

* Fixstars Corporation, Tokyo, Japan † Preferred Networks, Inc., Tokyo, Japan ‡ Kobe University, Kobe, Japan

Performance Analysis of the NICAM Benchmark on MN-Core Processor

Hikaru Takayashiki*, Natsuko Saito*, Hiroto Imachi†, Ryo Sakamoto† and Junichiro Makino†‡

Background

- Nonhydrostatic Icosahedral Atmospheric Model (NICAM)
 - Global atmospheric circulation simulations that uses the icosahedral division of a sphere
 - The main part is stencil calculations to apply differentiations
 - Memory bandwidth-demanding application in traditional supercomputers
- Near-memory computing
 - A promising technique for addressing memory wall problem
 - Places data in on-chip memory close to arithmetic unit to reduce off-chip access

• MN-Core Processor

- A near-memory processor that has large addressable memory near arithmetic unit
- Originally developed for deep learning and machine learning
- Can be attractive for bandwidth-demanding HPC applications



Objective & Approach

- Objective
 - Achieve high performance for the memory bandwidth-demanding HPC application
- Approach
 - Implement the benchmark kernels of NICAM to near-memory architecture processor
 - Give analysis of MN-Core performance for the benchmark kernels



MN-Core Architecture



- Distributed memory SIMD processor with hierarchical structure
 - Totally 8192 PE exist among 4 chips in a package
 - Processor Element (PE) has 4-way SIMD unit for double precision
 - Massively parallel arithmetic unit with no jump instruction
 - All loop structure must be unrolled at compile time
- No cache mechanism
 - Each PE has 74kB addressable local memory and 4.6kB register files
 - Addressable L1B memory / L2B memory



NICAM Benchmark Implementation

- Memory allocation to each PE
 - The computational space is a 4-dimensional structure, while stencil calculation is 3-dimensional
 - Each dimension is divided and assigned to hardware unit
 - Assignments where halo exchange is not a bottleneck
- Halo exchange for stencil calculation
 - Halo exchange between PEs is required because PE has distributed memory
 - Memory is hierarchical so that halo exchange is also hierarchical











Evaluation Result



MN-Core (on-chip)

When assuming the data are initially in the on-chip memory and written back to it

MN-Core (DRAM) When assuming the data are initially in DRAM and written back to DRAM

A64FX

Comparison for Fugaku supercomputer's processor

- On-chip benchmark shows very high efficiency (25-50% of the peak performance)
 - \circ On-chip memory bandwidth is high enough to match the computing power
- About half of peak performance in MN-Core (on-chip) in dyn diffusion/dyn vert adv limiter
 - In the real application, data between kernels could be passed in the near memory of MN-Core
- Higher performance than A64FX even in MN-Core(DRAM) in dyn_diffusion
 - programmer can explicitly place reusable data in near memory
- MN-Core (DRAM) is lower than A64FX in dyn_vert_adv_limiter due to less reusable data

Reflecting to the difference of memory bandwidth Copyright© Fixstars Group

Summary

- Near-memory architecture is promising for NICAM applications
 - MN-Core is distributed near-memory SIMD architecture processor
- Objective
 - Achieve high performance for the memory bandwidth-demanding HPC application
- Approach
 - Implement the benchmark kernels of NICAM to near-memory architecture processor
- Result
 - Obtained 986 GFLOPS, which is 13.4% of peak when we assume the data are initially in DRAM at max
 - Obtained 4,178 GFLOPS, which is 56.7% of the peak performance when we assume the data are initially in the on-chip memory at the maximum
 - Near-memory computing with distributed-memory SIMD architecture shows potential for high performance in memory-demanding HPC applications
- Future work
 - Implementing other kernels or applications to MN-Core





Thank you