SC14 Poster @ New Orleans, LA, USA 16-21, Nov. 2014

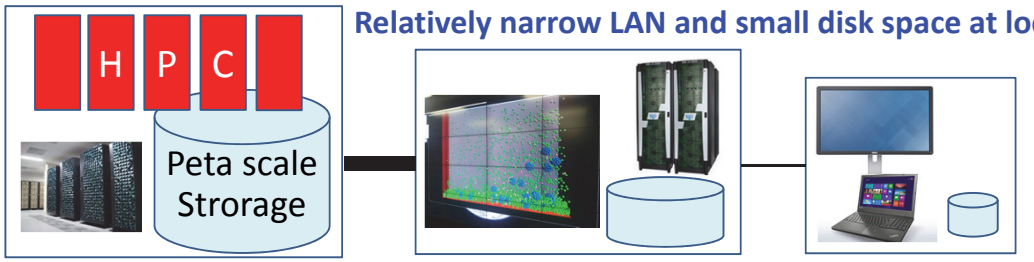
Partially supported by
Joint Usage / Research Center for Interdisciplinary Large-scale Information Infrastructures

Efficient Data Compression by Efficient Use of HDF5 Format with Bit Segmentation Method for Visualization and Analysis

Katsumi Hagita(National Defense Academy of JAPAN), Manabu Omiya(Hokkaido Univ.), Takashi Honda(Zeon), Masao Ogino(Nagoya Univ.)

Data Transfer Problem

Minimize file size for VR visualization.

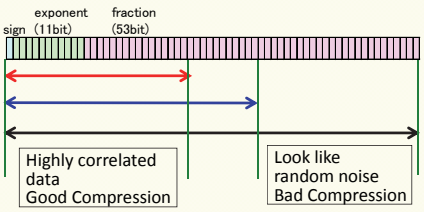


Relatively narrow LAN and small disk space at local environment

Concept of JHPCN-DF

Jointed Hierarchical Precision Compression Number - Data Format (JHPCN-DF)

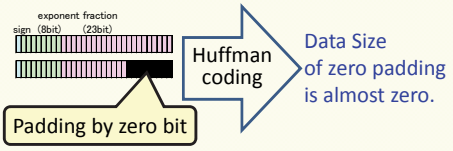
IEEE754



Visualization
Analysis
Lossless

Local workstation (PC)
Neighboring supercomputers
Top Supercomputers

Reduction of Data Size



Example of implementation

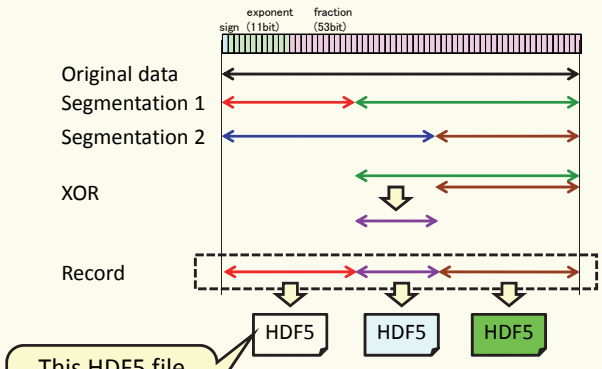
```
union fi32{
float f;
int i32;
};
 union fi32 fival;
 double logallo=log(allowerr)/log(2.0);
```

XOR

```
fival.f=fval0;
fival1.f=fval1;
 ixval=(fival1.i32  ^ fival.i32);
```

Segmentation

```
fval=frexp(fval0,&ival);
ival2=(int)(-logallo+ival-1);
sval=(int)(23-ival2+1);
if(sval>23) sval=23;
do {
sval--;
fival.f=fval0;
fival.i32=(fival.i32 >> sval);
fival.i32=(fival.i32 << sval);
fval1=fival.f;
} while ((fval1-fval0)*(fval1-fval0)>allowerr*allowerr);
```



Related method of standard HDF5

- N-bit Filter
  packing n-bit data on output by stripping off all unused bits
  and
  unpacking on input (restoring extra bits)
- Scale-Offset Filter
  Truncate value to a minimum number of bits before storing by scale and/or offset
  Relative error < 5*(D-scale_factor+1))

Advantage of JHPCN-DF

- Decodable without special API.
  - JHPCN-DF is effective for almost software and user application supporting HDF5.
- Splitting a few HDF5 files by simple converter application.
- Recoverable with lossless.

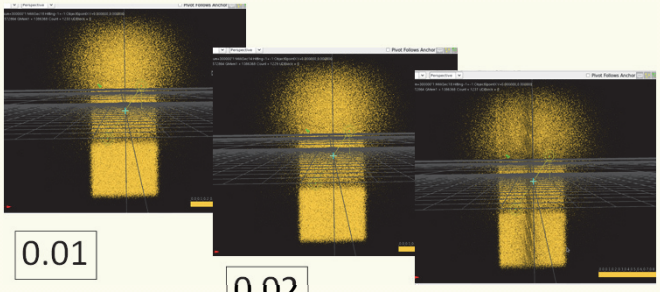
Acknowledgement

This work is partially supported by "Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures" and "High performance Computing Infrastructure" in Japan.
JHPCN 14-NA28, HPCI hp140191 (JHPCN 10-MD01, 11-MD02, 12-MD03, jh130028-NA19, HPCI hp130056, hp130062, hp130122)
This work is partially supported by HPC project of Nagoya University.

Test for plasma PIC simulation

150,000 ions and 150,000 electrons

Comparison for allowerr =0.01, 0.02, 0.05



0.01 0.02 0.05

Size of 20 HDF5 files of 100 timesteps

	exponent fraction sign (8bit) (23bit)	exponent fraction sign (8bit) (23bit)
0.0	6,336 MB	----
0.001	3,369 MB	4,503 MB
0.01	1,940 MB	5,298 MB
0.02	1,413 MB	5,586 MB
0.05	883.1 MB	5,712 MB
0.1	568.5 MB	5,857 MB

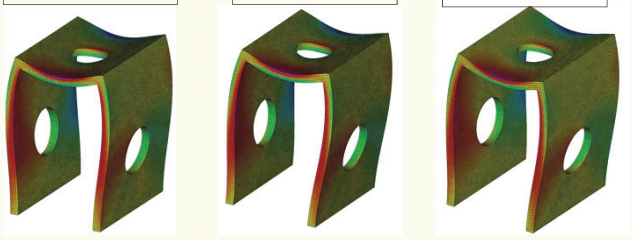
Test for FEM simulation

Finite element analysis of 0.24 million elements and 0.38 million nodes

8.3MB for double

3.3MB for 1.0e-6

1.3MB for 1.0e-3

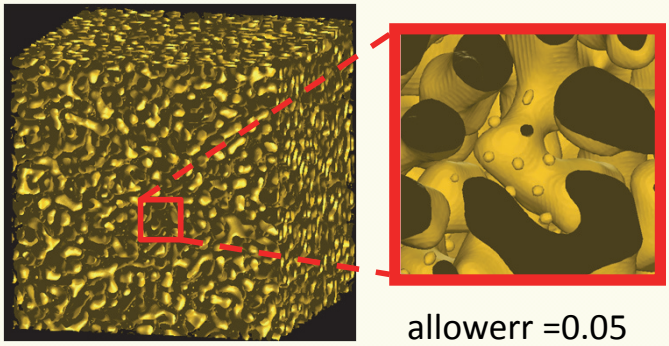


Good for analysis

Good for visualization

Test for block copolymer

Reduce data size from 1182MB to 42MB to present isosurface of phase separated block copolymers. (mesh size: 512^3)



allowerr =0.05

Conclusion and Future

- Concept of JHPCN-DF works and shows excellent results for various type of simulation! JHPCN-DF is a powerful tool for all large scale simulations.
- Distribute know-how of JHPCN-DF for all supercomputing.
  - Increase number of examples for various simulation fields (in progress)
    - Nonhydrostatic ICosahedral Atmospheric Model (NICAM)
    - Fluid dynamics simulation
    - 0.1 billion powder simulation
- We will upgrade as a filter of HDF5.
- We consider post process of HDF5 using MPI-IO to be compressed by JHPCN-DF as system software coupled with job scheduler.
- Treatment of the last bit is related to rounding mode control of IEEE 754.