Performance Evaluation of a Vector Supercomputer SX-Aurora TSUBASA

Kazuhiko Komatsu, S. Momose, Y. Isobe, O. Watanabe,
A. Musa, M. Yokokawa, T. Aoyama, M. Sato, H. Kobayashi

Tohoku University

14 November, 2018

SC18

Outline

- Background
- Overview of SX-Aurora TSUBASA
- Performance evaluation
  - Benchmark performance
  - Application performance
- Conclusions

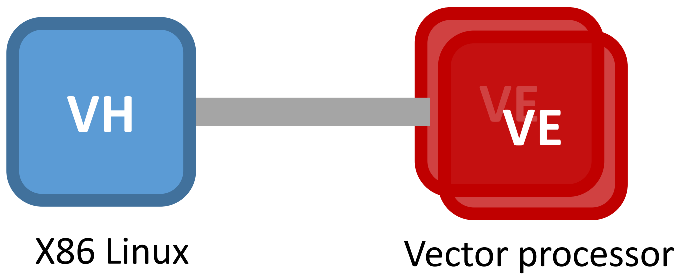
Background

- Supercomputers become important infrastructures
  - Widely used for scientific researches as well as various industries
  - Top1 Summit system reaches 143.5 Pflop/s
- Big gap between theoretical performance and sustained performance
  - ◎ **Compute-intensive** applications stand to benefit from high peak performance
  - ✖ **Memory-intensive** applications are limited by lower memory performance

Memory performance has gained more and more attentions

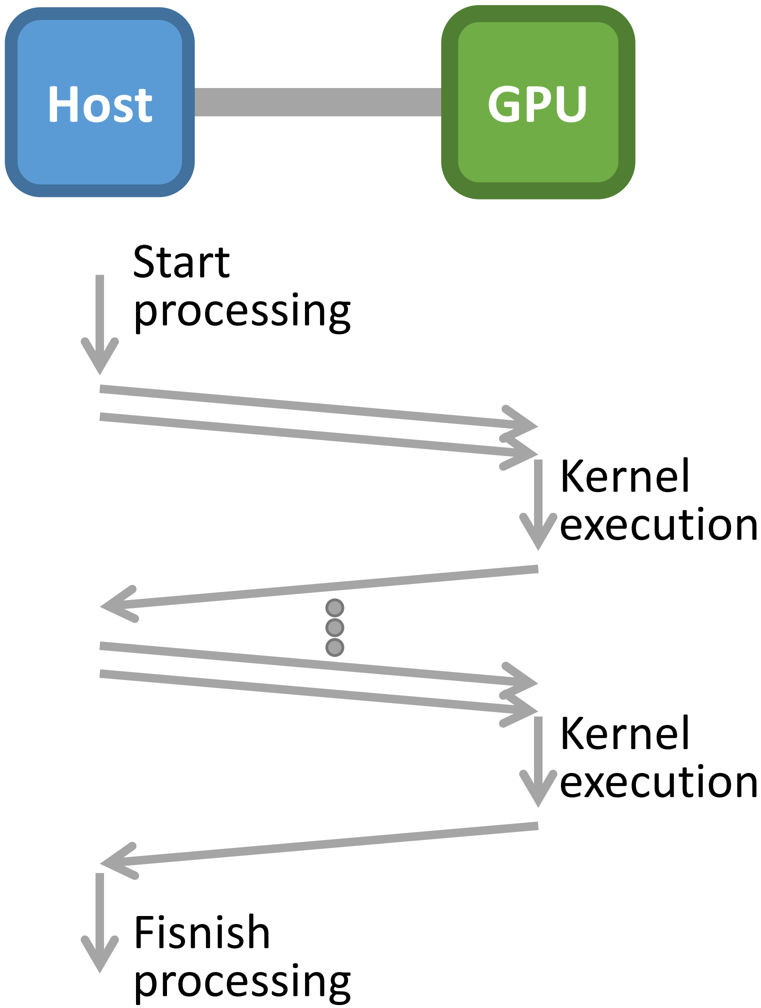
A new vector supercomputer
SX-Aurora TSUBASA

- Two important concepts of its design
  - **High usability**
  - **High sustained performance**
- New memory integration technology
  - Realize the world's highest memory bandwidth
- New architecture
  - **Vector host (VH)** is attached to **vector engines (VEs)**
    - VE is responsible for executing an entire application
    - VH is used for processing system calls invoked by the applications

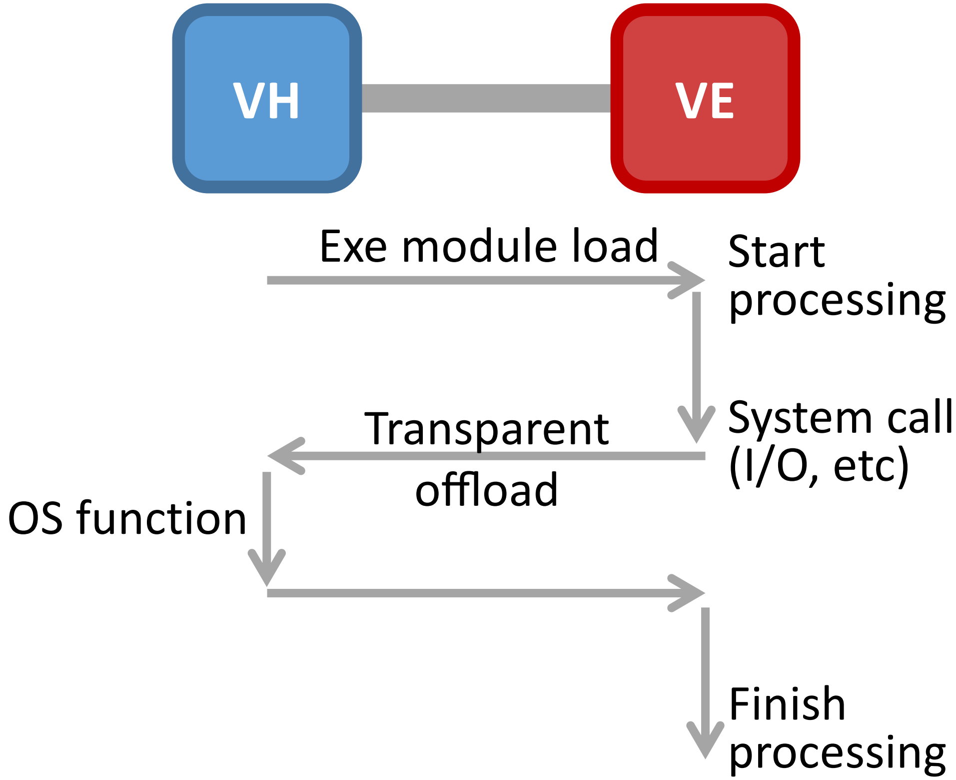


New execution model

- Conventional model



- New execution model



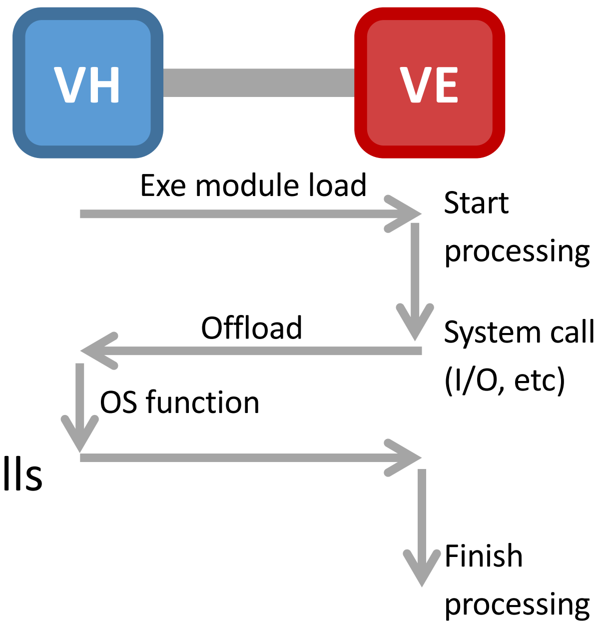
Highlights of the execution model

- Two advantages over conventional execution model
  - Avoid frequent data transfers between VE and VH
    - Applications are entirely executed on VE
    - Only necessary data for system calls need to be transferred

  →**High sustained performance**

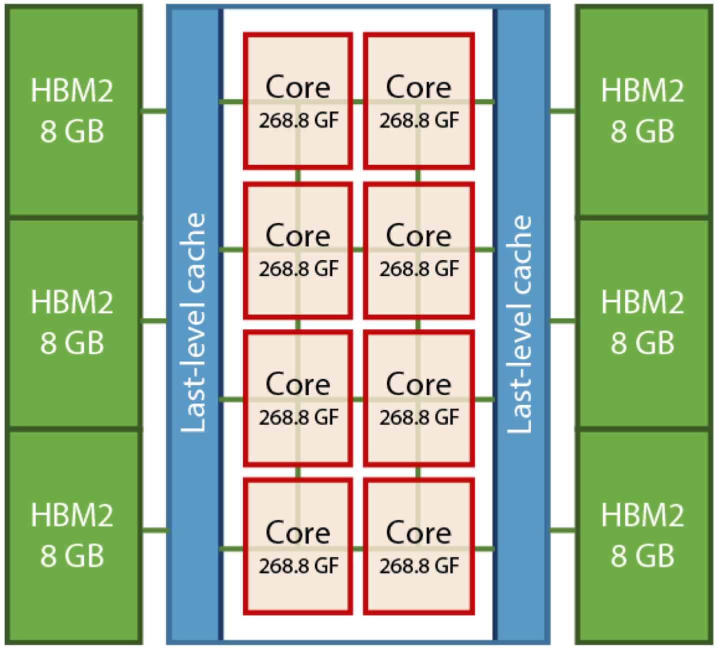
  - No special programming
    - Explicit specifications of computation kernels are not necessary
    - System calls are transparently offloaded to the VH
      - Programmers do not need to care system calls

  →**High usability**



Specification of SX-Aurora TSUBASA

- Memory bandwidth
  - 1.22 TB/s world's highest memory bandwidth
    - Six HBM2 memory modules integration
  - 3.0 TB/s LLC bandwidth
    - LLC is connected to cores via 2D mesh network
- Computational capability
  - 2.15 Tflop/s@1.4 GHz
    - 8 powerful vector cores
    - 16 nm FINFET process technology
    - 4.8 billion transistors
    - 14.96 mm x 33.00 mm



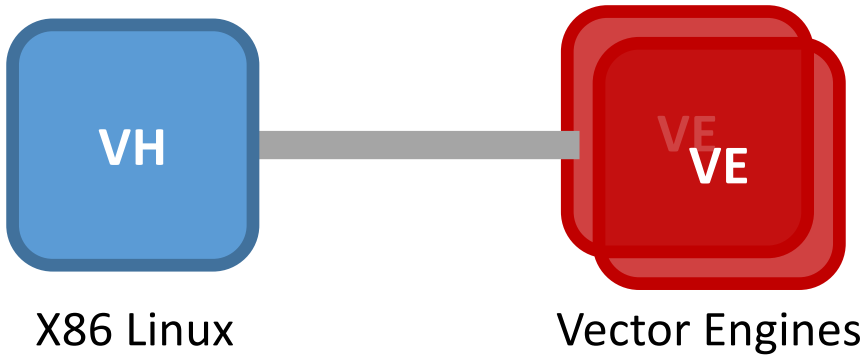
Block diagram of a vector processor

Outline

- Background
- Overview of SX-Aurora TSUBASA
- Performance evaluation
  - Benchmark performance
  - Application performance
- Conclusions

Experimental environments

- SX-Aurora TSUBASA A300-2
  - 2x VEs Type 10B
  - 1x VH



VH	Intel Xeon Gold 6126
Frequency	2.60 GHz / 3.70 GHz (Turbo)
Peak FP / core	83.2 Gflop/s
# cores	12
Peak DP Flops	998.4 Gflop/s
Mem BW	128 GB/s
Mem Capacity	96 GB
Mem config	DDR4-2666 DIMM 16GB x 6

VE	Type 10B
Frequency	1.4 GHz
Peak FP / core	268.8 Gflop/s
# cores	8
Peak DP Flops / socket	2.15 Tflop/s
Memory BW	1.2 TB/s
Memory capacity	48 GB
Memory config	HBM2

Experimental environments cont.

Processor	SX-Aurora Type 10B	Xeon Gold 6126	SX-ACE	Tesla V100	Xeon Phi KNL 7290
Frequency	1.4 GHz	2.6 GHz	1.0 GHz	1.245 GHz	1.5 GHz
# of cores	8	12	4	**5120**	72
DP flop/s (SP flop/s)	2.15 T (4.30 T)	998.4 GF (1996.8 GF)	256 GF	**7 TF (14 TF)**	3.456 TF (6.912 TF)
Memory subsystem	**HBM2 x6**	DDR4 x6ch	DDR3 x16ch	HBM2 x4	MCDRAM DDR4
Memory BW	**1.22 TB/s**	128 GB/s	256 GB/s	900 GB/s	450+ GB/s 115.2 GB/s
Memory capacity	48 GB	**96 GB**	64 GB	16 GB	16 GB 96 GB
LLC BW	**2.66 TB/s**	N/A	1.0 TB/s	N/A	N/A
LLC capacity	16 MB shared	19.25 MB shared	1 MB private	6 MB shared	1 MB shared by 2 cores

Applications used for evaluation

- SGEMM/DGEMM
  - Matrix-matrix multiplications to evaluate the Peak flop/s
- Stream benchmark
  - Simple kernels (copy, scale, add, triad) to measure sustained memory performance
- Himeno benchmark
  - Jacobi kernels with a 19-point stencil as a memory-intensive kernels
- Application kernels
  - Kernels of practical applications of Tohoku univ in Earthquake, CFD, Electromagnetic
- Microbenchmark to evaluate the execution model
  - Mixture with vector-friendly Jacobi kernels and I/O kernels

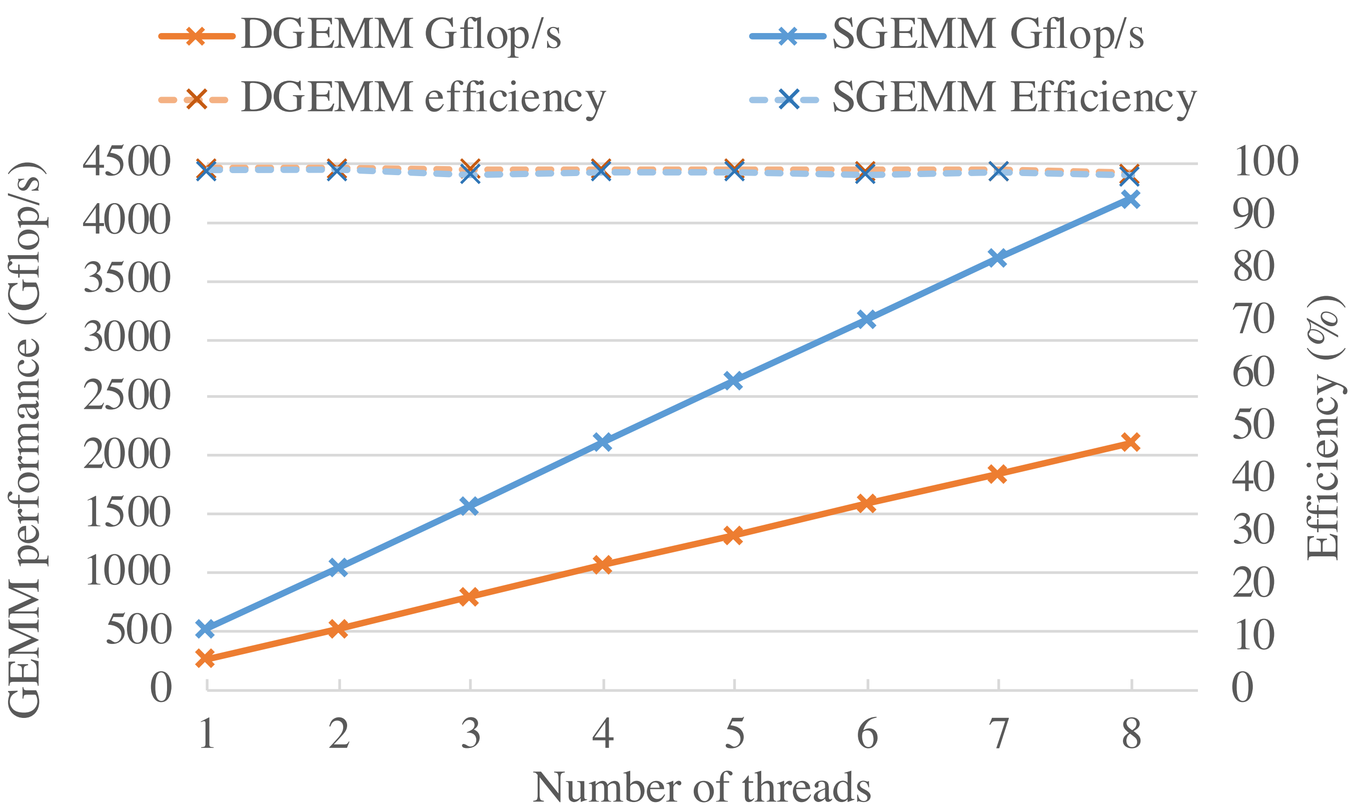
Overview of application kernels

Kernels	Fields	Methods	Memory access	Mesh size	Code B/F	Actual B/F
Land mine	Electro magnetic	FDTD	Sequential	100x750x750	6.22	5.15
Earthquake	Seismolo gy	Friction Law	Sequential	2047x2047x256	4.00	4.00
Turbulent Flow	CFD	Navier-Stokes	Sequential	512x16384x512	1.91	0.35
Antenna	Electro magnetic	FDTD	Sequential	252755x9x97336	1.73	0.98
Plasma	Physics	Lax-Wendroff	Indirect	20,048,000	1.12	0.075
Turbine	CFD	LU-SGS	Indirect	480x80x80x10	0.96	0.0084

Outline

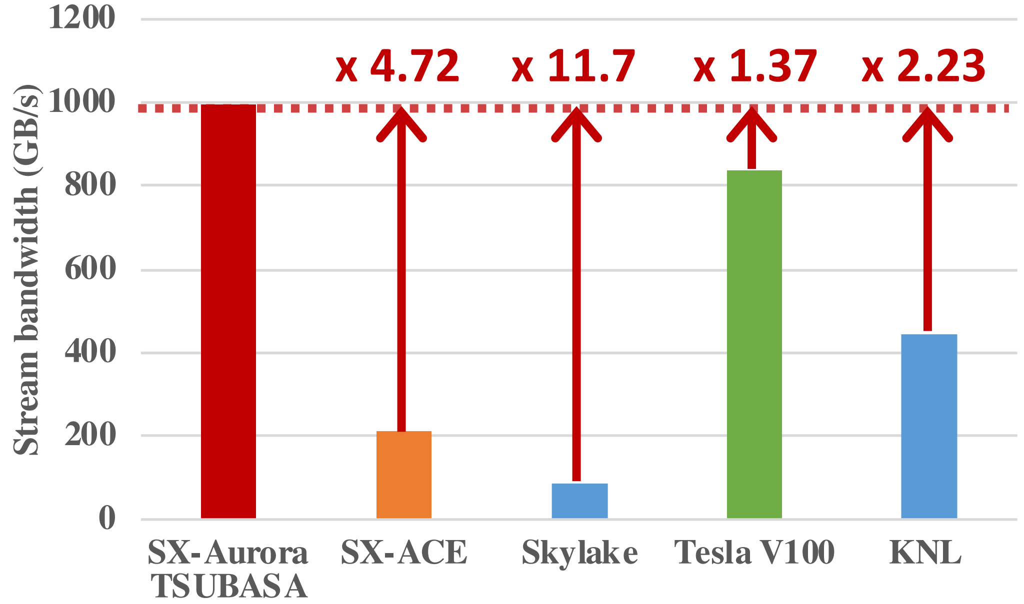
- Background
- Overview of SX-Aurora TSUBASA
- Performance evaluation
    - Benchmark performance
    - Application performance
- Conclusions

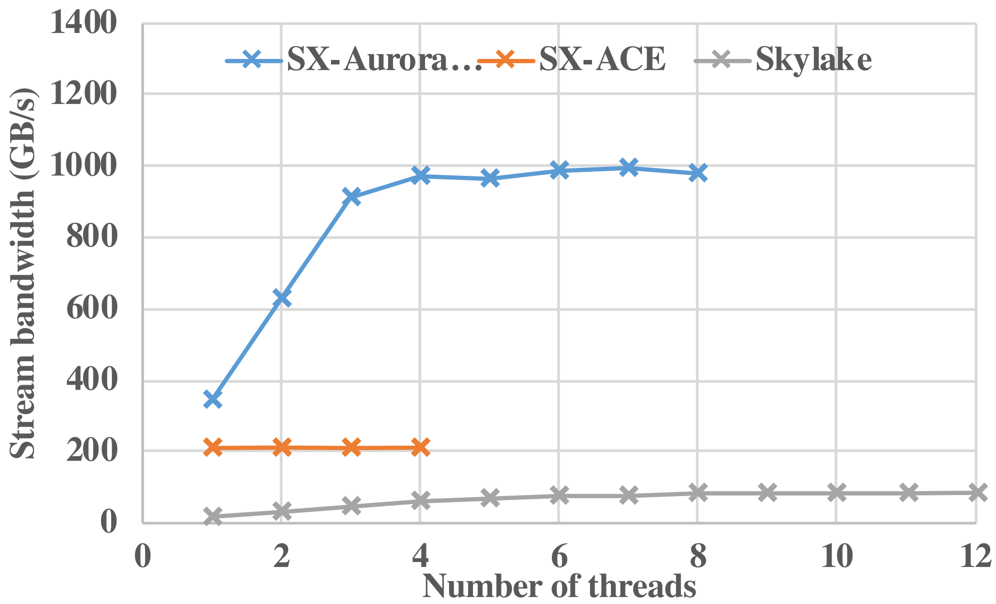
SGEMM/DGEMM Performance



- High scalability up to 8 threads
  - High vectorization ratio 99.36%, good vector length 253.8
- High efficiency
  - Efficiency 97.8~99.2%

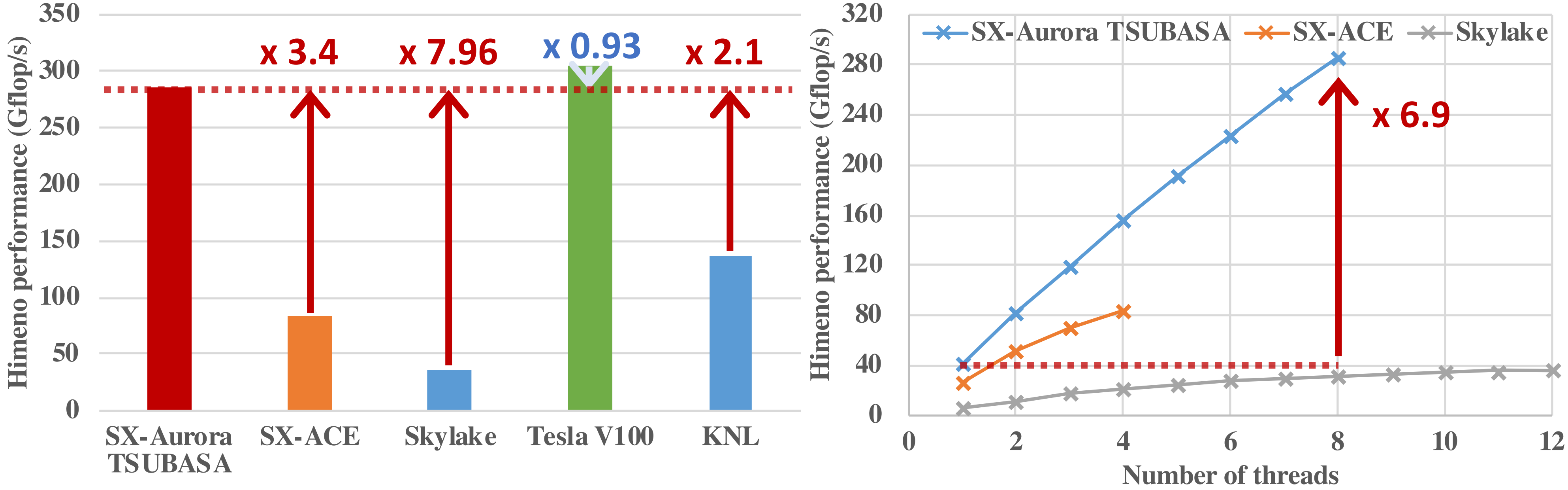
Memory performance(Stream Triad)





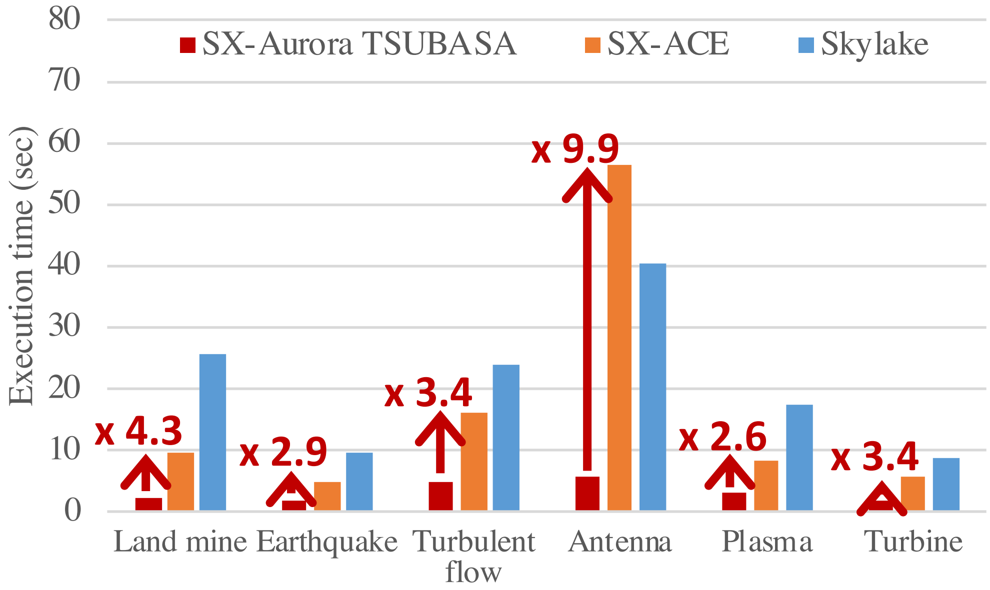
- High sustained memory bandwidth of SX-Aurora TSUBASA
  - Efficiency: Aurora 79%, ACE 83%, Skylake 66%, V100 81%
- Scalability
  - Saturated even when the number of threads is less than half

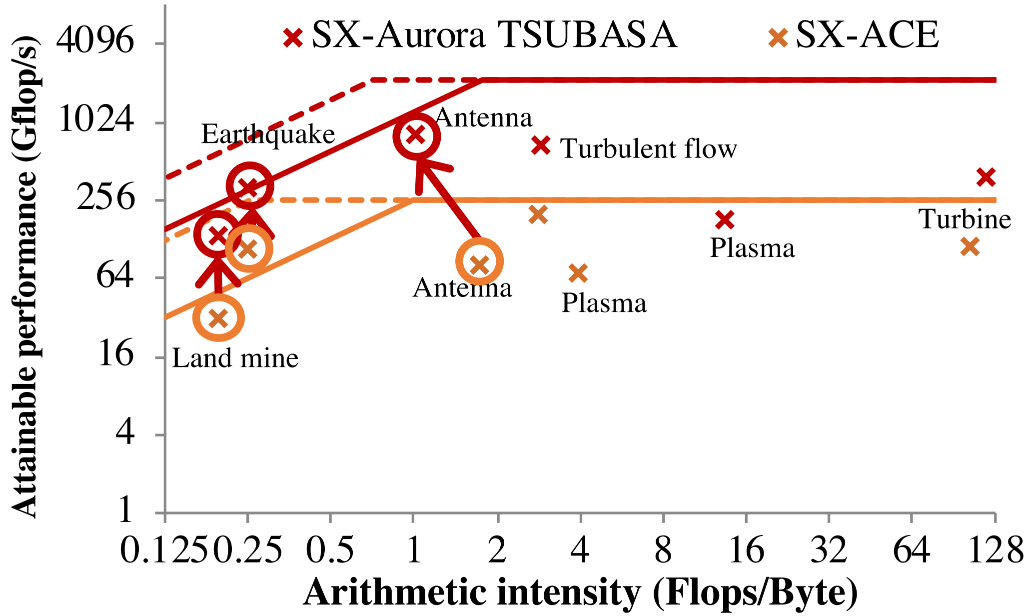
Himeno (Jacobi) performance



- Higher performance… except GPU
  - Vector reduction becomes bottleneck due to copy among vector pipes
- Nice thread scalability
  - 6.9x speedup in 8 threads => 86% parallel efficiency

Application kernel performance





- SX-Aurora TSUBASA could achieve high performance
  - Plasma, Turbine => Indirect access, memory latency-bound
  - Antenna => computation-bound to memory BW-bound
  - Land mine, Earthqauke, Turbulent flow =>memory or LLC BW-bound

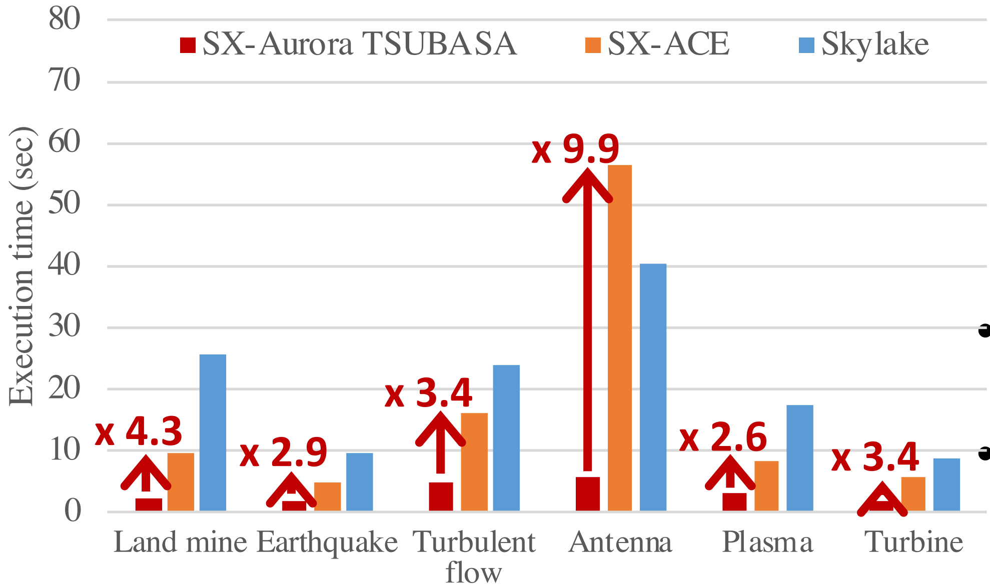
Memory bound? or LLC bound?

- Further analysis using 4 types of Bytes/Flop ratio
  - **Memory B/F** = (memory BW) / (peak performance)
  - **LLC B/F** = (LLC BW) / (peak performance)
  - **Code B/F** = (necessary data in Byte) / (# FP operations)
  - **Actual B/F** = (# block memory access) * (block size) / (# FP operations)

B/F ratio	Actual < Memory	Memory > Actual
Code < LLC	Computation-bound	Memory BW-bound
Code > LLC	LLC BW-bound	Memory or LLC bound *

  - Code B/F > Actual B/F * LLC BW / Memory BW => **LLC bound**
  - Code B/F < Actual B/F * LLC BW / Memory BW => **memory bound**

Application kernel performance



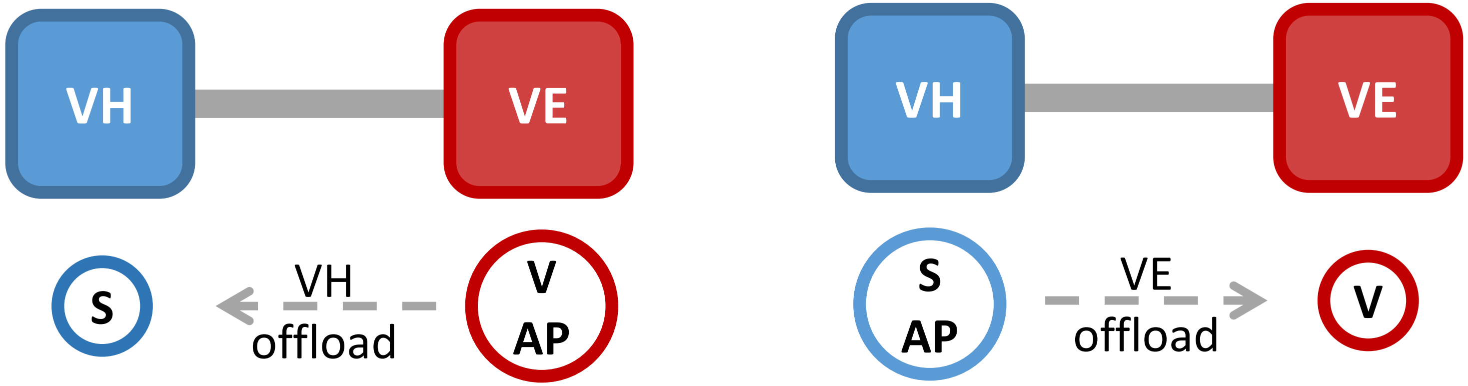
B/F ratio	Actual < Memory	Actual > Memory
Code < LLC	Computation-bound	Memory BW-bound
Code > LLC	LLC BW-bound	Memory or LLC bound

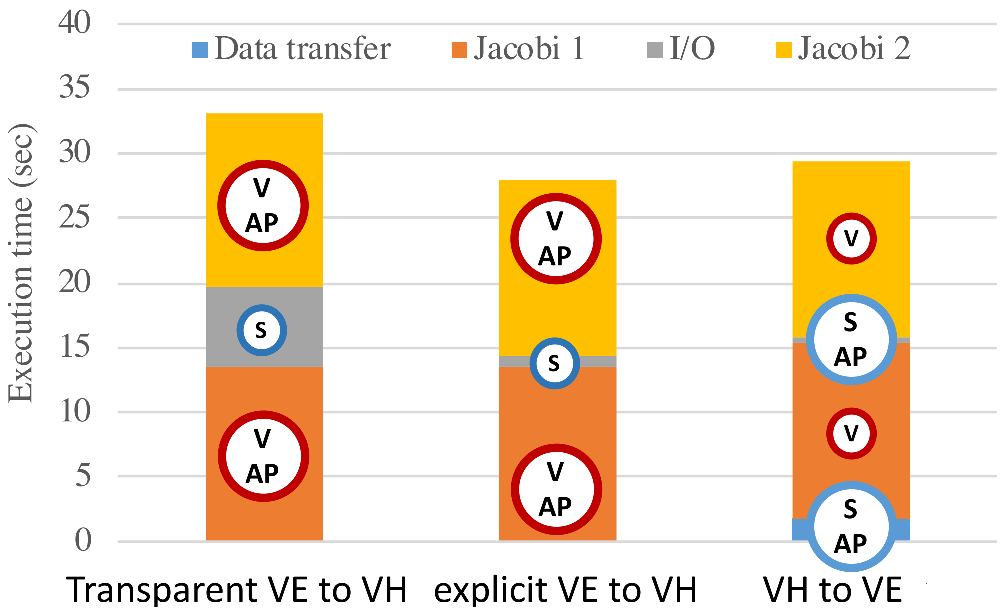
- Memory B/F = 1.22 TB/s / 2.15 TF
  = 0.57
- LLC B/F = 2.66 TB/s / 2.15 TF
  = 1.24

- Land mine (Code 6.22, Actual 5.79) => LLC bound
- Earthqauke (Code 6.00, Actual 2.00) => LLC bound
- Turbulent flow (Code 1.91, Actual 0.35) => memory BW bound
- Antenna (Code 1.73, Actual 0.98) => memory BW bound

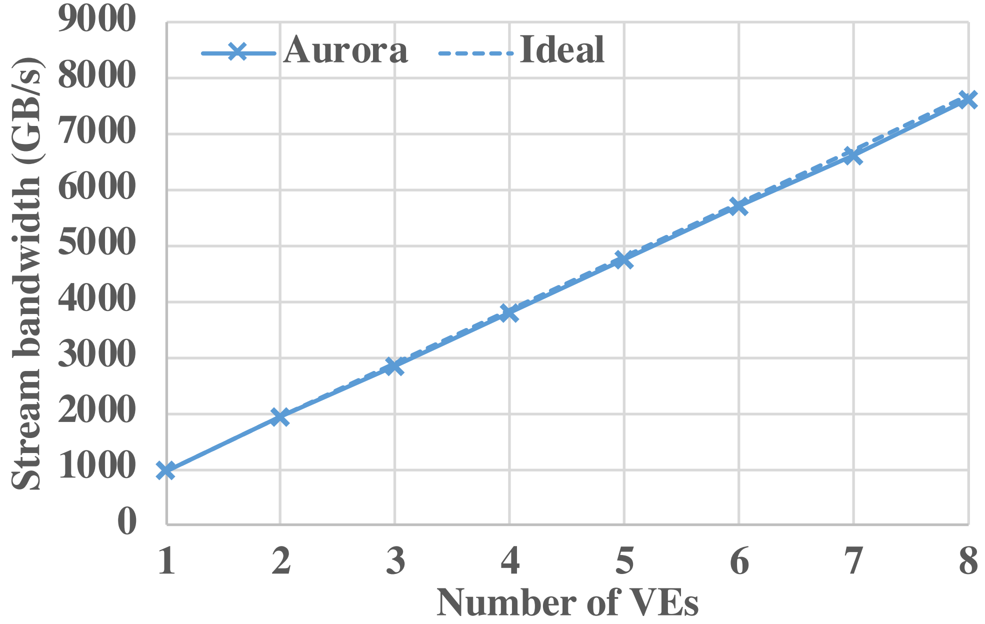
Evaluation of the execution model

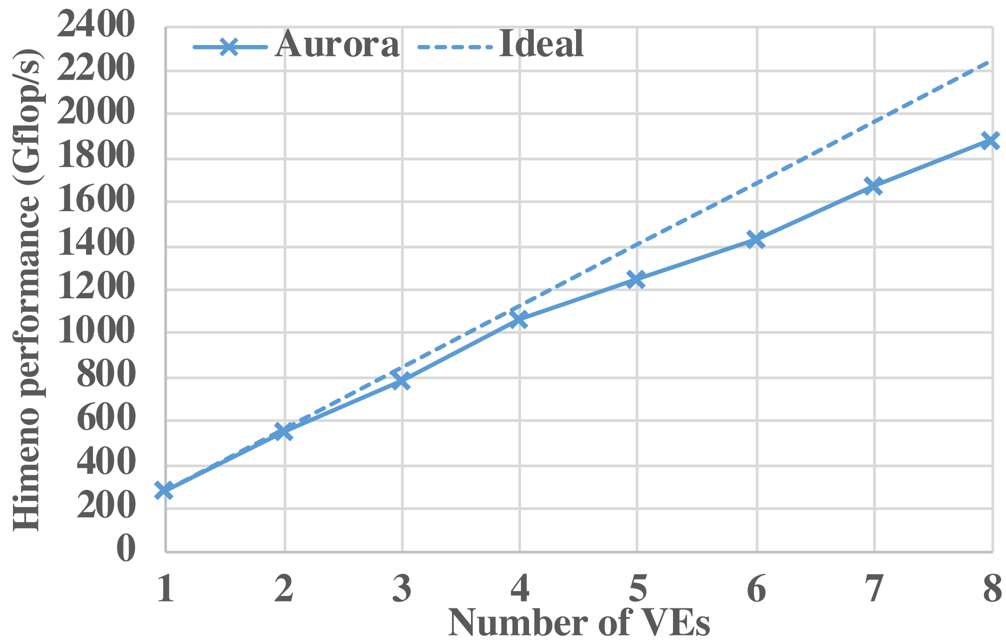
- (Transparent/Explicit) Offload from VE to VH
- Offload from VH to VE





Multi-VE performance on A300-8



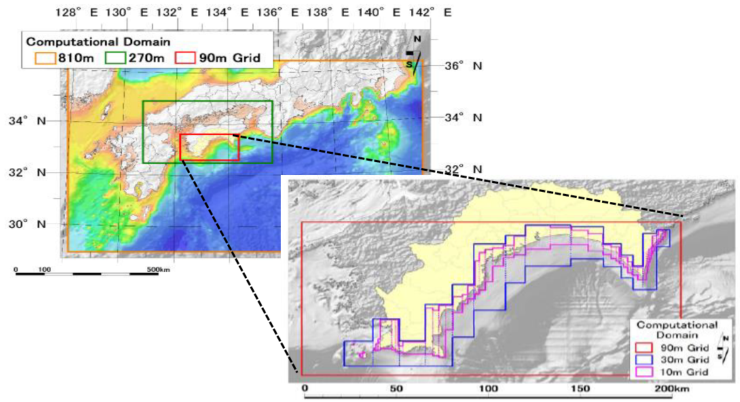


- Stream VE-level scalability
  - Almost ideal scalability up to 8 VEs
- Himeno VE-level scalability
  - Good scalability up to 4VEs
  - Lack of vector lengths when more than 5VEs
    - Problem size is too small

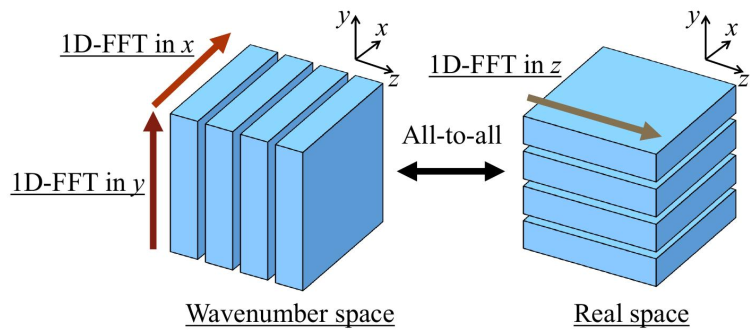
Outline

- Background
- Overview of SX-Aurora TSUBASA
- Performance evaluation
  - Benchmark performance
  - Application performance
- Conclusions

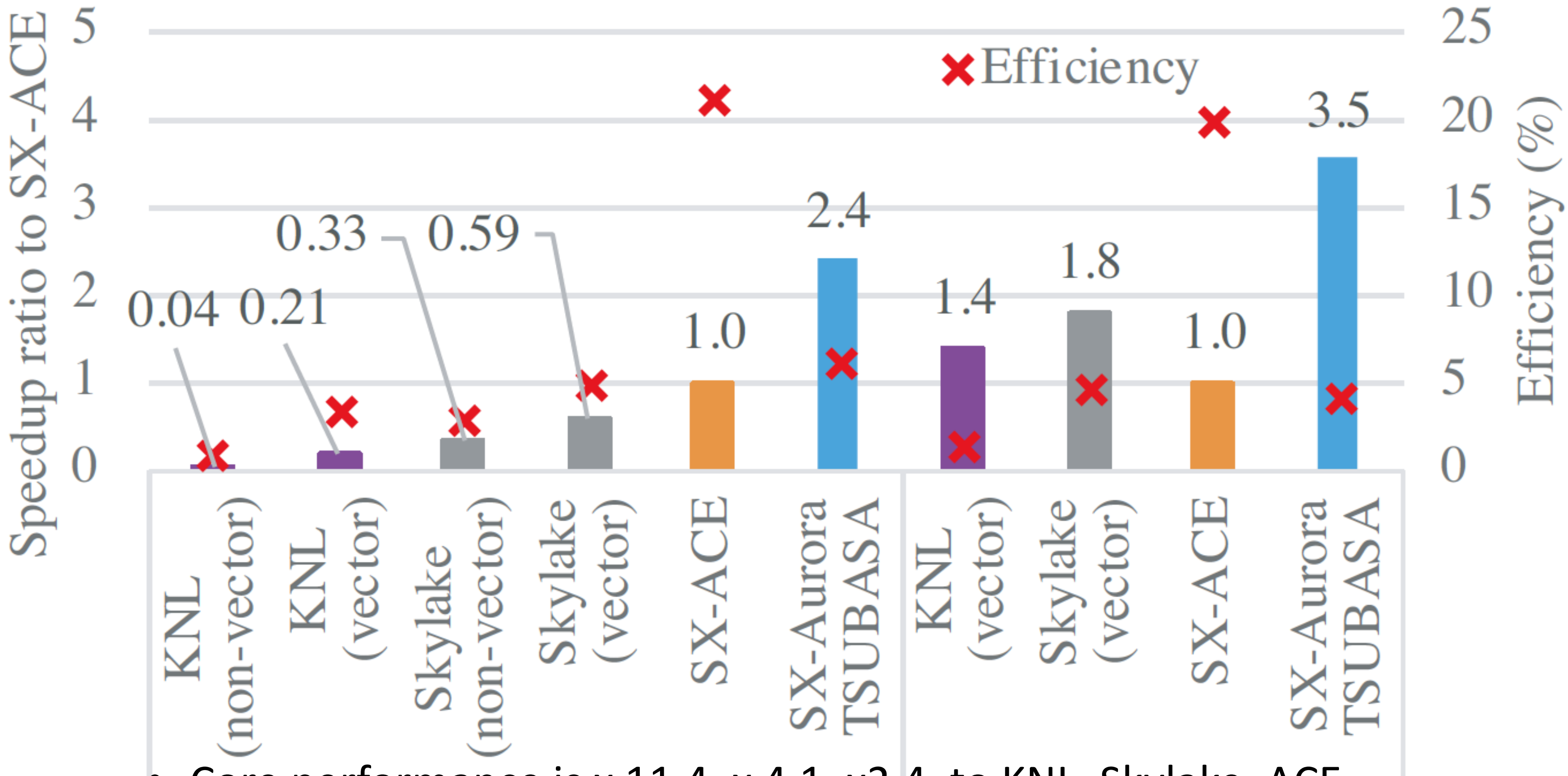
Application evaluation



- Applications
  - Tsunami simulation code
    - Mainly consists on ground fluctuation and Tsunami inundation prediction
    - Used for a real-time tsunami inundation system
    - A costal region of Japan(1244x826km) with an 810m resolution
  - Direct numerical simulation code of turbulent flows
    - Incompressible Navier-Stokes equations and a continuum equation are solved by 3D FFT
    - MPI parallelization to 4VEs (1~32 processes)
    - 128^3, 256^3, 512^3 grid points

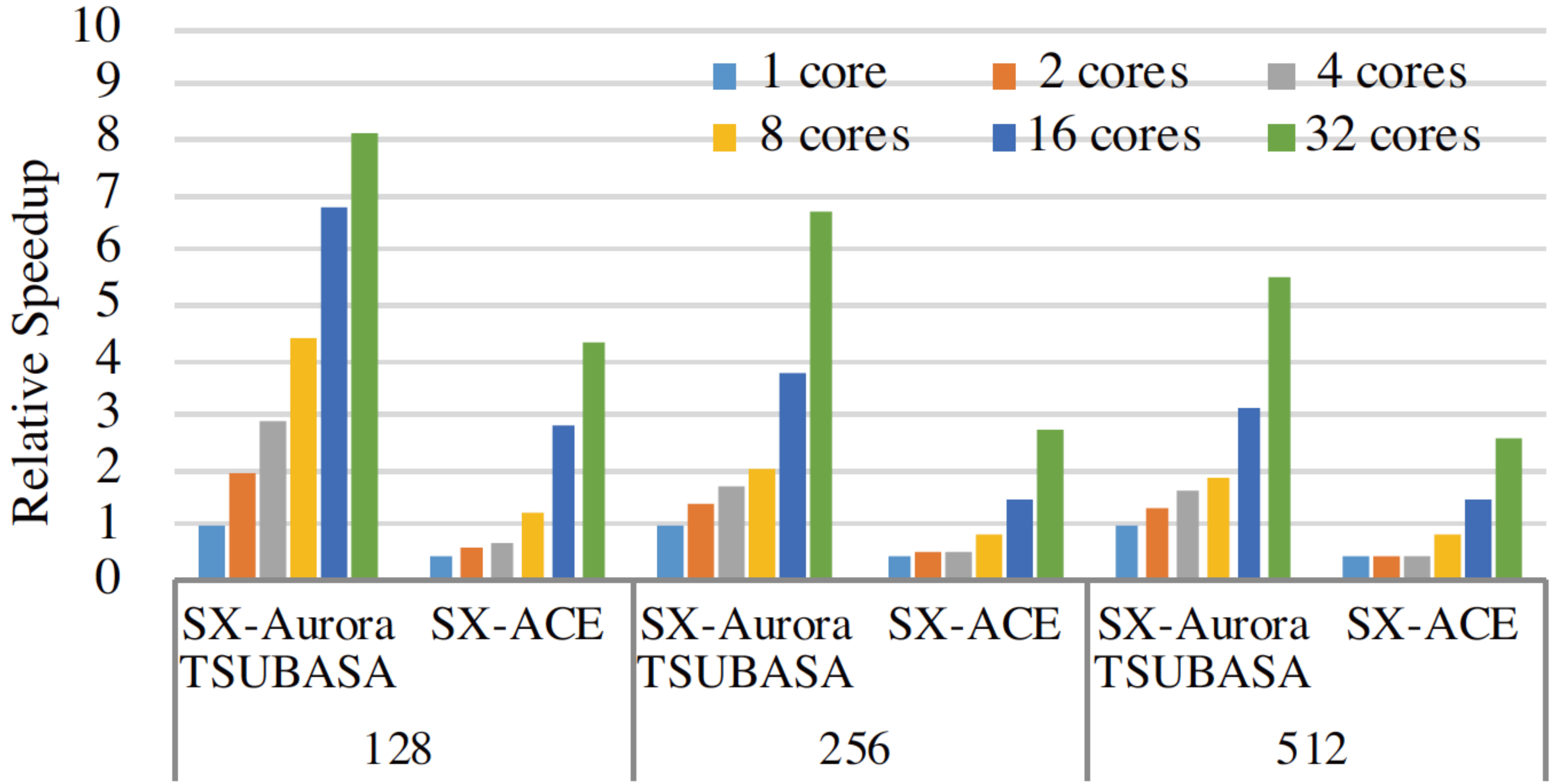


Performance of the Tsunami code



- Core performance is x 11.4, x 4.1, x2.4, to KNL, Skylake, ACE.
- Socket performance is x 2.5, x 1.9, x 3.5

Performance of the DNS code



- About 8.14, 6.73, 5.48 speedup by 32cores to 1core in 128^3, 256^3, 512^3 grids
  - LLC hit ratios of 256^3 and 512^3 are lower than that of 128^3
  - 10% is bank conflicts

Conclusions

- A vector supercomputer SX-Aurora TSUBASA
  - The highest bandwidth 1.22 TB/s by six HBM2 integration
  - New execution model to achieve **high usability** and **high sustained performance**
- Performance evaluation
  - Benchmark programs

    → **High sustained memory performance**

    → **effectiveness of a new execution model**
  - Application programs

    → **High sustained performance**
- Future work
  - Further optimizations for SX-Aurora TSUBASA

Acknowledgements

- People
  - Hiroyuki Takizawa
  - Ryusuke Egawa
  - Souya Fujimoto
  - Yasuhisa Masaoka
- Projects
  - The closed beta program for early access to Aurora
  - High performance computing division (jointly-organized with NEC), Cyberscience Center, Tohoku university