Optimizations for the Himeno Benchmark on Vector Computing System SX-Aurora TSUBASA

Background

Importance of Vector Processing

- Recent CPUs improve computational performance by vector processing
  - ✓ NEC SX-Aurora TSUBASA, Intel AVX-512, ARM SVE, etc.

→ **The use of vector processing is necessary for accelerating programs**

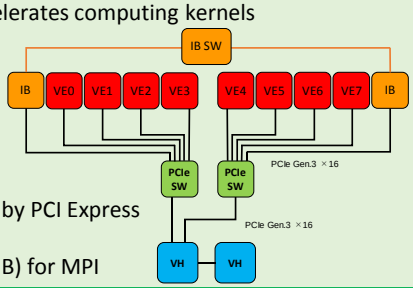
- Many applications can be calculated by using vector processing
  - ✓ Large-scale numerical simulations, big data analysis, etc.

Vector Computing System SX-Aurora TSUBASA

- Composed of Vector Hosts (VHs) and Vector Engines (VEs) [1]
  - ✓ VH: Standard x86 processor that is responsible for executions of OS-related tasks
  - ✓ VE: Vector processor that accelerates computing kernels
- Applications are executed on VEs
  - ✓ Only when applications need system calls, VEs request VHs to execute them

◎ System Overview of A300-8

- 2 VHs and 8 VEs
- One of VH and VEs are connected by PCI Express
  - ✓ Via 2 PCIe switches
- VEs are connected by Infiniband (IB) for MPI



◎ Architecture of a VE

- 8-vector cores
- Vector registers for 256-elements data

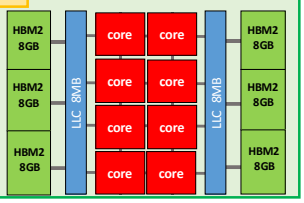
Need to utilize the multiple cores and the long vector length

- High bandwidth Last Level Cache (LLC)

Need to utilize the LLC bandwidth

- 6-module High Bandwidth Memory (HBM2)

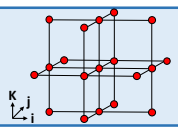
Need to utilize the main memory bandwidth



Optimizations for the Himeno Benchmark

Overview of the Himeno Benchmark

- The benchmark solves Poisson's equation using the Jacobi iteration method [2]
- Jacobi iteration method
  - ✓ 19-point stencil calculation nested by i, j, and k
  - ✓ The memory-intensive kernel (Code B/F: 3.74)



Three Optimizations for VE

① Store highly reusable data in the LLC

- ✓ Utilize the LLC whose bandwidth is higher than that of HBM2
- ✓ An array of the pressure variables is referenced 19 times in the kernel
  - ➢ Place the array in the LLC on a priority basis

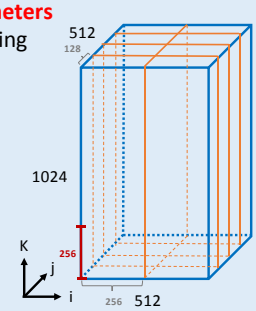
② Reduce the loop overheads

- ✓ Apply loop unrolling
  - ➢ The kernel has long loops with nested structures
  - ➢ Use of plenty of vector registers of vector processors
    - Set the unroll time as large as possible without register spilling

③ Tune the domain decomposition parameters

- ✓ Important for efficient vector processing
  - ➢ Longer vector length
  - ➢ Higher LLC hit ratio

- ☐ Keeping the length in the k direction larger than 256, while
- ☐ Increasing the decomposition of the j direction compared with i direction to achieve the high LLC hit ratio

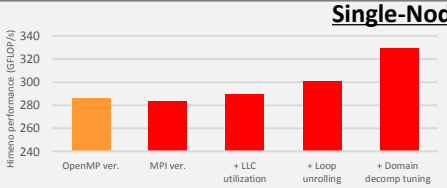


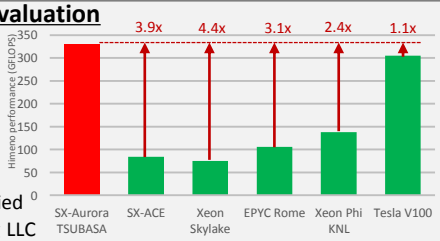
Performance Evaluation

Experimental Environment

- SX-Aurora TSUBASA A300-8
  - ✓ VH: Intel Xeon Gold 6126 x 2
    - ➢ Peak performance / socket: 1.0 TFLOPS
    - ➢ Peak memory bandwidth: 0.13 TB/s
  - ✓ VE: Type 10B x 8
    - ➢ Peak performance / node: 2.15 TFLOPS
    - ➢ Peak memory bandwidth / node: 1.2 TB/s
    - ➢ Peak LLC bandwidth / node: 3.2 TB/s
  - ✓ Infiniband FDR
- Cent OS 7.5.1804
- VEOS 2.4.0
- NEC C/C++ Compiler 3.0.1
- NEC MPI 2.5.0
- Himeno benchmark version 3.0
  - ✓ C language
  - ✓ MPI static allocate version
  - ✓ Problem size: XL (1024x512x512)
  - ✓ Initial decomposition: (i,j,k)=(2,2,2)

Single-Node Evaluation

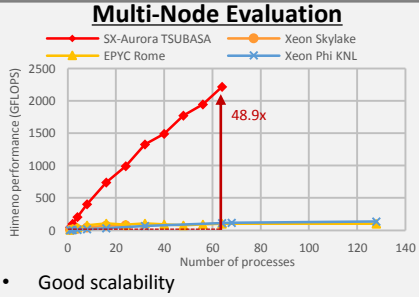




- The performance improves as each optimization is applied
  - ✓ LLC hit ratio increases from 44.3% to 49.6% by utilizing LLC
  - ✓ Overhead decreases greatly by loop unrolling
  - ✓ Achieve vector length of 255 by tuning the decomposition
- 1.15x performance improvement compared with that of OpenMP version [3]
- 15% of the peak performance in the MPI version

- SX-Aurora TSUBASA achieves the highest performance
  - ✓ Utilize HBM2 and LLC bandwidth efficiently

Multi-Node Evaluation



- Good scalability
  - ✓ 48.9x speedup and 76% parallel efficiency with 64 processes
- 13% of the peak performance with 8VEs

Conclusions & Future Work

- SX-Aurora TSUBASA achieves high performance of the Himeno benchmark by the optimizations
  - ✓ Important to execute efficient vector processing with effective use of hardware resources
- Hybrid execution with VHs and VEs has the potential to achieve higher performance

References

[1] Y. Yamada et al., "Vector Engine Processor of NEC's Brand-New Supercomputer SX-Aurora TSUBASA", Hot Chips 30, Aug 2018

[2] Himeno benchmark, http://i.riken.jp/en/supercom/documents/himenobmt/

[3] K. Komatsu et al., "Performance Evaluation of a Vector Supercomputer SX-Aurora TSUBASA", SC18, Nov 2018