OSCAR Automatic Paralleling and Power Reducing Compiler for Multicores

Hironori Kasahara
Professor, Dept. of Computer Science & Engineering
Director, Advanced Multicore Processor Research Institute
Waseda University, Tokyo, Japan
IEEE Computer Society Multicore STC Chair
URL: http://www.kasahara.cs.waseda.ac.jp/

Waseda Univ. GCSC
Multicore for Performance and Low Power

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

\[
\text{Power} \propto \text{Frequency} \times \text{Voltage}^2 \\
(\text{Voltage} \propto \text{Frequency}) \\
\Rightarrow \text{Power} \propto \text{Frequency}^3
\]

If **Frequency** is reduced to \(\frac{1}{4}\) (Ex. 4GHz\(\rightarrow\)1GHz),

- **Power** is reduced to \(\frac{1}{64}\) and
- **Performance** falls down to \(\frac{1}{4}\).

**<Multicores>**

If **8cores** are integrated on a chip,

- **Power** is still \(\frac{1}{8}\) and
- **Performance** becomes **2 times**.

Compiler is the Key Technology

---

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”
With 128 cores, OSCAR compiler gave us 100 times speedup against 1 core execution and 211 times speedup against 1 core using Sun (Oracle) Studio compiler.
Renesas-Hitachi-Waseda Low Power 8 core RP2
Developed in 2007 in METI/NEDO project

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”
8 Core RP2 Chip Block Diagram

- Core #0
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #1
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #2
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #3
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #4
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #5
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #6
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- Core #7
  - I$: 16K
  - D$: 16K
  - CPU
  - FPU
  - Local memory: I: 8K, D: 32K
  - URAM: 64K
  - CCN
  - BAR

- LCPG: Local clock pulse generator
- PCR: Power Control Register
- CCN/BAR: Cache controller/Barrier Register
- URAM: User RAM (Distributed Shared Memory)
Power Reduction of MPEG2 Decoding to 1/4 on 8 Core Homogeneous Multicore RP-2 by OSCAR Parallelizing Compiler

**MPEG2 Decoding with 8 CPU cores**

Without Power Control
(Voltage : 1.4V)

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

Avg. Power
5.73 [W]

73.5% Power Reduction

Avg. Power
1.52 [W]
Demo of NEDO Multicore for Real Time Consumer Electronics at the Council of Science and Engineering Policy on April 10, 2008

CSTP Members
Prime Minister: Mr. Y. FUKUDA
Minister of State for Science, Technology and Innovation Policy: Mr. F. KISHIDA
Chief Cabinet Secretary: Mr. N. MACHIMURA
Minister of Internal Affairs and Communications: Mr. H. MASUDA
Minister of Finance: Mr. F. NUKAGA
Minister of Education, Culture, Sports, Science and Technology: Mr. K. TOKAI
Minister of Economy, Trade and Industry: Mr. A. AMARI
Green Computing Systems R&D Center
Waseda University
Supported by METI (Mar. 2011 Completion)

<R & D Target>
Hardware, Software, Application for Super Low-Power Manycore Processors
➢ More than 64 cores
➢ Natural air cooling (No fan)
   Cool, Compact, Clear, Quiet
➢ Operational by Solar Panel

<Industry, Government, Academia>
Hitachi, Fujitsu, NEC, Renesas, Olympus, Toyota, Denso, Mitsubishi, Toshiba, etc

<Ripple Effect>
➢ Low CO₂ (Carbon Dioxide) Emissions
➢ Creation Value Added Products
   ➢ Consumer Electronics, Automobiles, Servers

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

Beside Subway Waseda Station,
Near Waseda Univ. Main Campus
Industry-government-academia collaboration in R&D and target practical applications

Solar Powered Smart phones
Solar Powered, Non-fan, cool, quiet servers
Operation/recharging by solar cells

Robots

Multicore Engine ECU, ADAS (Driver Assistance), Self Driving, HV, EV, FCV

Consumer electronic
Internet TV/DVD

Cameras

Camcorders

Disaster Survival Supercomputer (Earthquakes, tsunami, Fire-spreading)

Many-core system technologies with ultra-low power consumption

Waseda University: R&D

For smart life

Protect Environment

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

Supercomputers and servers

Waseda University: R&D

Many-core system technologies with ultra-low power consumption

OSCAR Multicore

Vector Acc.

OSCAR Technology

Many-core System technologies with ultra-low power consumption

Optimization, Compiler, API

Solar Powered

Medical servers

Operation/recharging by solar cells

Heavy particle radiation planning, cerebral infarction

OSCAR Technology

Disaster Survival

Capstone servers

Supercomputers

National Institute of Radiological Sciences

Solar Powered, Non-fan, cool, quiet servers

For smart life

Industry

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

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

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

Data Transfer Overlapping
Data transfer overlapping using Data Transfer Controllers (DMAs)

Power Reduction
Reduction of consumed power by compiler control DVFS and Power gating with hardware supports.
Multicore Program Development Using OSCAR API V2.0

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

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

Low Power Homogeneous Multicore Code Generation
- Proc0
  Code with directives
  Thread 0
- Proc1
  Code with directives
  Thread 1

Low Power Heterogeneous Multicore Code Generation
- Accelerator 1
  Code
- Accelerator 2
  Code

Server Code Generation
OpenMP Compiler

OpenMP Compiler

Generation of parallel machine codes using sequential compilers

Homegeneous Multicores from Vendor A (SMP servers)

Heterogeneous Multicores from Vendor B

Shred memory servers

OSCAR: Optimally Scheduled Advanced Multiprocessor API: Application Program Interface
Cancer Treatment
Carbon Ion Radiotherapy
(Previous best was 2.5 times speedup on 16 processors with hand optimization)

8.9 times speedup by 12 processors
Intel Xeon X5670 2.93GHz 12 core SMP (Hitachi HA8000)

55 times speedup by 64 processors
IBM Power 7 64 core SMP (Hitachi SR16000)
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
OSCAR Compile Flow for Simulink Applications

1. Generate MTG → Parallelism
2. Generate gantt chart → Scheduling in a multicore
3. Generate parallelized C code using the OSCAR API → Multiplatform execution (Intel, ARM and SH etc)
Speedups of MATLAB/Simulink Image Processing on Various 4core Multicores

(Intel Xeon, ARM Cortex A15 and Renesas SH4A)

Road Tracking, Image Compression: http://www.mathworks.co.jp/jp/help/vision/examples
Vessel Detection: http://www.mathworks.co.jp/matlabcentral/fileexchange/24990-retinal-blood-vessel-extraction/
Performance of OSCAR Compiler on Intel Core i7 Notebook PC

- OSCAR Compiler accelerate Intel Compiler about **2.0 times** on average

CPU: Intel Core i7 3720QM (Quad-core)
MEM: 32GB DDR3-SODIMM PC3-12800
OS: Ubuntu 12.04 LTS
Parallelization of 2D Rendering Engine SKIA on 3 cores of Google NEXUS7

http://www.youtube.com/channel/UCS43lNYElkC8i_KIgFZYQBQ

On Nexus7, 3 core parallelization gave us

for DrawRect: 1.91 speedup
for DrawImage: 1.95 speedup
110 Times Speedup against the Sequential Processing for GMS Earthquake Wave Propagation Simulation on Hitachi SR16000 (Power7 Based 128 Core Linux SMP)
Automatic Parallelization of Still Image Encoding Using JPEG-XR for the Next Generation Cameras and Drinkable Inner Camera

TILEPro64

Speed-ups on TILEPro64 Manycore

55 times speedup with 64 cores against 1 core
Parallel Processing of Face Detection on Manycore, Highend and PC Server

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

- DTU: Data Transfer Unit
- LPM: Local Program Memory
- LDM: Local Data Memory
- DSM: Distributed Shared Memory
- CSM: Centralized Shared Memory
- FVR: Frequency/Voltage Control Register
An Image of Static Schedule for Heterogeneous Multi-core with Data Transfer Overlapping and Power Control
33 Times Speedup Using OSCAR Compiler and OSCAR API on RP-X
(Optical Flow with a hand-tuned library)

Y. Yuyama, et al., “A 45nm 37.3GOPS/W Heterogeneous Multi-Core SoC”, ISSCC2010

Speedups against a single SH processor:

- 1SH: 1
- 2SH: 2.29
- 4SH: 3.09
- 8SH: 5.4
- 2SH+1FE: 18.85
- 4SH+2FE: 26.71
- 8SH+4FE: 32.65

111[fps]
Low-Power Optimization with OSCAR API

Scheduled Result by OSCAR Compiler

VC0
MT1
MT2
VC1
Sleep
MT3
MT4

Generate Code Image by OSCAR Compiler

void main_VC0() {
    MT1
    Sleep
    MT2
    MT4
}

void main_VC1() {
    MT1
    #pragma oscar fvcontrol ¥ 
    ((OSCAR_CPU(),0))
    MT2
    #pragma oscar fvcontrol ¥ 
    (1,(OSCAR_CPU(),100))
    Sleep
    MT3
    MT4
}
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] → 0.54[W]

1cycle: 33[ms] → 30[fps]
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 (Deadline 66.6 [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コンパイラによるHaswellマルチコア上での
自動低消費電力化(Intel 4コア)
- 消費電力を2/5に削減 -

Intel Haswell 4コア上での顔認識プログラム並列化

[低消費電力化時の平均消費電力]
- 電力制御なし - 電力制御あり

<table>
<thead>
<tr>
<th>コア数</th>
<th>平均消費電力（W）</th>
<th>約3/5に削減(-40.0%)</th>
<th>約2/5に削減(-60.2%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.36</td>
<td>25.51</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>41.35</td>
<td>16.46</td>
<td></td>
</tr>
</tbody>
</table>

顔認識プログラムの並列処理

入力処理 → カメラ → 画像を変えて探索ループ → OSCARコンパイラ

次フレームの処理 → 描画処理 → ディスプレイ

Intel Haswell 4コアの電力測定

CPU: Intel Core i7 4770k
コア数: 4
周波数: 3.5GHz〜0.8GHz
マザーボード: ASUS H81M-A

約2.44倍に速度向上 38.08[msec]

約2.44倍に速度向上 48.80[msec]

约2.44倍に速度向上 93.06[msec]
### Performance of OSCAR Compiler Software Coherence Control

- Faster or Equal Processing Performance up to 4 cores with hardware coherent mechanism on RP2.
- Software Coherence gives us correct execution without hardware coherence mechanism on 8 cores.

#### Hardware Coherence vs. Software Coherence

<table>
<thead>
<tr>
<th>No. of processor cores</th>
<th>SMP</th>
<th>Non-Coherent Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAC Encoder</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>1.89</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>5.90</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>3.59</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>3.54</td>
<td>3.34</td>
</tr>
<tr>
<td>MPEG2 Decoder</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>2.10</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.36</td>
<td></td>
</tr>
<tr>
<td>MPEG2 Encoder</td>
<td>1.02</td>
<td>1.00</td>
</tr>
</tbody>
</table>

#### Diagram

- Diagram showing hardware and software coherence performance comparison for different number of processor cores (1, 2, 4, 8) for AAC Encoder, MPEG2 Decoder, and MPEG2 Encoder.
Target:

- Solar Powered with compiler power reduction.
- Fully automatic parallelization and vectorization including local memory management and data transfer.
Summary

- Waseda University Green Computing Systems R&D Center supported by METI has been researching on low-power high performance Green Multicore hardware, software and application with government and industry including Hitachi, Fujitsu, NEC, Renesas, Denso, Toyota, Olympus and OSCAR Technology.

- OSCAR Automatic Parallelizing and Power Reducing Compiler has succeeded speedup and/or power reduction of scientific applications including “Earthquake Wave Propagation”, medical applications including “Cancer Treatment Using Carbon Ion”, and “Drinkable Inner Camera”, industry application including “Automobile Engine Control”, “Smartphone”, and “Wireless communication Base Band Processing” on various multicores from different vendors including Intel, ARM, IBM, AMD, Qualcomm, Freescale, Renesas and Fujitsu.

- In automatic parallelization, 110 times speedup for “Earthquake Wave Propagation Simulation” on 128 cores of IBM Power 7 against 1 core, 55 times speedup for “Carbon Ion Radiotherapy Cancer Treatment” on 64 cores IBM Power7, 1.95 times for “Automobile Engine Control” on Renesas 2 cores using SH4A or V850, 55 times for “JPEG-XR Encoding for Capsule Inner Cameras” on Tilera 64 cores Tile64 manycore.

- The compiler will be available on market from OSCAR Technology.

- In automatic power reduction, consumed powers for real-time multi-media applications like Human face detection, H.264, mpeg2 and optical flow were reduced to 1/2 or 1/3 using 3 cores of ARM Cortex A9 and Intel Haswell and 1/4 using Renesas SH4A 8 cores against ordinary single core execution.
IEEE Computer Society COMPSAC 2016 Plenary Panel
Rebooting Computing: Future of Architecture and Software
13:30-15:30, June 11, 2016, Atlanta, Georgia

Moderator:
Prof. Thomas M. Conte, 2015 IEEE CS President, Georgia Tech.,
Manycores, Microprocessor Architectures, Back-end Compiler

Panelists:
➢ Prof. Yale N. Patt, COMPSAC Keynote Speaker, Univ. Texas Austin,
  Microarchitectures & High Performance Systems
➢ Prof. Jean-Luc Gaudiot, 2017 CS President, UC Irvine,
  Data Flow Architectures and Languages
➢ Dr. Kunio Uchiyama, CS BoG, Chief scientist of Hitachi, Ltd.
  Embedded Multicore Processors, Embedded Applications
➢ Prof. Vladimir Getov, CS BoG, U.Westminster
  Energy Efficient Computing
➢ Prof. Hironori Kasahara, Chair, Multicore STC, Ex-CS BoG, Waseda U.
  Multicore Architecture, Parallelizing and Power Reducing Compiler