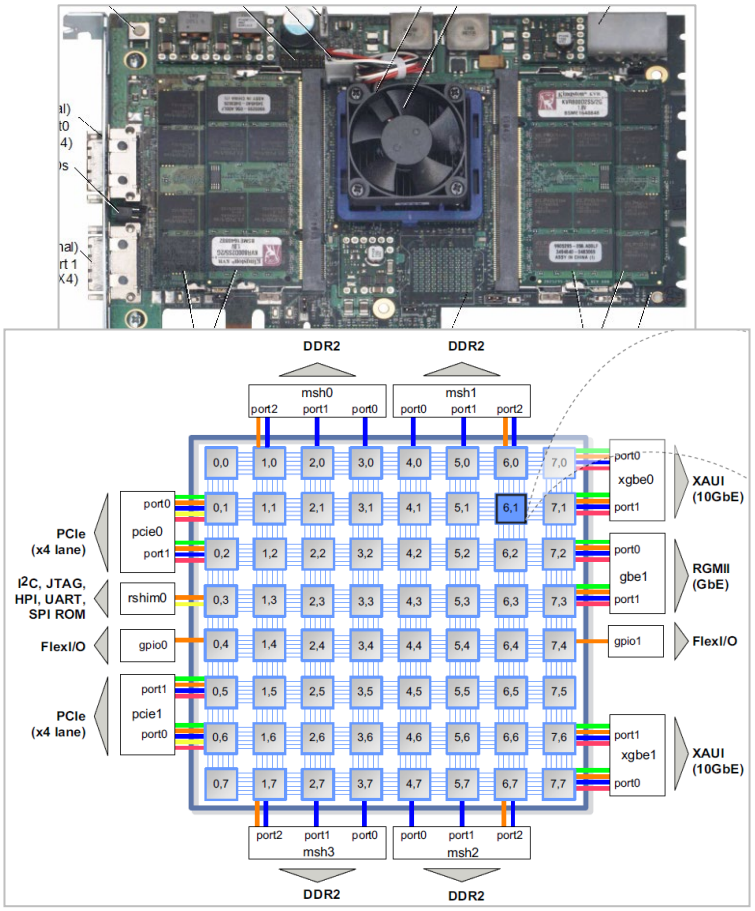
Fortran:15 thousand lines

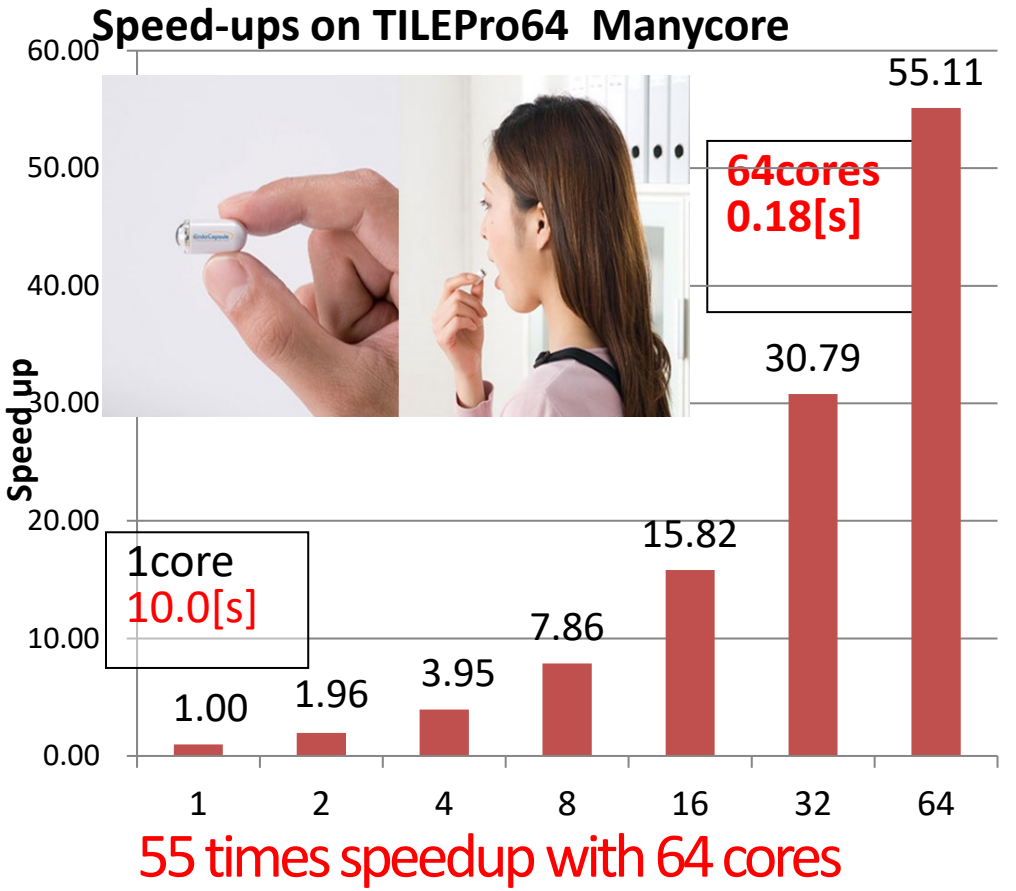


First touch for distributed shared memory and cache optimization over loops are important for scalable speedup

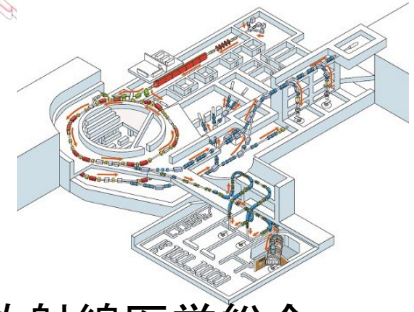
飲むカプセル内視鏡の研究開発
高精細画像圧縮プログラム JPEG-XRの並列化
64プロセッサコア集積のチップで55倍処理を高速化

- TILEPro64（MITマサチューセッツ工科大学発ベンチャーの64コアチップ）



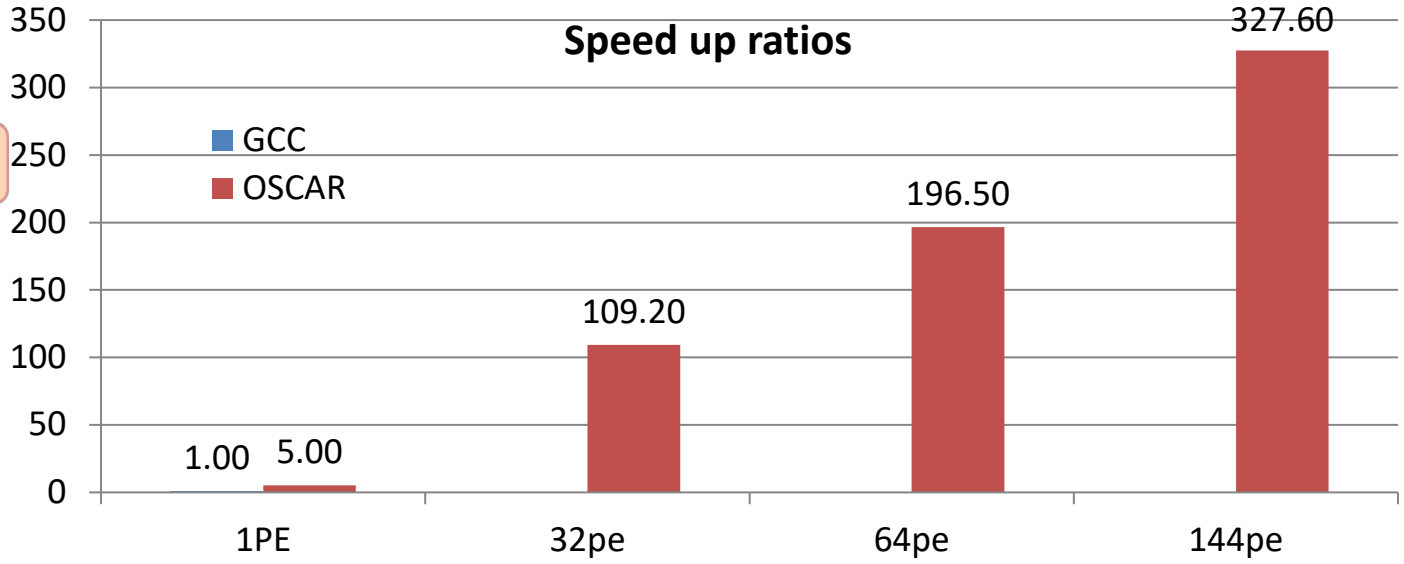


重粒子線がん治療計算の日立BS500ブレードサーバ上での並列化



放射線医学総合研究所サイトより
http://www.nirs.qst.go.jp/rd/cpt/index.html

日立 SMPブレードサーバ BS500:
Xeon E7-8890 V3(2.5GHz　18core/chip) x8 chip 計144cores



- オリジナル逐次実行時間2948秒（約50分）が、OSCARコンパイラによる144コア並列処理で、9秒に短縮され、327.6倍の速度向上

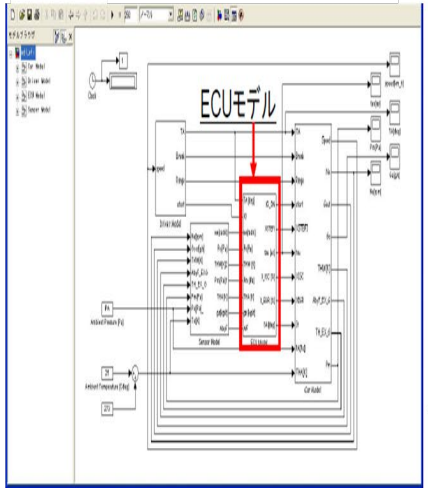
日本乗用車のエンジン制御計算をデンソー2コアECU上で、1．95倍の速度向上に成功。

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。

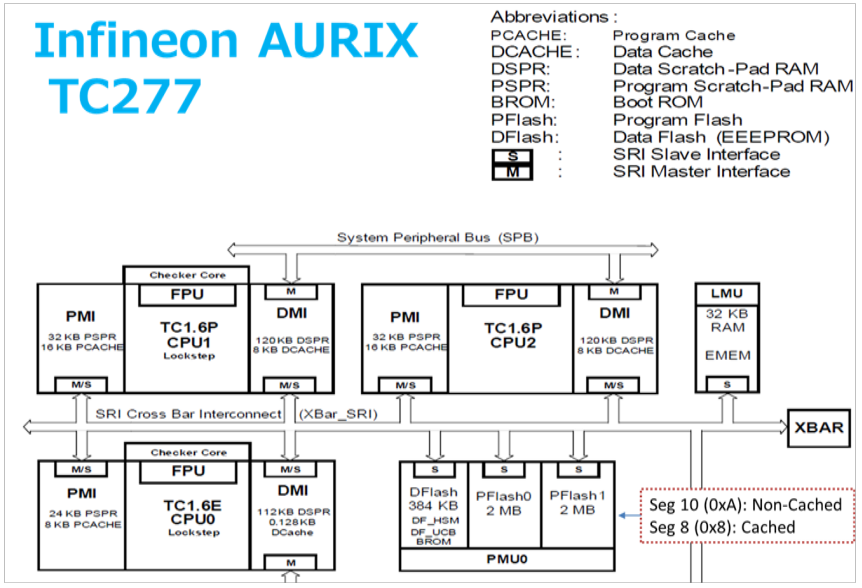
Engine Control by multicore with Denso

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



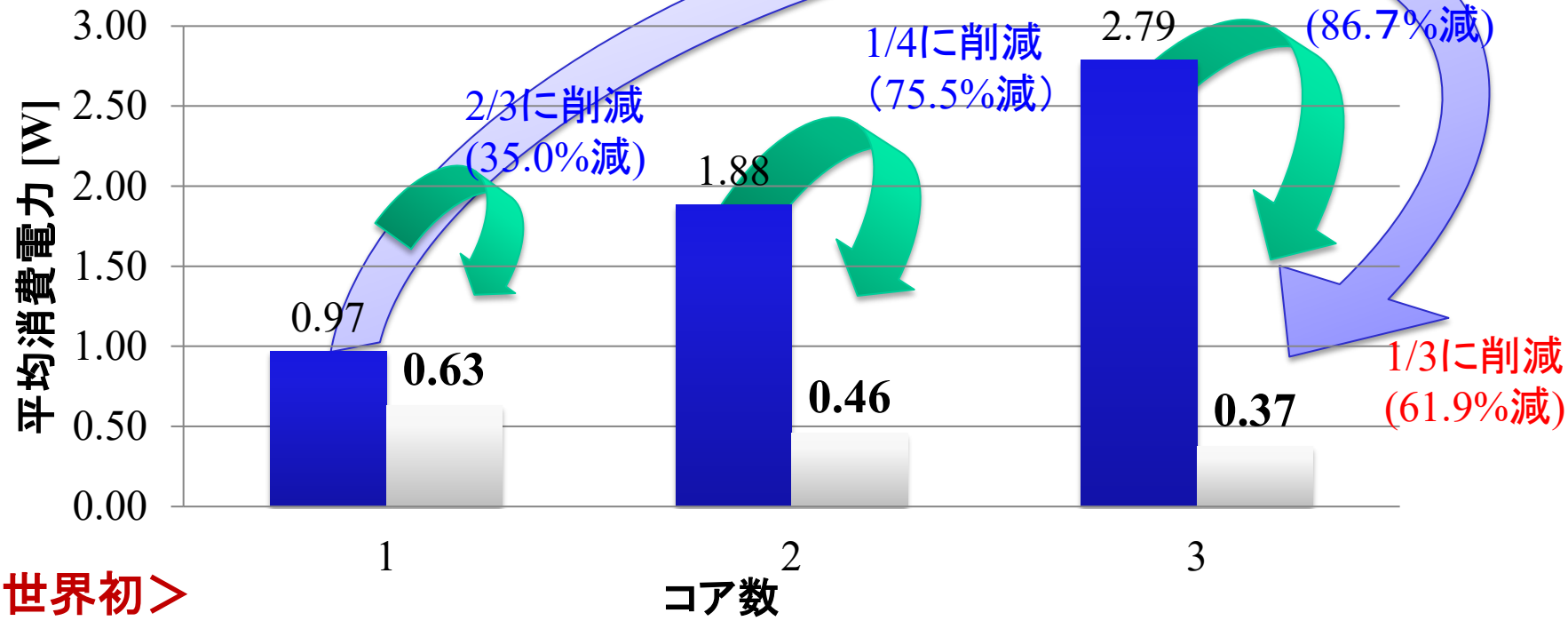
Automatic Parallelization of an Engine Control C Program with 400 thousands lines on AUTOSAR on 2 cores of Infineon AURIX TC277



Androidスマートフォン上での電力削減

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

週1回以下の充電,さらには
太陽光充電を目指して

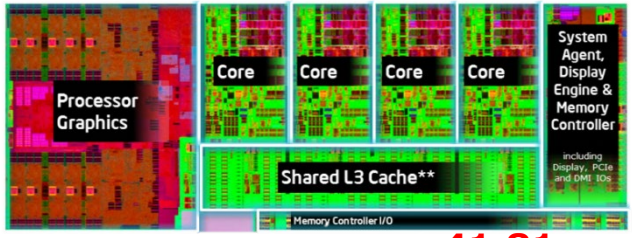


＜世界初＞

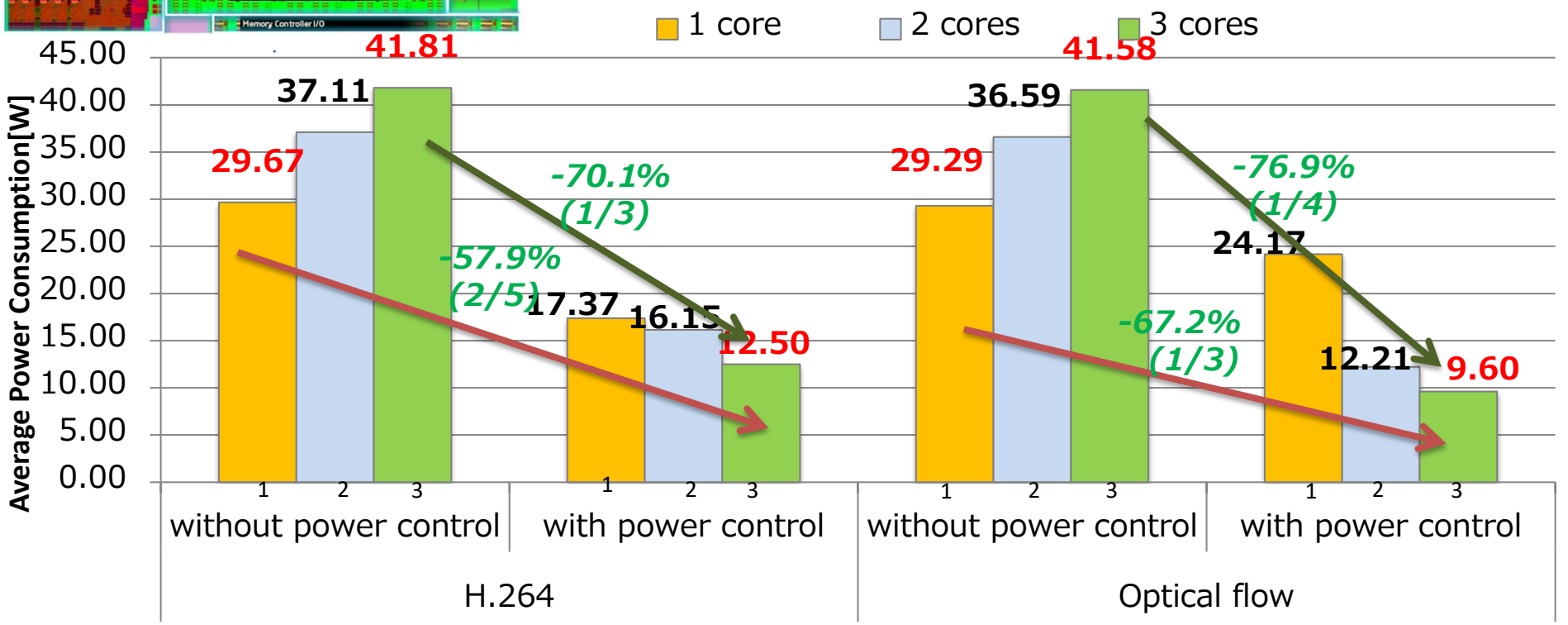
- 3PE電力制御なしと3PE電力制御ありで電力を最大**1/7**に削減
- 1PE電力制御なしと3PE電力制御ありで電力を**1/3**に削減

Automatic Power Reuction on Intel Haswell

H.264 decoder & Optical Flow (3cores)

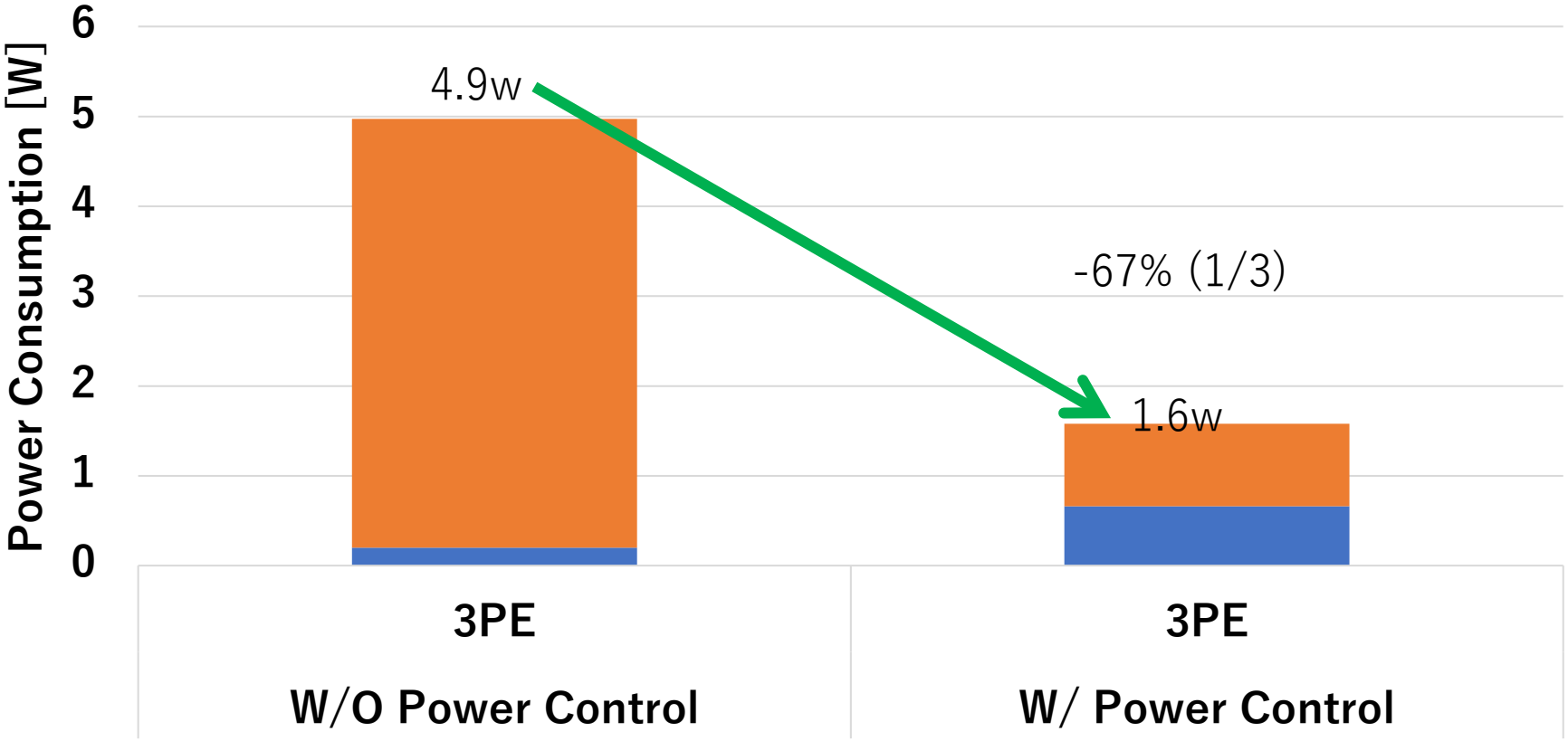


H81M-A, Intel Core i7 4770k
Quad core, 3.5GHz～0.8GHz



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

年200億個以上のプロセッサが生産

マルチコアプロセッサ：スマホ,タブレット,IoTデバイス,自動車制御,サーバ,スパコン等

例：ARM,IBM Power,Renesas RH850, Infineon,SPARC,RISC V

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

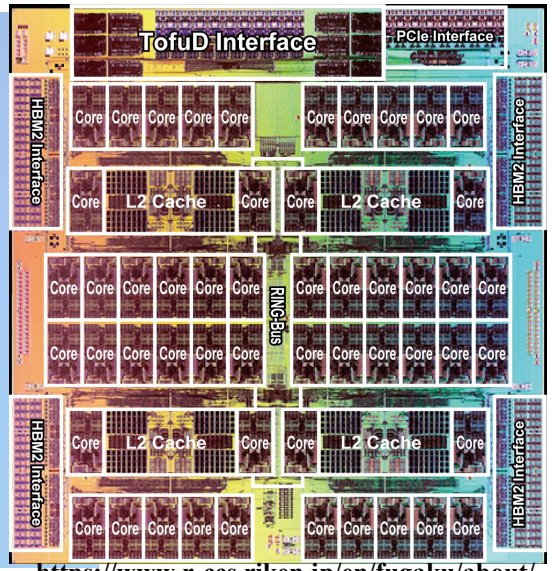
64-bit
iPhone 13
2021

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“高性能コア & 4 ”Blizzard”省エネコア
Instruction set: A64, Transistors: 15 billion (15億個)
GPU(s): Apple-designed 5 core GPU in iPhone 13

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

理研富岳スーパーコンピュータ 2020年6月から2021年11月まで世界No.1

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

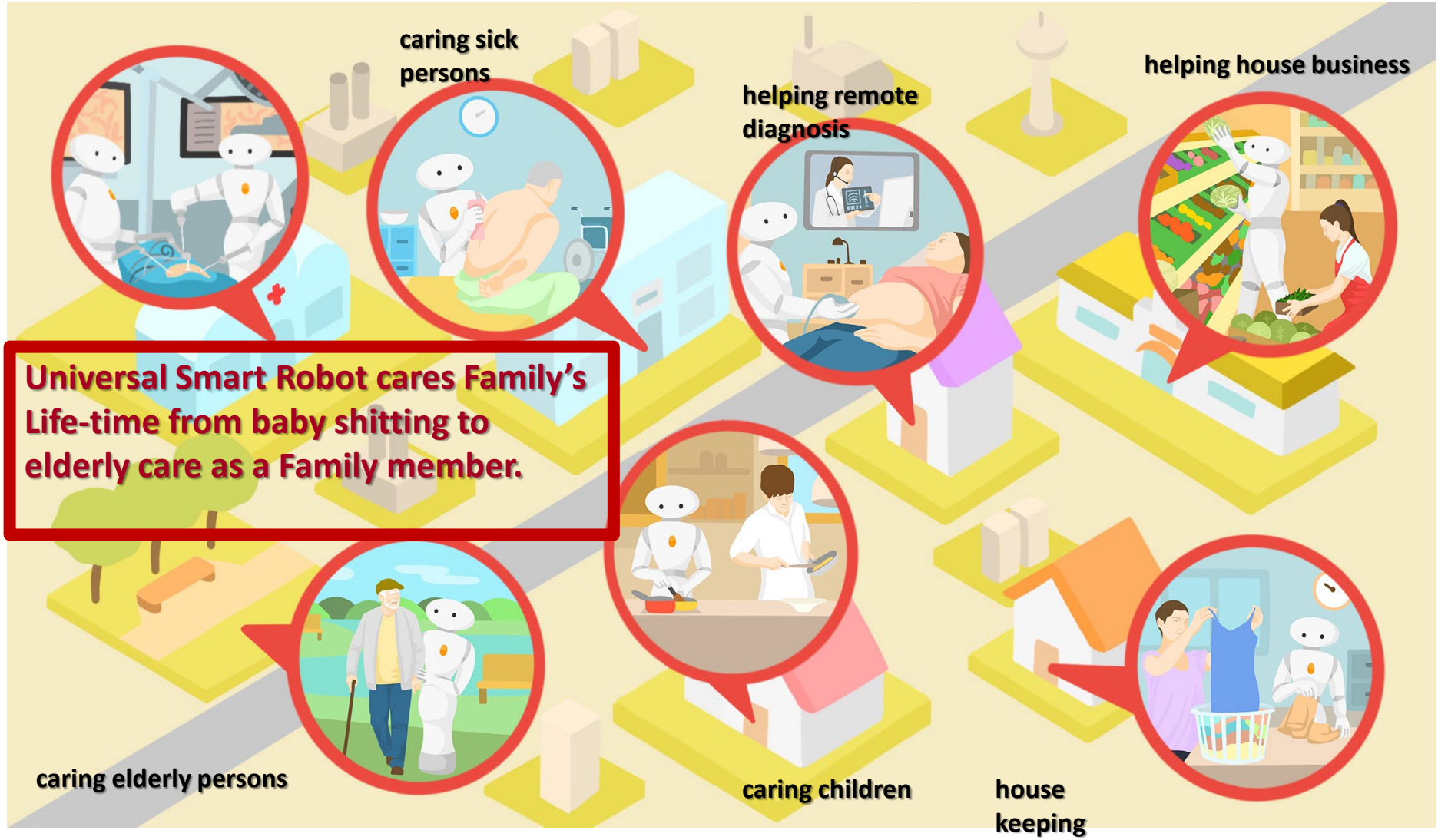
48コア/チップ, 2.2GHz, 7 nm FinFET,
約7百30万コア, 28MW
理論最高性能:51京回浮動小数点演算/秒,
2020年6月時点

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

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

Supported by Japanese Government "Moonshot" Project from 2020

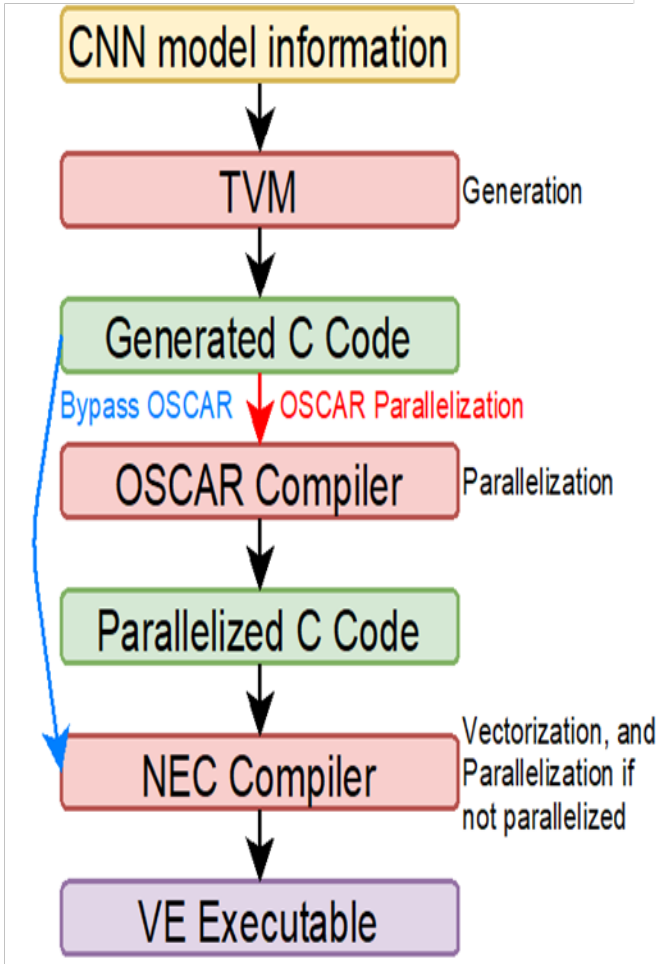
人生に寄り添うAIロボット：育児,家事,家業補助,看護,遠隔診断,介護

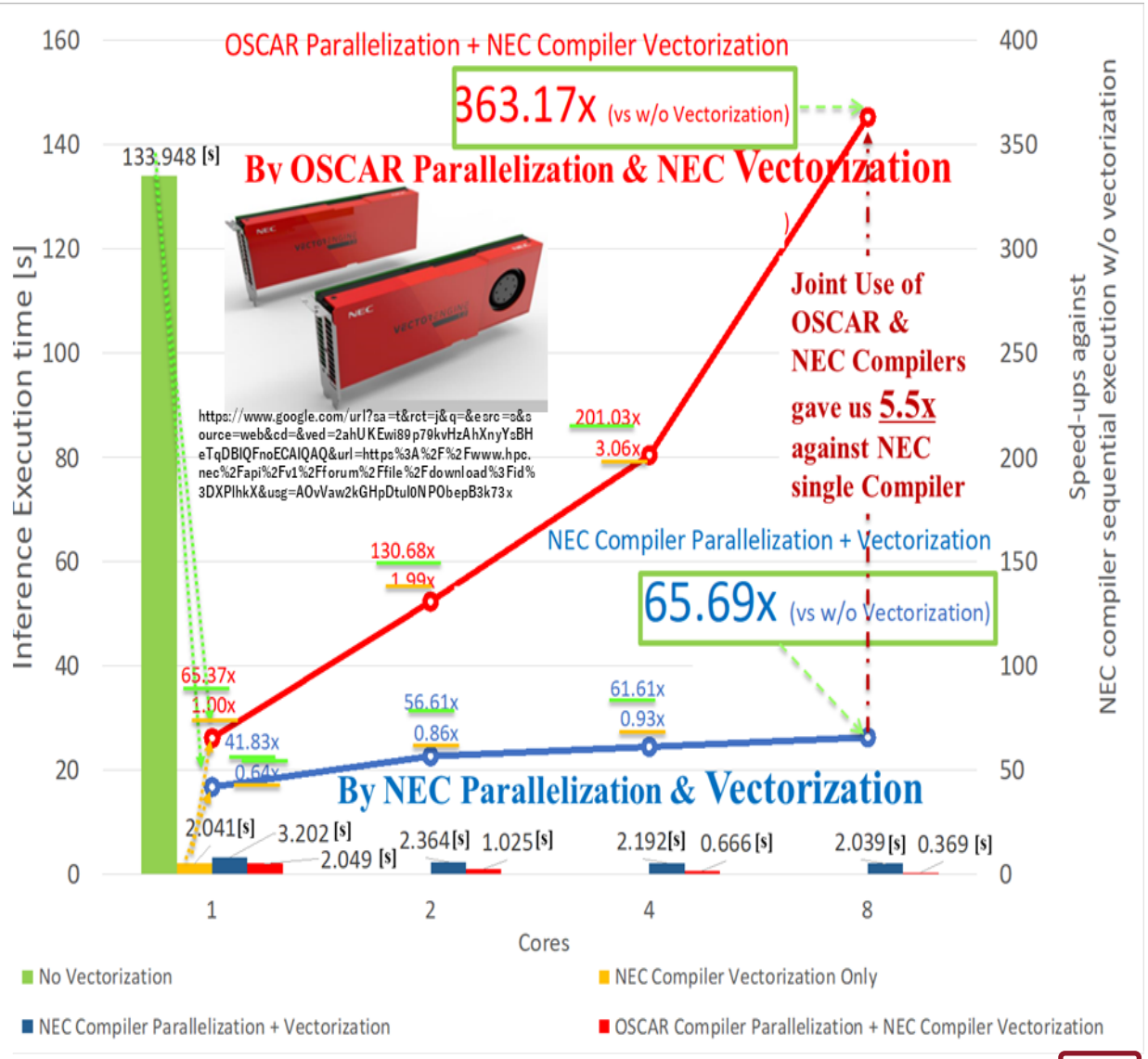


Speedups of Deep Learning Winograd 2D-Convolution generated by TVM on NEC Personal Vector Supercomputer SX-Aurora TSUBASA 8 Core Type 10C

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

Parallelization of Deep Leaning C Code generated by TVM

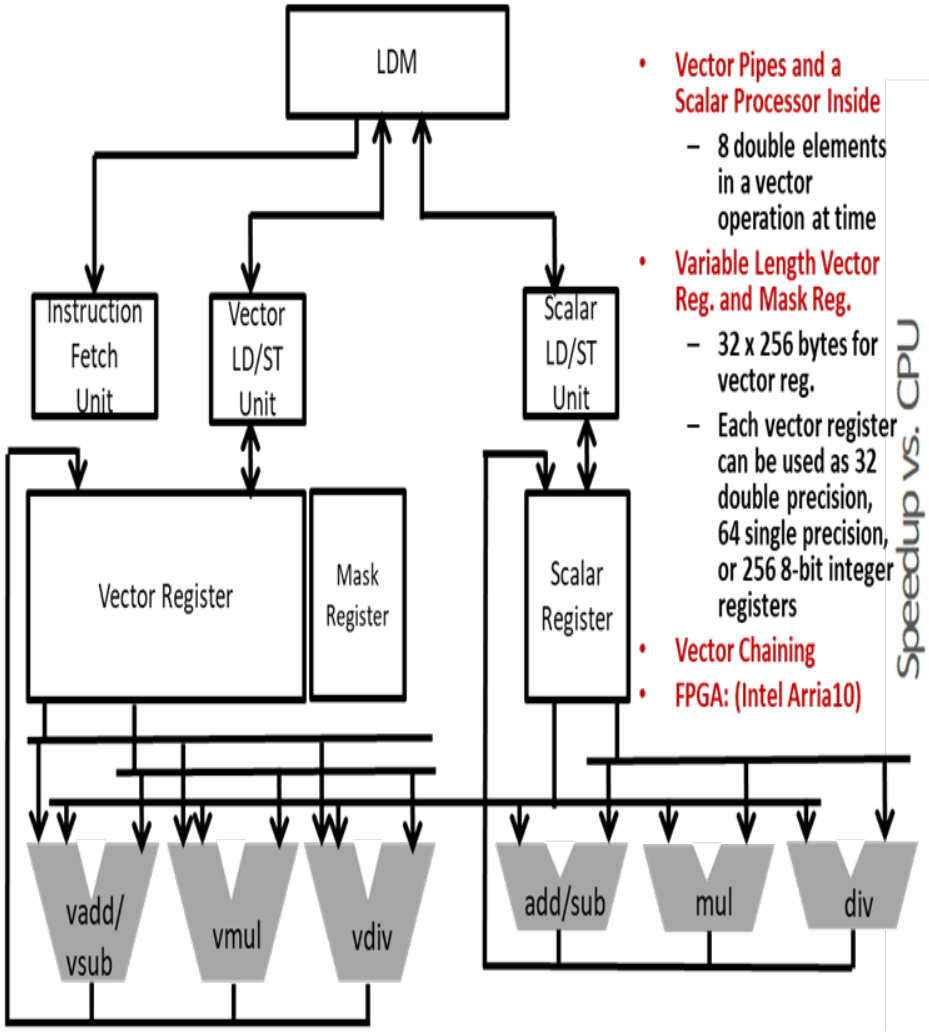




Performance for Multimedia and Scientific Applications on OSCAR Vector Accelerator

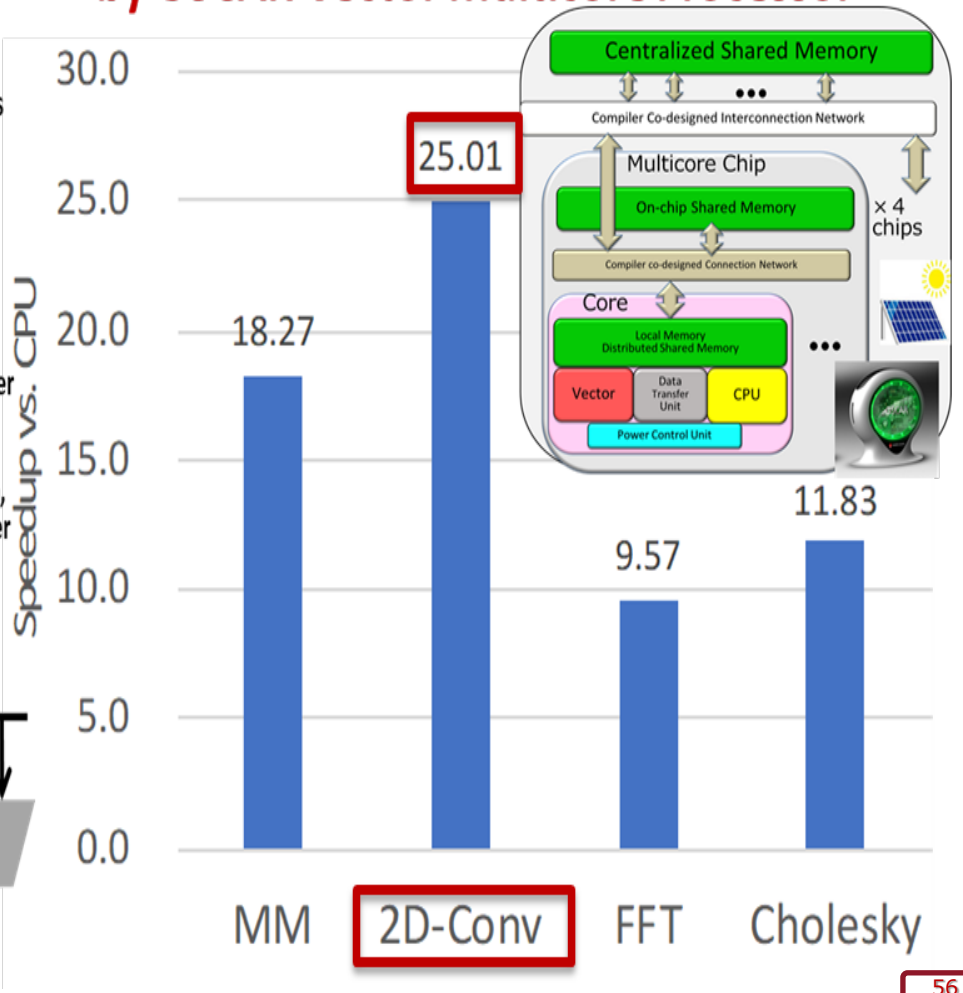
(A Vector Processor with Local Memory or Distributed Shared Memory and DMA Controller Managed by the OSCAR Compiler Redesigned Improving Japanese Supercomputer Technology in 1980-2000)

OSCAR Vector Accelerator on FPGA



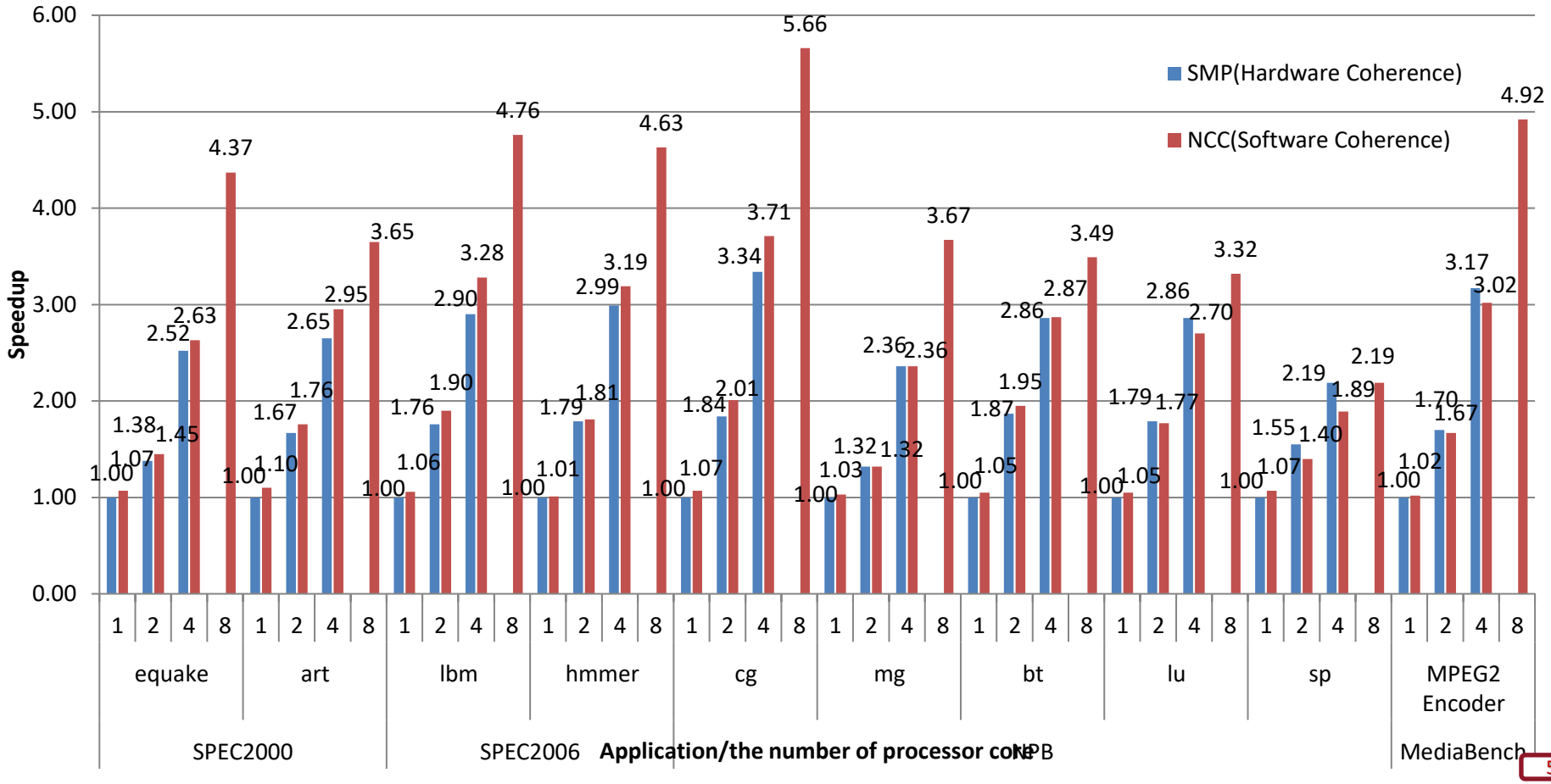
- Vector Pipes and a Scalar Processor Inside
  - 8 double elements in a vector operation at time
- Variable Length Vector Reg. and Mask Reg.
  - 32 x 256 bytes for vector reg.
  - Each vector register can be used as 32 double precision, 64 single precision, or 256 8-bit integer registers
- Vector Chaining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

Speedups against General Purpose Processor by OSCAR Vector Multicore Processor



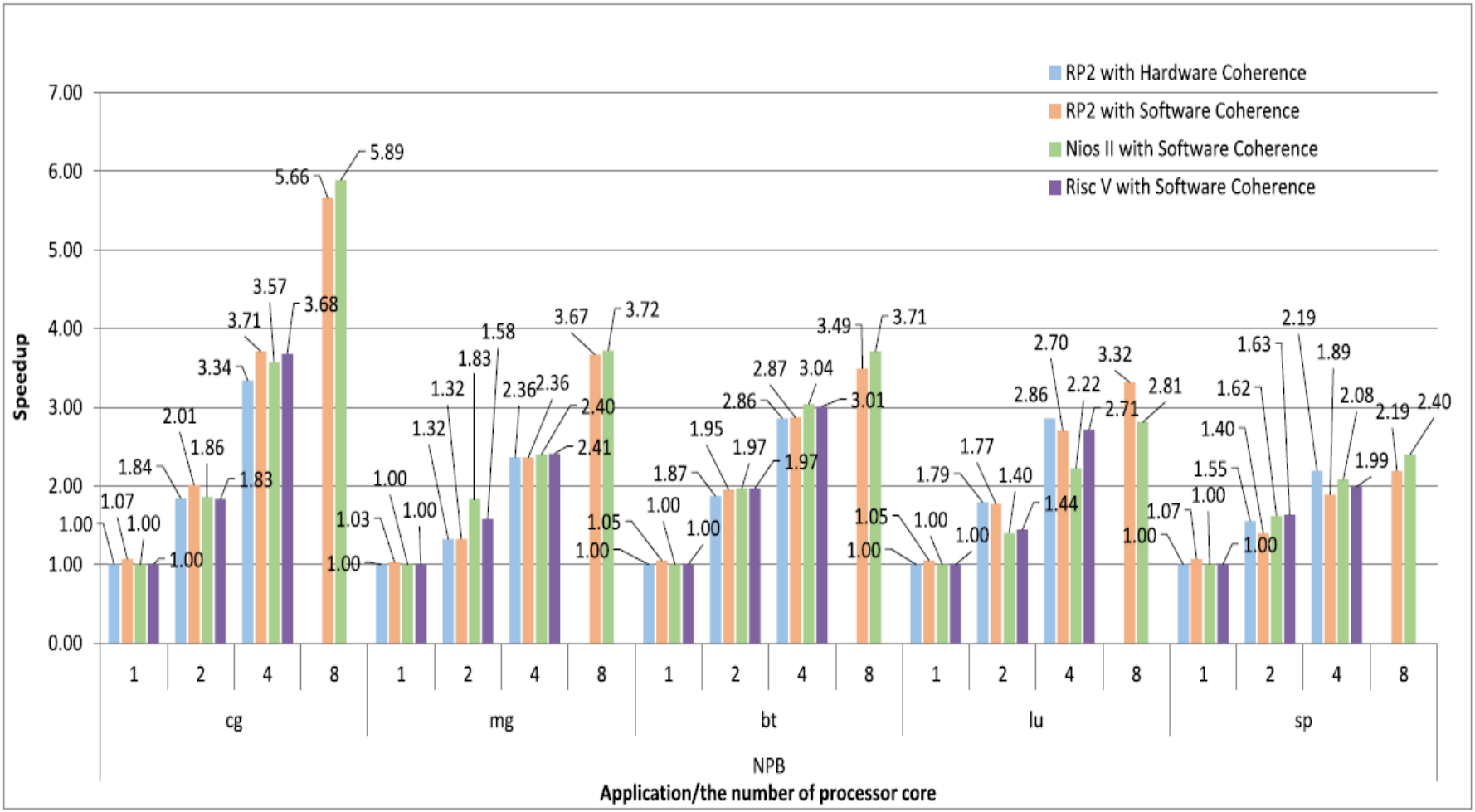
Automatic Software Coherent Control for Manycores

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



OSCAR Software Cache Coherent Control for NIOS and RISCV cores on FPGA

3.57x Speedups for NIOS and 3.68x for RISCV using 4 cores for NPB CG

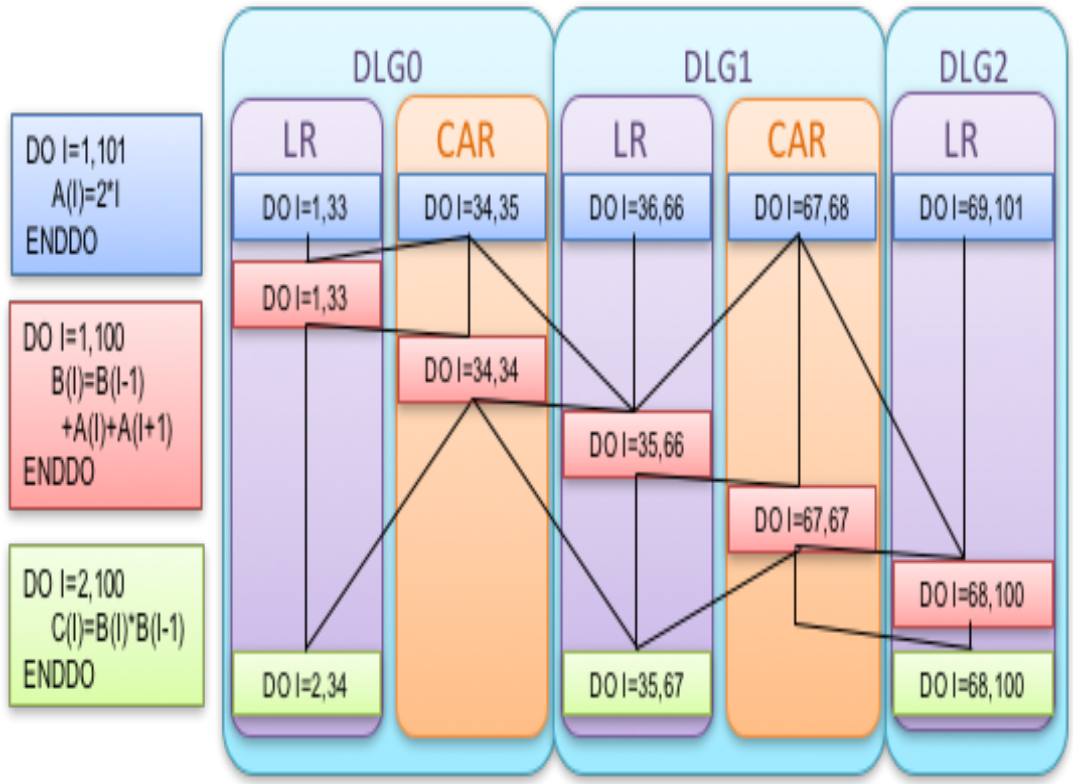


Automatic Local Memory Management

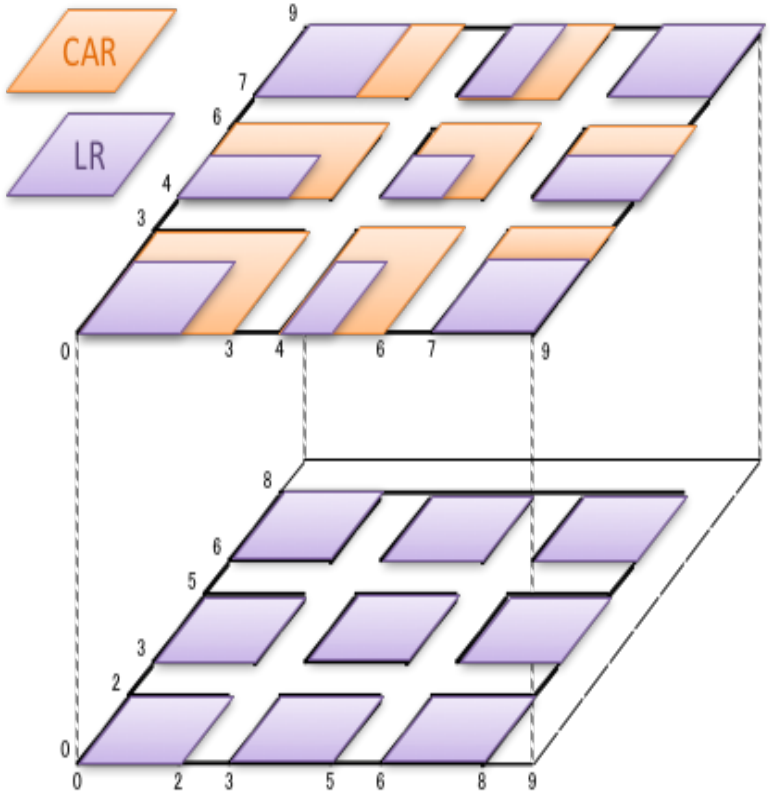
Data Localization: Loop Aligned Decomposition

- **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

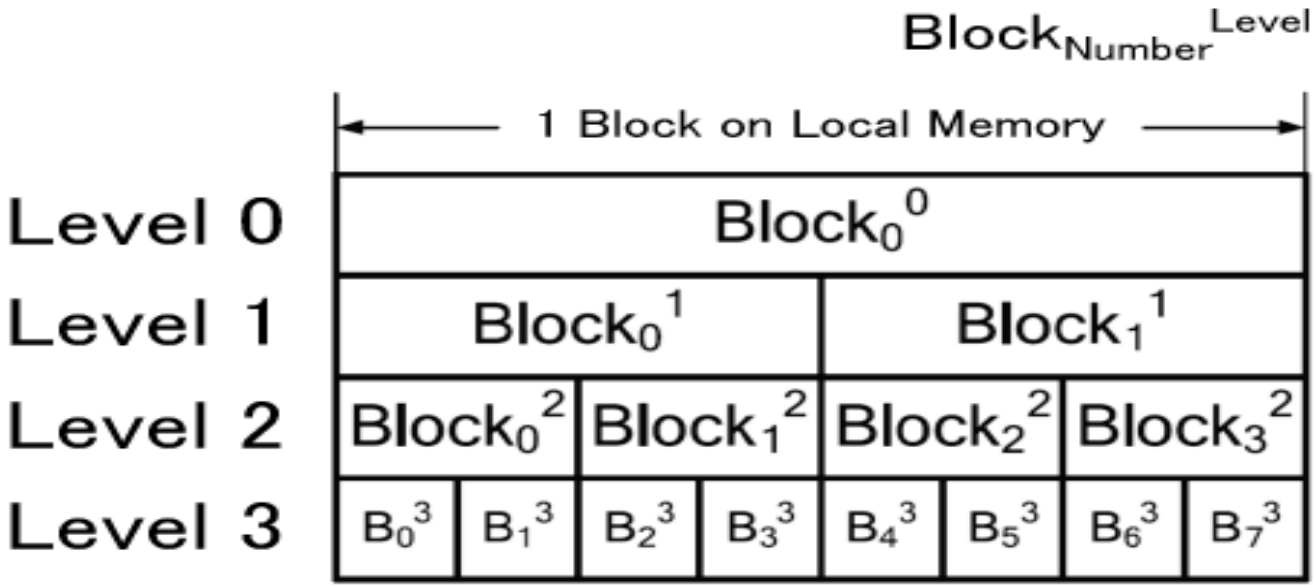


Multi-dimension Decomposition



Adjustable Blocks

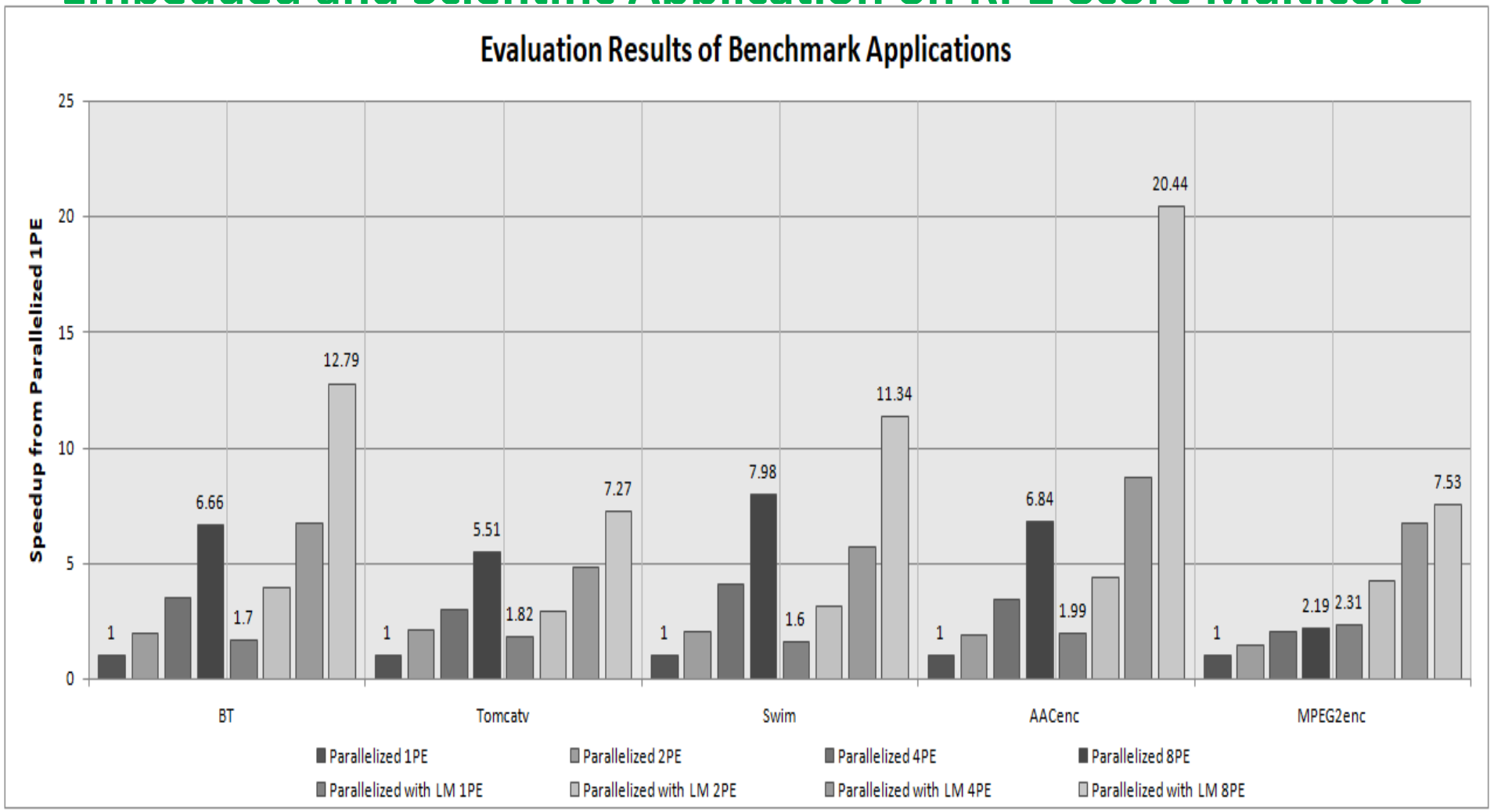
- Handling a suitable block size for each application
  - different from a fixed block size in cache
  - each block can be divided into smaller blocks with integer divisible size to handle small arrays and scalar variables



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 OSCAR Automatic Local Memory Management compared to Executions Utilizing Centralized Shared Memory on Embedded and Scientific Application on RP2 8core Multicore



Maximum of 20.44 times speedup on 8 cores using local memory against sequential execution using off-chip shared memory

グリーンコンピューティングセンターでの国際イベント

The 25th International Workshop on Languages and Compilers for Parallel Computing (LCPC2012), September 11-13, 2012

MULTICORE VIDEO SERIES

Practical Innovation

Multicore processors have become pervasive, but most organizations struggle to use them efficiently. That's why we brought together renowned experts in the field for this video series to examine the innovative techniques they use to improve reliability and performance while reducing costs, time, and power consumption.

Hear about some of the most advanced power-reduction, parallelization, and vectorization technologies used in a range of industry applications, including automobiles, big data, cloud computing, cluster computing, medical image processing, multimedia, smartphones, and supercomputing.

World's best educational content

Learn from the World's Leading Multicore Compiler Experts


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


Vectorization
P. Sadayappan


Vectorization/Parallelization in the IBM Compiler
Yaoqing Gao


Vectorization/Parallelization in the Intel Compiler
Peng Tu


Roundtable Discussion
All Presenters

Who Should Watch these Videos?

Professionals in any industry that demands real-time processing, high performance, and speed will find these videos an important tool for getting better results from their multicore processing systems and future-proofing their applications.

Educators and graduate students will also find inspiration from this window into the minds of some of the most accomplished experts in multicore.

www.computer.org/multicore-video

IEEE computer society

IEEE

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

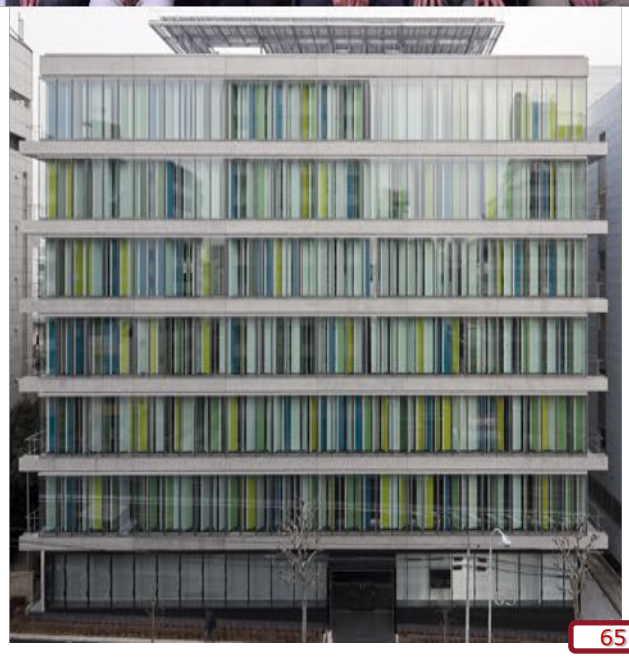
Keynote: Paul Messina ANL, USA

- 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



Oxford University, 11/12-13,2019(CSでの招待講演及び連携協議)

Oxford大は、2022年104現在、THE大学ランキング7年連続No.1

Vice Chancellor Prof. Louise Richardson
(WoI 2020での基調講演(予定))
Head of Astrophysics:Prof. Rob Fender
Dept. of Physics: Prof. Ian Shipsey
Astrophysics: Prof. H.Falche, et. al.

Merton College
Warden: Prof. Irene Tracy
Fellow: Dr. Peter Braam
Sub Warden: Prof. Judy Armitage
CS: Prof. Jeremy Gibbons

Choral Evensong, 750th Anniversary Room

ソーラーパワー・パーソナル・スパコン：新アクセラレータ・グリーンマルチコア

（AI、ビッグデータ、自動運転車、交通制御、ガン治療、地震、ロボット）

世界最高性能・低電力化機能OSCARコンパイラとの協調

ベクトルアクセラレータ併置・
共有メモリ型マルチコアシステム

性能：8TFLOPS, 主メモリ：8TB

電力: 40W, 効率：200GFLOPS/W

- 命令拡張などのプロセッサにも付加できるベクトルアクセラレータ
  - 低消費電力で高速に立ち上がるベクトルで、低コスト設計
- コンパイラによる自動ベクトル・並列化及び自動電力削減
  - 周波数・電源電圧制御機能
  - バリア高速同期・ローカル分散メモリで無駄削減
  - ローカルメモリ利用で低メモリコスト
  - 誰でもチューニングなく使用でき、低コスト短期間ソフト開発可能

Future Automatic Parallelizing Compiler Codesigned Multicore Sustainable Products

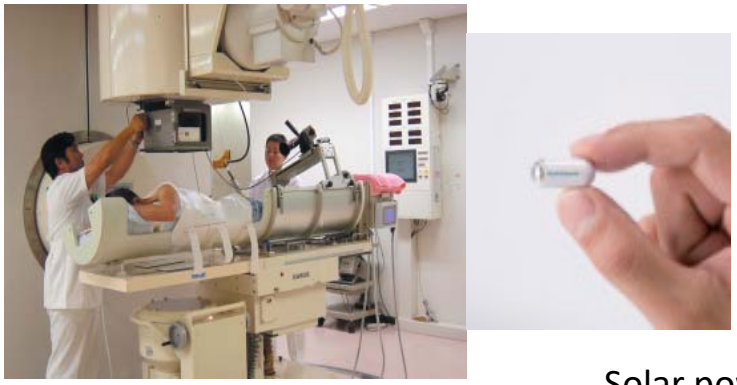
Next Generation Automobiles

- 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