Optimizing Memory Layout of Hyperplane Ordering for Vector Supercomputer SX-Aurora TSUBASA

MCHPC'19: Workshop on Memory Centric High Performance Computing

Osamu Watanabe[1,2]
Yuta Hougi[1]
Kazuhiko Komatsu[1]
Masayuki Sato[1]
Akihiro Musa[1,2]
Hiroaki Kobayashi[1]

[1]Tohoku University, [2]NEC Corporation

18th November, 2019

CONTENTS

INTRODUCTION

Backgrounds

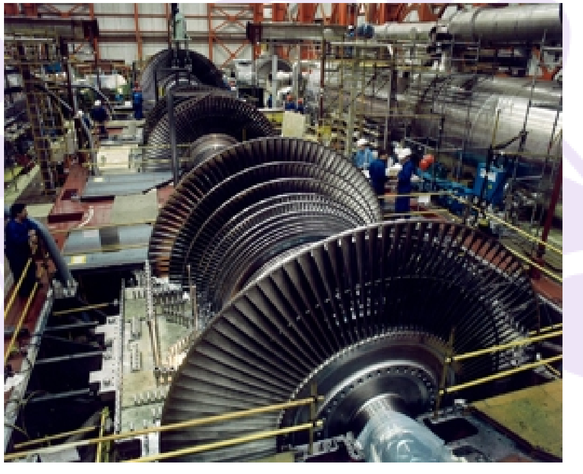
Various phenomena can be reproduced more realistically by large-scale simulations using supercomputer systems.

There are still many issues to be addressed, and the impact of problems with social infrastructures (ex. gas and steam turbines) on our society is immeasurable.

Demand for High Performance Turbine Simulation

Necessary to conduct a numerical simulation of a turbine using a supercomputer to simulate various phenomena occurring in the turbine in order to predict failures in advance.

- Complex wet-steam flow field analysis is required, but
- the cost for experiments is expensive.



http://www.mhi-global.com/

Numerical Turbine code using NEC SX-Aurora TSUBASA is the best solution.

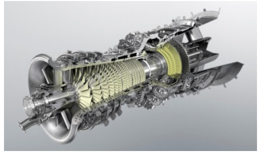
NUMERICAL TURBINE CODE ON SX-Aurora TSUBASA

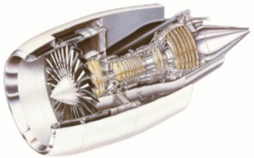
Numerical Turbine

High Performance Turbine Design Code on SX Systems

Numerical Turbine developed by Prof. Yamamoto of Tohoku University

- is a simulation code realizing High-performance and High-reliable Future Turbines and
- has been accelerated on the SX series of Cyberscience Center at Tohoku University.


Gas turbine for plants

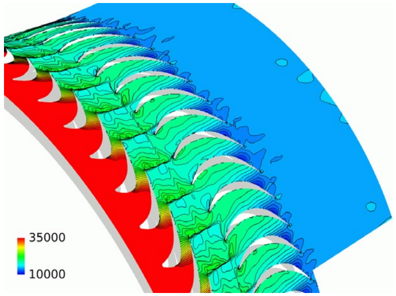

Gas turbine for airplanes


Steam turbine

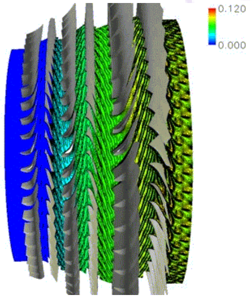
Only Numerical Turbine has achieved the following simulations in the world.

- Unsteady flows with wetness and shocks in actual gas turbines and steam turbines
- Full annulus simulation for resolving unsteady wet-steam and moist-air flows in actual turbines and compressors

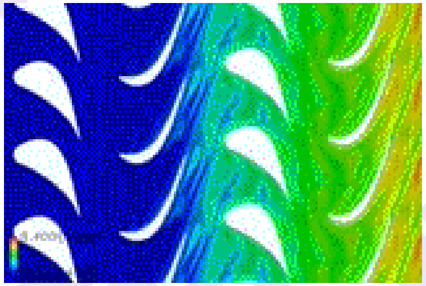
Resolving such complex flows is crucial for developing high-performance and high-reliable turbines



Unsteady shocks generated in turbine stages

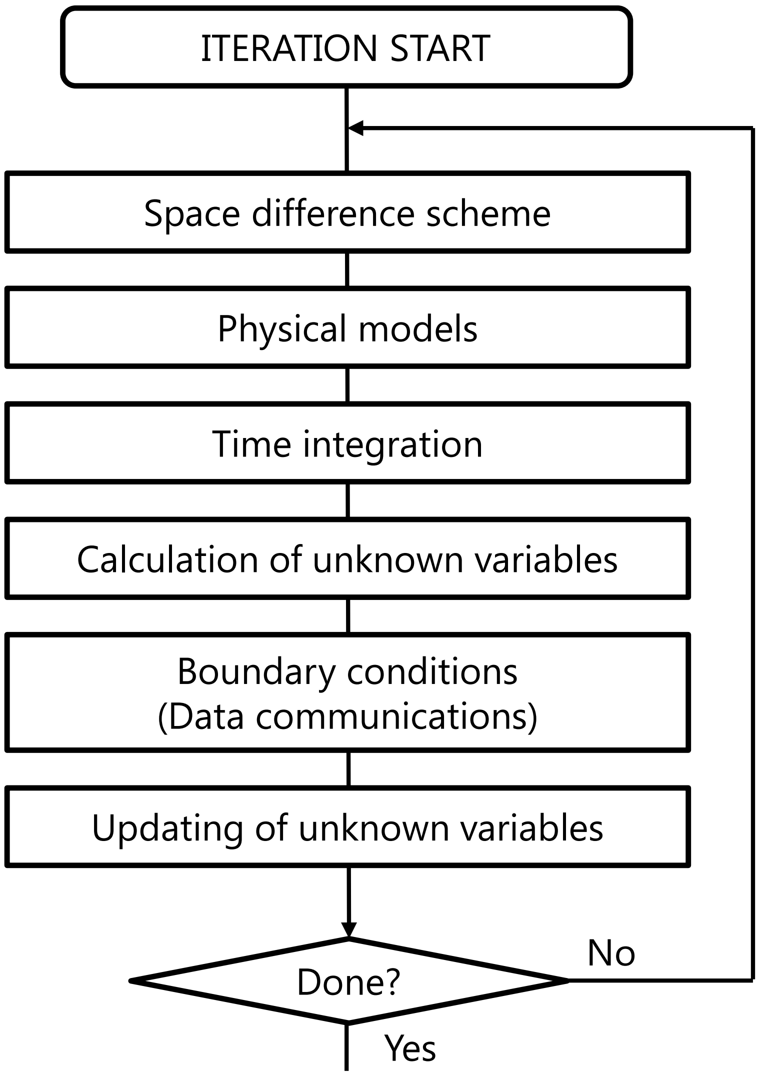


Unsteady wetness in full annuals turbine stages


Unsteady wet-steam flow in turbine stages

Numerical Schemes

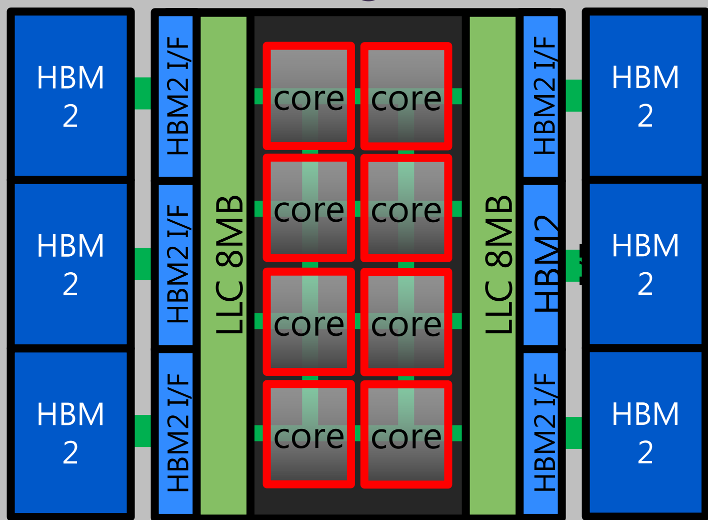
Flowchart of the Iteration Loop



- Space Difference of Convective and Pressure Terms
  - Roe's Approximate Riemann Solver
    (P.L.Roe, J.Comp.Phys., 1981)
  - Compact MUSCL
    (S.Yamamoto and H.Daiguji, Computer & Fluids, 1993)
- Viscous and Diffusion Terms
  - Second-order Central Difference
- Time Integration
  - LU-SGS Scheme
    (S.Yoon and A.Jameson, AIAA Jurnal, 1988)
- Turbulence Model
  - SST Turbulence Model
    (F.R.Menter, AIAA Journal, 1994)

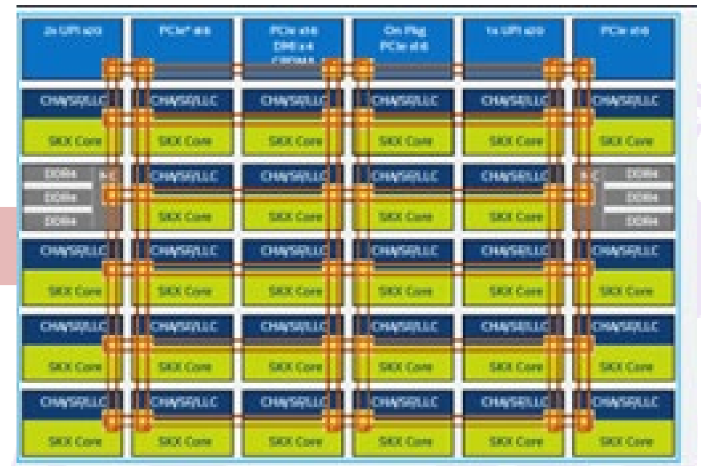
Hardware Specification of NEC SX-Aurora TSUBASA

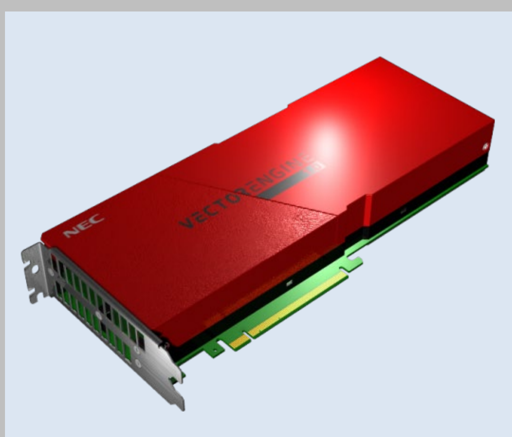
Vector Engine (VE)



PCIe

x86 Processor (Xeon)



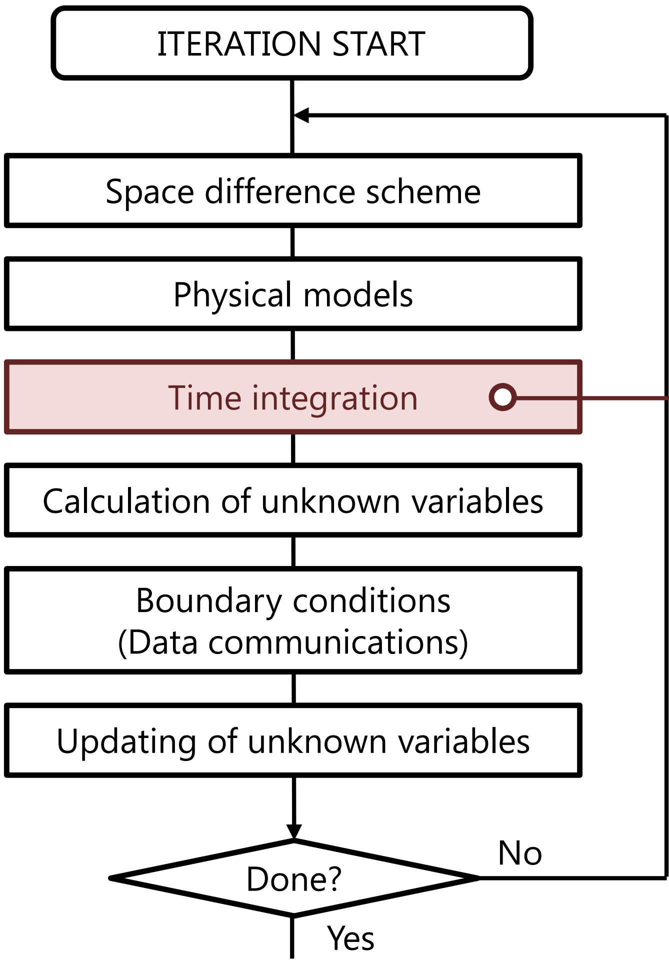


Vector Engine	Type 10B
Vector Cores	8
Frequency	1.4 GHz
Performance/core	537.6 GF (SP), 268.8 GF (DP)
Performance/processor	4.30 TF (SP), 2.15 TF (DP)
Cache Capacity	16 MB (shared)
Mem. Bandwidth	1.2 TB/s
Mem. Capacity	48 GB (HBM2 8GB x 6)

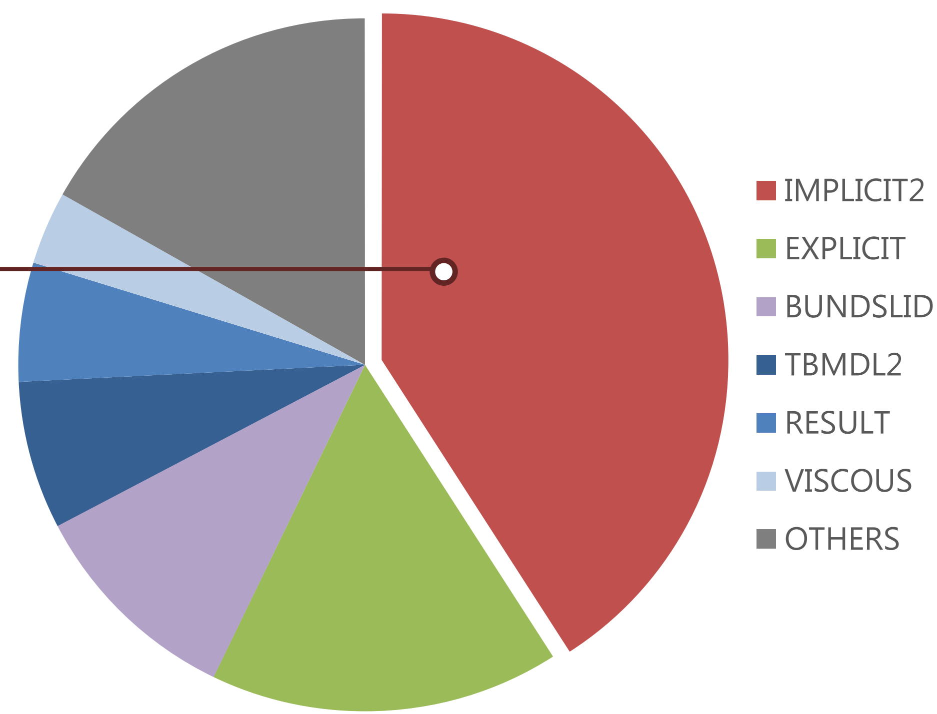
Long vector loop length is exploit the high performance of SX-Aurora TSUBASA.

Cost Distribution of Main Iteration Routines

Flowchart of the Iteration Loop



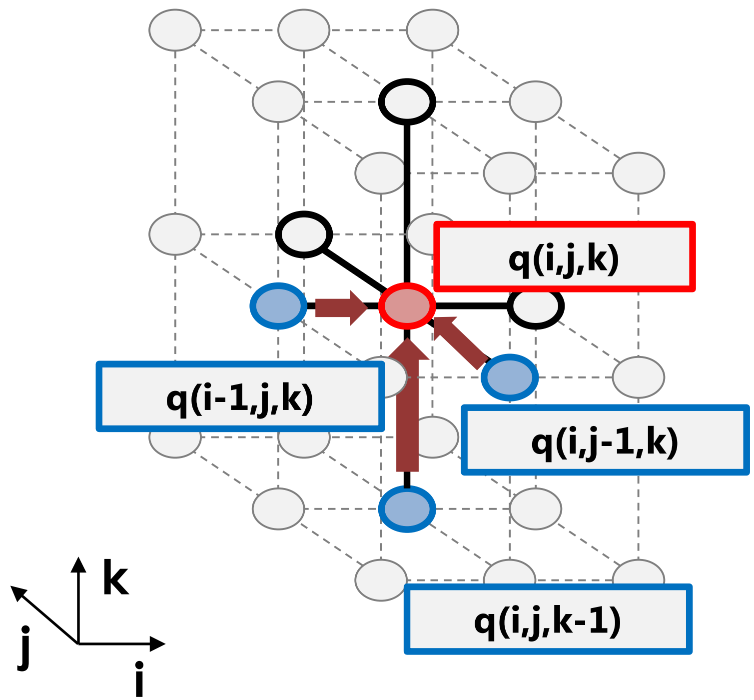
Cost Distribution



The time integration routine (LU-SGS)is the most dominant.

Data Dependency on the LU-SGS Scheme

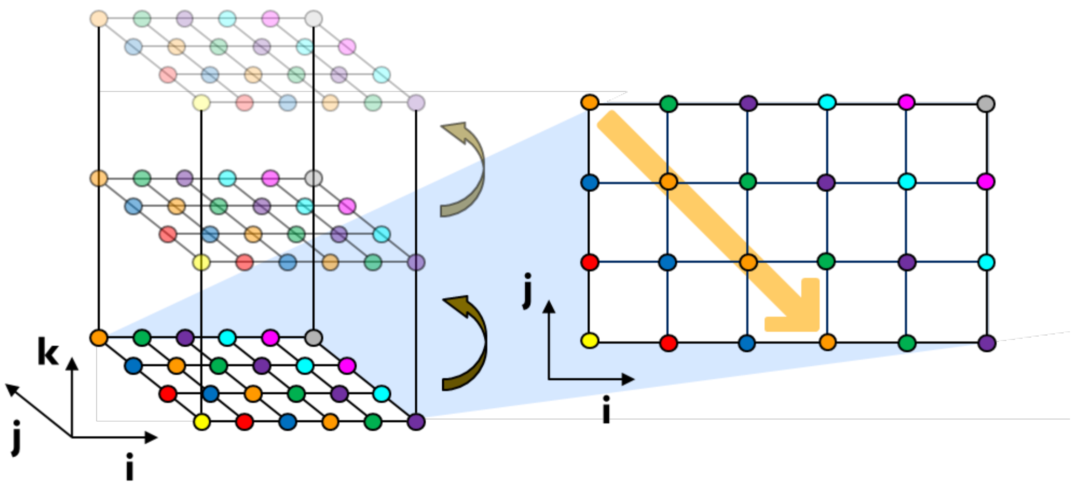
- Numerical Turbine adopts the LU-SGS scheme for time-integration
- Data dependency of q(i,j,k) on q(i-1,j,k), q(i,j-1,k) and q(i,j,k-1)



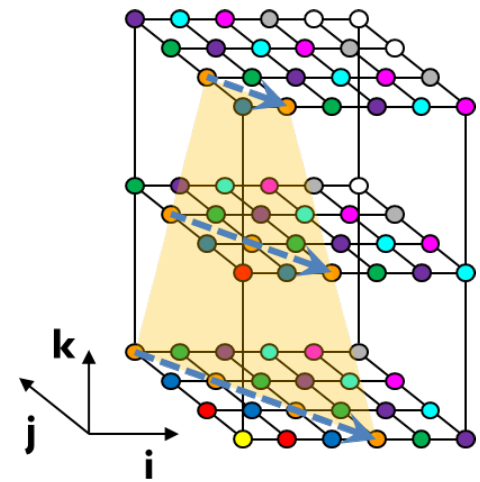
Multi-dimensional hyperplane method resolves this data dependency.

2D and 3D Hyperplane Ordering

2D Hyperplane



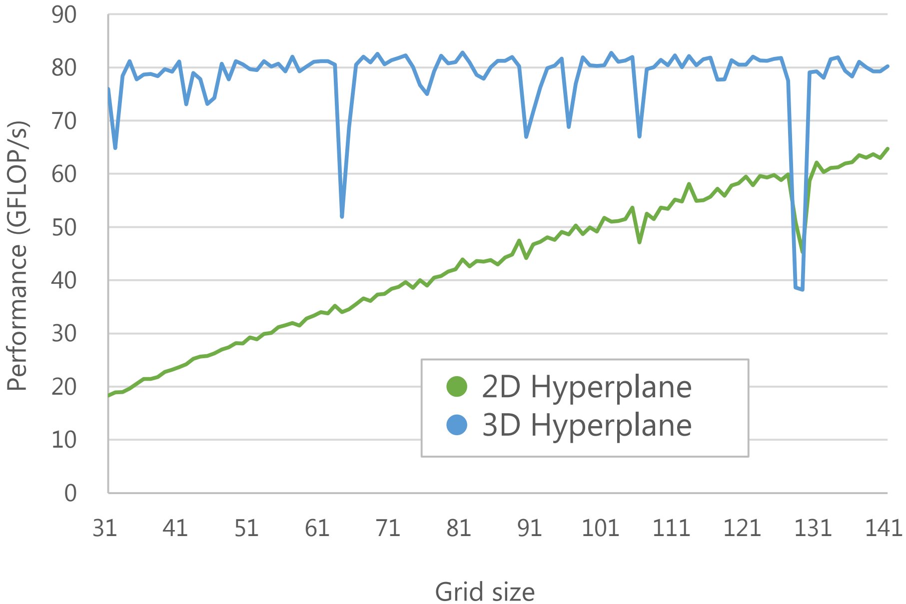
3D Hyperplane

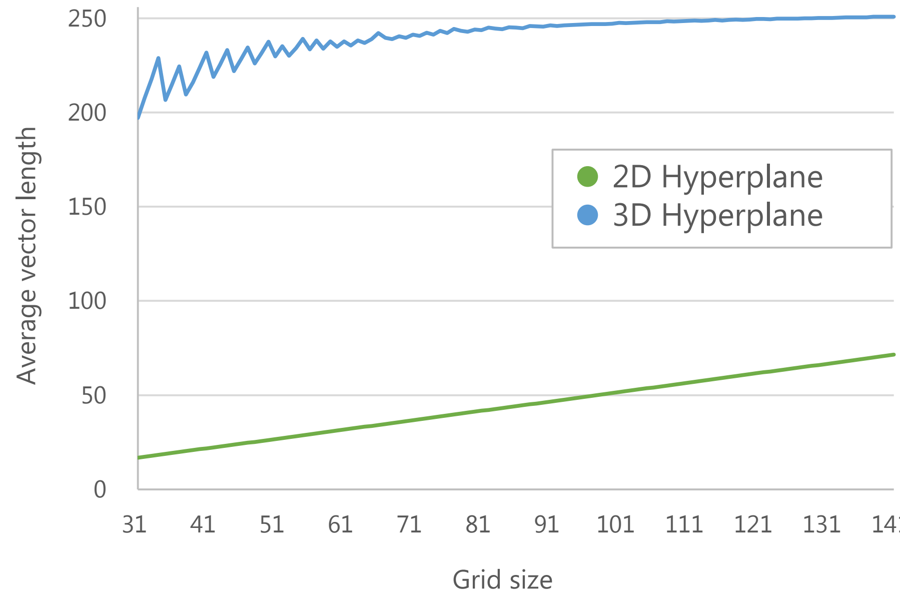


The sum of i, j, and k (in case of 2D-HP, the sum of I and j) of each grid point on a hyperplane is the same value.

	Vector Length	Memory Access
2D Hyperplane	Short	Direct Access
3D Hyperplane	Long	Indirect Access

2D Hyperplane vs. 3D Hyperplane





3D hyperplane provides long vector lengths. However...
The method is accompanied by a high memory load with indirect memory accesses.

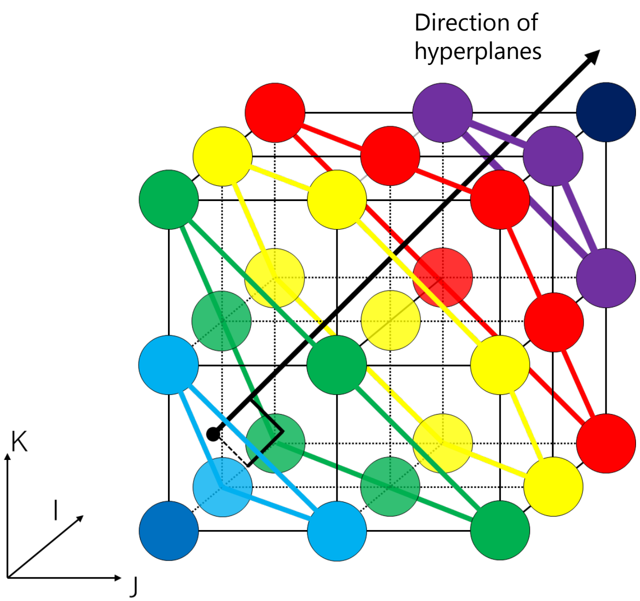
- Latency of the indirect accesses is high.
- Some blips downward in performance occurs due to bank-conflict.

Need to reduce indirect memory accesses in order to extract more performance from SX-Aurora TSUBASA.

OPTIMIZING
MEMORY-ACCESS PATTERN OF
3D HYPERPLANE ORDERING METHOD

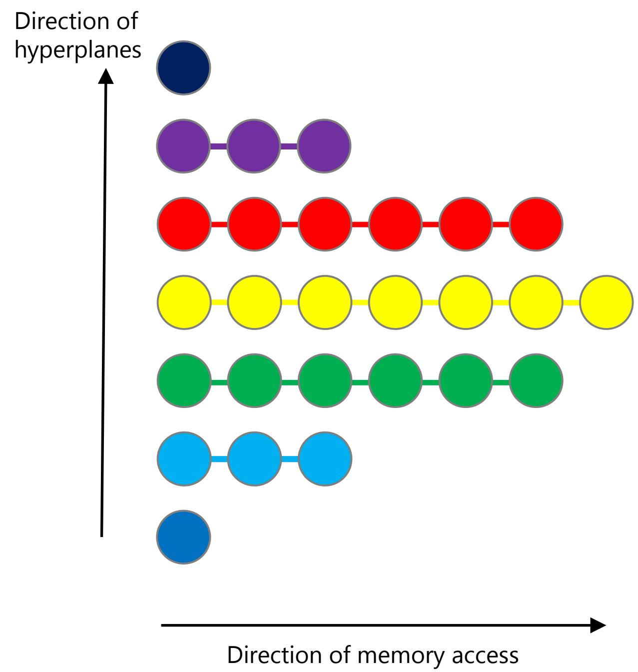
Changing 3D Data Layout to 1D Data Layout

Original 3D Data Layout



Indirect memory accesses

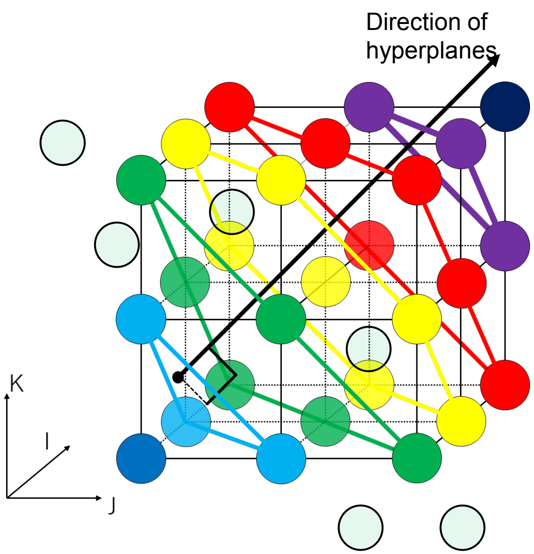
Proposed 1D Data Layout



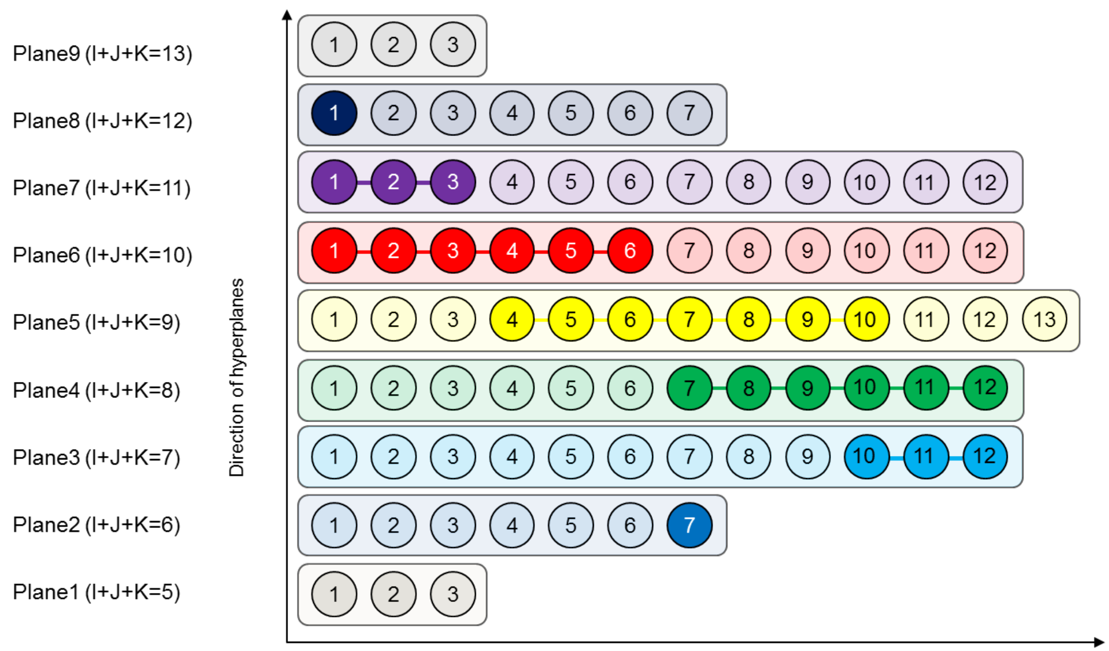
Sequential memory accesses

1D Data Layout and Each Plane Number

Original 3D Data Layout



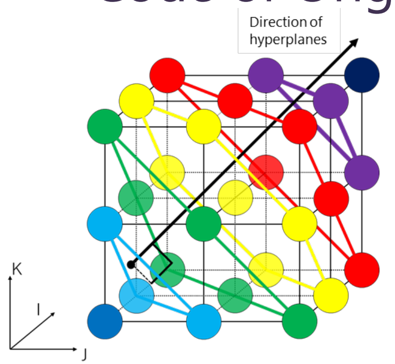
Proposed 1D Data Layout



When storing grid points in each 1D array, it is necessary to store the grid points on the boundary area on each hyperplane.

Source Code Image of Changing the Data Layout

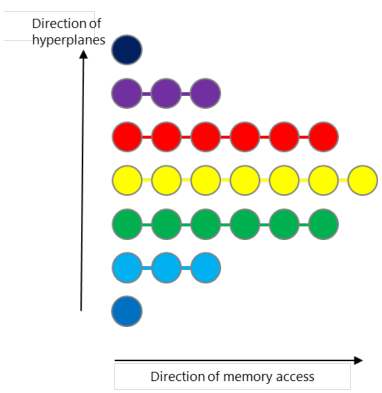
Code of Original 3D Data Layout




```
DO LHP=1,LIST_3DC(M,L)
  I = LIST_3DI(LHP,M,L)
  J = LIST_3DJ(LHP,M,L)
  K = LIST_3DK(LHP,M,L)
  …
  AIM = (A(I-1,J,K)*A(I,J,K))*0.5D0
  …
ENDDO
```

Indirect Access

Code of Proposed 1D Data Layout




```
DO LHP=1,LIST_3DC(M,L)
  I = LIST_3DI(LHP,M,L)
  J = LIST_3DJ(LHP,M,L)
  K = LIST_3DK(LHP,M,L)
  IM = LIST_3DA(I-1,J,K,L)
  …
  AIM = (A(M-1)%ARRAY(IM)*A(M)%ARRAY(LHP))*0.5D0
  …
ENDDO
```

Indirect Access

Sequential Access

Number of Gather/Scatter Operations in Calculation Part

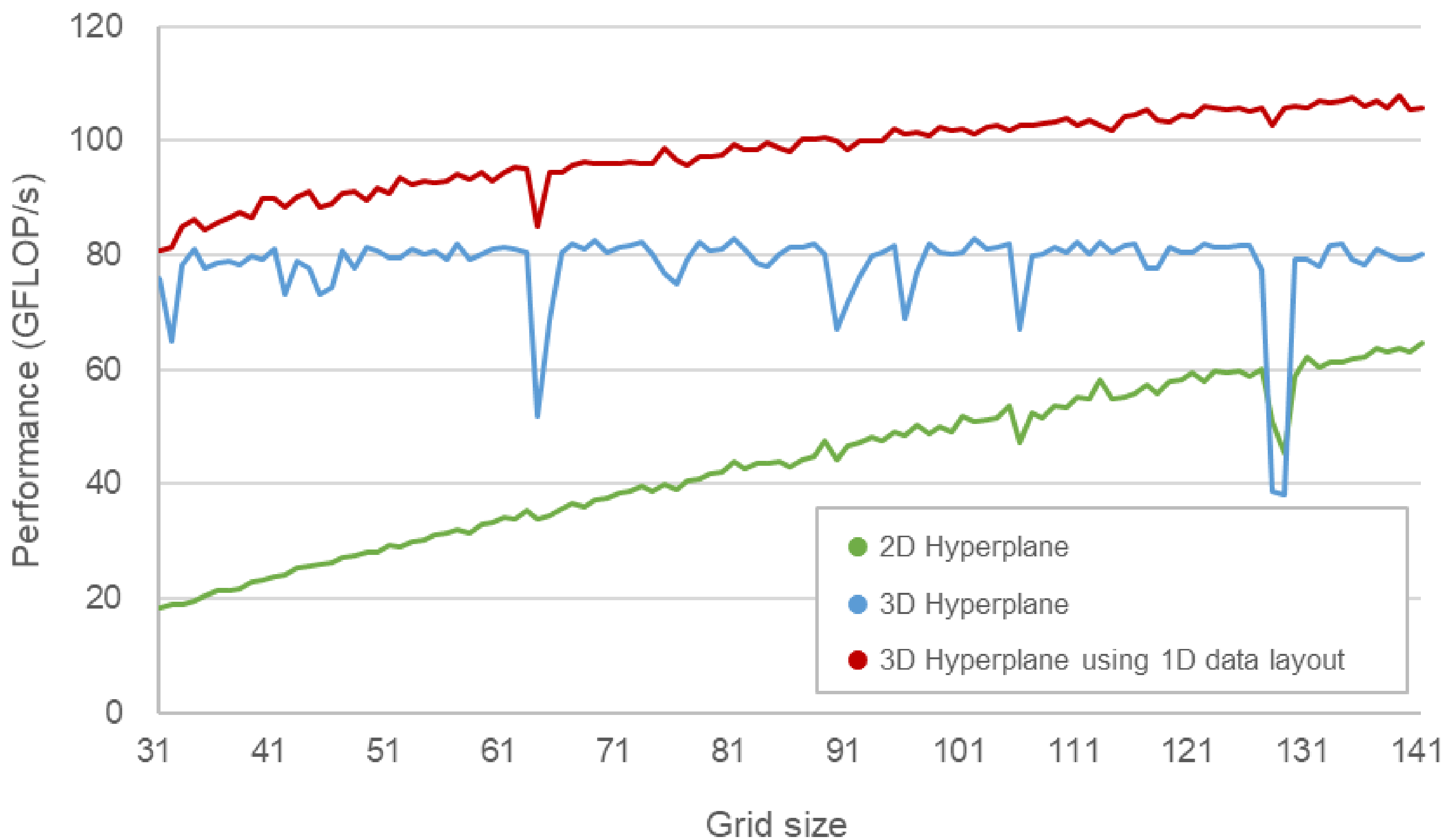
	Original 3D Hyperplane	Proposed 3D Hyperplane
Vector Gather op.	228	162
Vector Scatter op.	20	0

Reducing indirect memory accesses

PERFORMANCE EVALUATION

Performance of The Proposed Method

Perforance evaluation using Time-integration kernel codes.

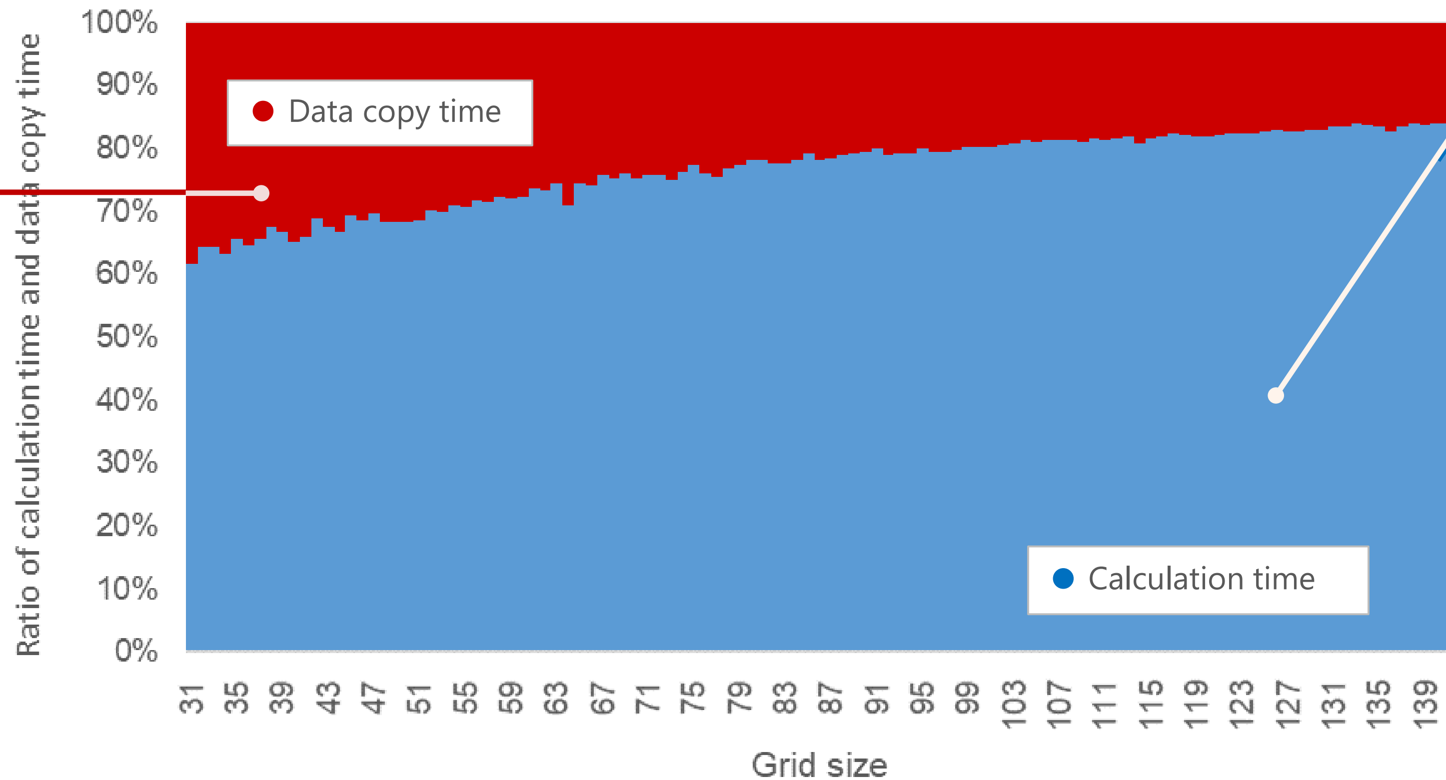


Average: 1.27x
Max : 2.77x
faster than original 3D Hyperplane

Some blips downward in performance are reduced.

Cost Distribution of Data-copy and Calculation

Calculation part performance: 123 ~146 Gflop/s

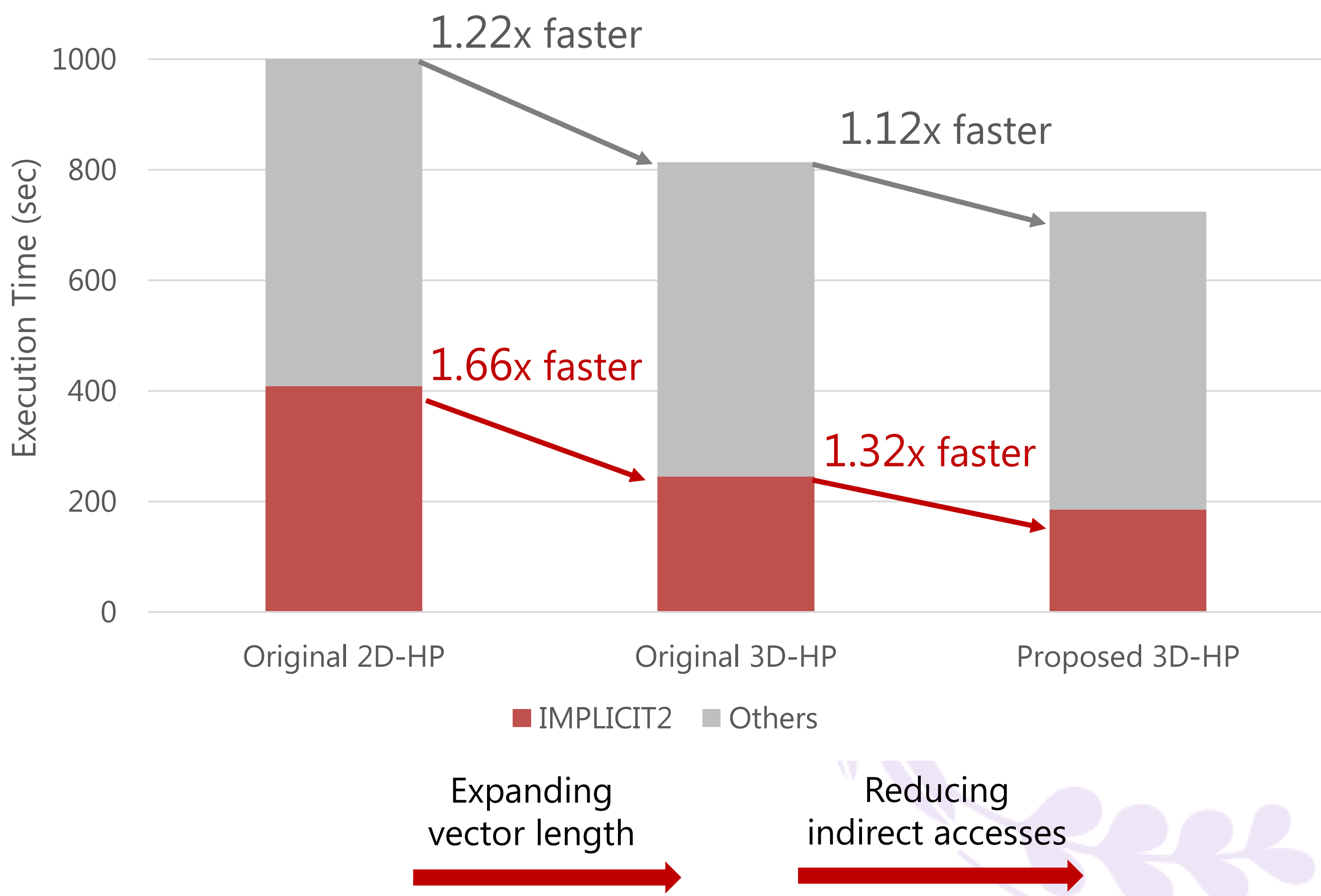


- The proposed method involves data-copies (3D layout ⇔ 1D layout).

Reducing the data-copy time can further improve the performance.

Performance Improvement of Full App.

- Numerical Turbine execution with 636 mpi
- 91x91x181 grid points for each stator and rotor passage



CONCLUSIONS

Conclusions and Future Work

- The 3D Hyperplane ordering method is suitable for SX-Aurora TSUBASA because the method can provide long vector lengths.
- Our proposed 3D Hyperplane ordering method further improves the performance by reducing indirect memory accesses.
- Future work: examine the effect of the method on other applications and major SIMD architectures.

Thank you for your kind attention!