# Early Evaluation of the SX-ACE Processor

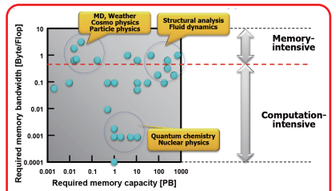
Ryusuke Egawa[a), d)], Shintaro Momose [b), a)], Kazuhiko Komatsu[a), d)], Yoko Isobe[b), a)], Akihiro Musa[b), a)], Hiroyuki Takizawa[c), d)], and Hiroaki Kobayashi[a)]
[a)]Cyberscience Center, Tohoku University, [b)] NEC Corporation, [c)]GSIS Tohoku University, [d)]JST CREST

## Background & Motivation

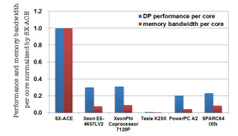
In November 2013, inheriting the advantages of conventional vector processors, SX-ACE has been launched to provide a high memory bandwidth commensurate with its high computational capability. The SX-ACE processor is designed so as to achieve
- a high sustained memory bandwidth for accelerating memory-intensive applications,
- a high single core performance for obtaining a high sustained performance with fewer cores,

under the limited power and silicon budgets.

**However, the potential of SX-ACE for practical applications is not clear.**




While the requirements for HPC supercomputer system are diversified, several important applications require a certain memory bandwidth.



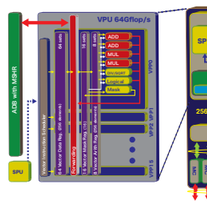

The SX-ACE System and Processor

Objective :
Unveiling the potential of the SX-ACE processor using practical applications!

## An Overview of the SX-ACE Processor

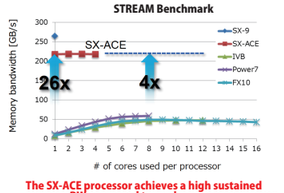
The SX-ACE processor employs a multi vector-cores architecture, which provides a high memory bandwidth with a high computational performance. As the successor of SX-9, SX-ACE introduces several architectural features as follows:

1. A multi vector-cores architecture
2. ADB and MSHR mechanisms to keep a high sustained memory bandwidth
3. Out-of order vector memory access mechanisms to improve the performance of indirect/stride memory accesses
4. A shorter memory access latency and a shorter vector pipe latency in chaining to realize efficient short-vector processing




Core Architecture

The SX-ACE Processor




**The SX-ACE processor achieves a high sustained memory BW compared to modern processors, even in the single core case !**

[1] Shintaro Momose, et.al., "The Brand-New Vector Supercomputer, SX-ACE," ISC' 14, June 2014.

### Specifications

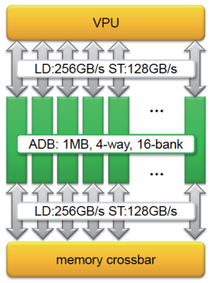
| | |
|---|---|
| Base Architecture | Multi Vector-cores |
| Core | |
| Clock Frequency | 1.0GHz |
| Peak Performance | 64Gflop/s |
| ADB Size | 1MB |
| ADB Bandwidth | 256GB/sec |
| Memory Bandwidth | 64GB/sec(4 cores) - 256GB/sec(1 core) |
| B/F | 1.0(4 cores) ~ 4.0(1 core) |
| CPU | |
| Number of Cores | 4 |
| Peak Performance | 256Gflop/s |
| Memory Bandwidth | 256GB/sec |
| B/F | 1.0 |
| Process Technology | 28nm |
| Chip Size | 23.05mm x 24.75mm |

### ADB & MSHR

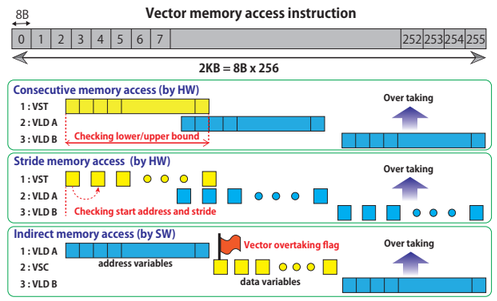
**ADB (Assignable Data Buffer)**
- Private, 1MB, 4-Way, 16-Bank
- 256GB/s bandwidth
- Software controllable
- Customized for fast random access

**Software Control Function**
- Compiler/User control
- A bypass flag in each instruction
- Avoiding cache pollution

**MSHR (with 512 entries)**
(Miss Status Handling Registers)
- Reducing redundant memory accesses by holding the information of in-flight memory accesses




### Out-of-order Vector Memory Access Mechanisms




## Evaluation Setup

### Evaluated Systems

| System | Perf./Socket (Gflop/s) | No. of Cores | Perf. /core (Gflop/s) | Mem. BW GB/sec | On-chip mem | NW BW (GB/sec) | Sys. B/F |
|---|---|---|---|---|---|---|---|
| SX-ACE | 256 | 4 | 64 | 256 | 1MB ADB /core | 2 x 4 IXS | 1.0 |
| SX-9 | 102.4 | 1 | 102.4 | 256 | 256KB ADB/core | 2 x 128 IXS | 2.5 |
| LX 406 | 230.4 | 12 | 19.2 | 59.7 | 256KB L2/core 30MB Shared L3 | 5 IB | 0.26 |
| SR16K M1 | 245.1 | 8 | 30.6 | 128 | 256KB L2/core 32MB Shared L3 | 2 x 24 - 96 custom NW | 0.52 |

### Benchmark Programs

| Applications | Fields | Methods | Mem Access Characteristics | Mesh Size | Code B/F |
|---|---|---|---|---|---|
| QSFDM GLOBE | Seismology | Spherical 2.5D FDM | Stencil with sequential memory accesses | $4.3 \times 10^7$ | 2.16 |
| Barotropic ocean | OGCM (Ocean General Circulation Model) | Shallow Water Model | Stencil with sequential memory accesses | 4322 x 216 | 1.97 |
| MHD (FDM) | MHD | Finite Difference Method | Stencil with sequential memory accesses | 200 x 1920 x 32 | 3.04 |
| SEISM3D | Seismology | Finite Difference Method | Stencil with sequential memory accesses | 1024 x 512 x 512 | 2.15 |
| MHD (Spectral) | MHD | Pseudo pectoral Method | Stride memory access | 900 x 768 x 96 | 2.21 |
| TURBINE | CFD | DNS | Indirect memory access with short vectors | 91 x 91 x 91 x 13 | 1.78 |
| BCM | CFD | Navier Stokes Equation | Stencil and iindirect memory access | (128 x 128 x 128 cells) x 64 Cubes | 7.01 |

### Definitions of Bytes/Flop (B/F) Ratios

To analyze the actual memory access behavior, three B/F ratios are considered. The system B/F is given by the system specification. While the code B/F can be obtained by analyzing application codes, this value does not always reflect actual memory access behavior due to the block data transfer between CPU and the memory subsystem. Therefore, the actual B/F is defined and used to analyzed the sustained performance in this study.

**System B/F = Peak Memory Bandwidth / Peak Performance**

$$\text{Code B/F} = \frac{\#\text{ of mem. accesses} \times \text{Data size}}{\text{Flops}}$$

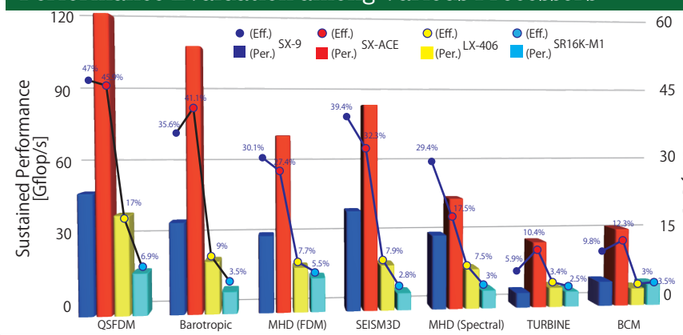
(Obtaining by analyzing codes)

$$\text{Actual B/F} = \frac{\#\text{ Of mem. accesses} \times (1\text{- hit rate}) \times \text{Data size} \times N_{bl}}{\text{Flops}}$$

$N_{bl}$ : # of elements transferred by block accesses

## Potential of SX-ACE

### Performance Evaluation among Various Processors




| | QSFDM | Barotropic | MHD (FDM) | Seism3D | MHD (Spectral) | TURBINE | BCM |
|---|---|---|---|---|---|---|---|
| Code B/F | 2.16 | 1.97 | 3.04 | 2.15 | 2.21 | 1.78 | 7.01 |
| Actual B/F | 0.78 | 1.11 | 1.41 | 1.68 | 2.18 | 5.47 | 5.86 |

Computation intensive ⟷ Memory intensive

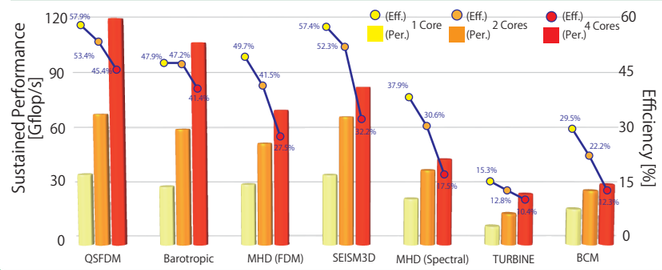
**The SX-ACE processor provides the higher sustained performance vs. scalar systems**

Since the actual B/F ratios of the applications are high, the memory performance of each system greatly affects the sustained performance of the applications. Thus, SX-ACE can achieve about 3.9x and 7.2x higher sustained performances than those of LX 406 and SR16K M1, respectively.

**vs. SX-9**

While the system B/F ratio degrades to 1.0 from 2.5 of SX-9, SX-ACE achieves higher sustained performance than SX-9. In the computation-intensive cases (QSFDM, Baratropic), the sustained performances depend on the peak performance. In the memory intensive cases(TURBINE, BCM), the new mechanism, such as ADB with MSHR, out of order vector load mechanisms, shorter memory latency work effectively to keep a high sustained performance with a high efficiency.

### Core Scalability on the SX-ACE Processor




**Steady core scalability :**
**The sustained performance increases as the number of core increases, due to a constantly high sustained memory bandwidth.**

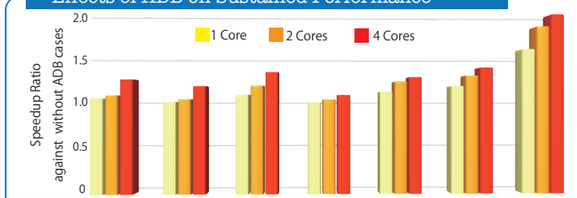
- As QSFDM and Barotropic are computation-intensive, a high core scalability with relatively high efficiencies is confirmed.

- Since the sustained performances of MHD (Spectral), Turbine and BCM are limited by off-chip memory bandwidth, their efficiencies are low. However, due to the new architectural features, these applications still keep high core scalability.

- High cost kernels of SEISM3D have well-balanced add and multiply operations, which can effectively fill the vector pipelines. Thus, SEISM3D achieves better performance compared with MHD(FDM), even though their actual B/F ratios are almost the same.

### Evaluations of the ADB & MSHR

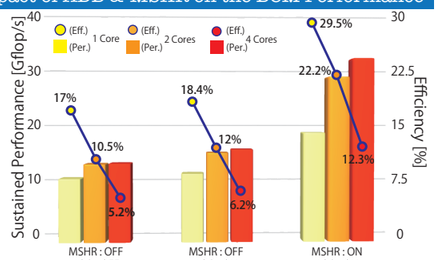
#### Effects of ADB on Sustained Performance




| | QSFDM | Barotropic | MHD (FDM) | SEISM3D | MHD (Spectral) | TURBINE | BCM |
|---|---|---|---|---|---|---|---|
| Hit Rate | 74.4% | 50.8% | 68.6% | 30% | 70.9% | 33.1% | 47.4% |

**For all the benchmark programs, ADB works more effectively for the 4-core configuration than 1- and 2- core ones.**
- Memory pressure increases with the number of cores
- ADB works well even though their hit-rates are not so high

#### Impact of ADB & MSHR on the BCM Performance

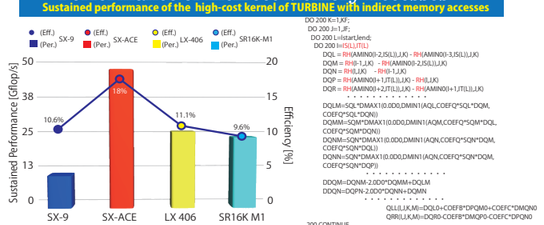



**The core scalability is remarkably improved by simultaneously enabling ADB and MSHR.**
- MSHR exploits the locality of vector load operations in the innermost loop

### Evaluations of the new features of SX-ACE - Overcoming the weakness of conventional vector architectures -

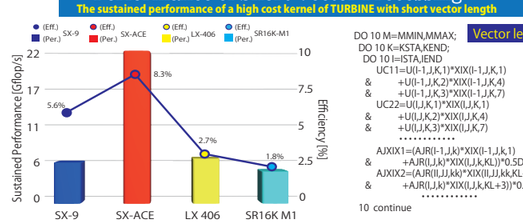
#### Performance of Indirect Memory Accesses

**Sustained performance of the high-cost kernel of TURBINE with indirect memory accesses**




#### Performance of Short Vector Processing

**The sustained performance of a high cost kernel of TURBINE with short vector length**




Vector length 46

```
DO 10 M=MMIN,MMAX;
DO 10 K=KSTA,KEND;
DO 10 I=ISTA,IEND
  UC11=U(I-1,J,K,1)*XIX(I-1,J,K,1)
 &     +U(I-1,J,K,2)*XIX(I-1,J,K,4)
 &     +U(I-1,J,K,3)*XIX(I-1,J,K,7)
  UC22=U(I,J,K,1)*XIX(I,J,K,1)
 &     +U(I,J,K,2)*XIX(I,J,K,4)
 &     +U(I,J,K,3)*XIX(I,J,K,7)
  ..............
  AJXIX1=(AJR(I-1,J,k)*XIX(I-1,J,k,1)
 &     +AJR(I,J,k)*XIX(I,J,k,KL))*0.5D0
  AJXIX2=(AJR(I,J,kk)*XIX(II,J,kk,KL+3)
 &     +AJR(I,J,k)*XIX(I,J,k,KL+3))*0.5D0
  ..............
10 continue
```

**The memory subsystem with a shorter latency, the out of order memory access operations, the data forwarding mechanism for a shorter latency in pipeline chaining, and ADB accelerate indirect memory accesses and short vector processing**

**While SX-ACE, LX460, and SR16K have comparable peak performance, SX-ACE achieves the highest sustained performance in the both cases!!**

## Conclusions and Future Work

To clarify a new-generation vector processor, SX-ACE, this poster examines the sustained performance of the SX-ACE processor using practical scientific and engineering applications. Evaluation results indicate that the SX-ACE processor has significant advantages against modern scalar processors and conventional vector processors. The newly introduced advanced memory subsystem, which employs ADB with MSHR and out-of-order memory access functions, effectively accelerates memory-intensive applications with indirect memory accesses and stride memory accesses. In addition, while inheriting the advantages of conventional vector processors, the SX-ACE processor achieves a higher sustained performance on processing short vector operations by shortening the memory access latencies and the chaining operation latencies.
Our future work includes performance evaluations of the multi-node system, power evaluation, and performance comparison with GPUs/accelerators.