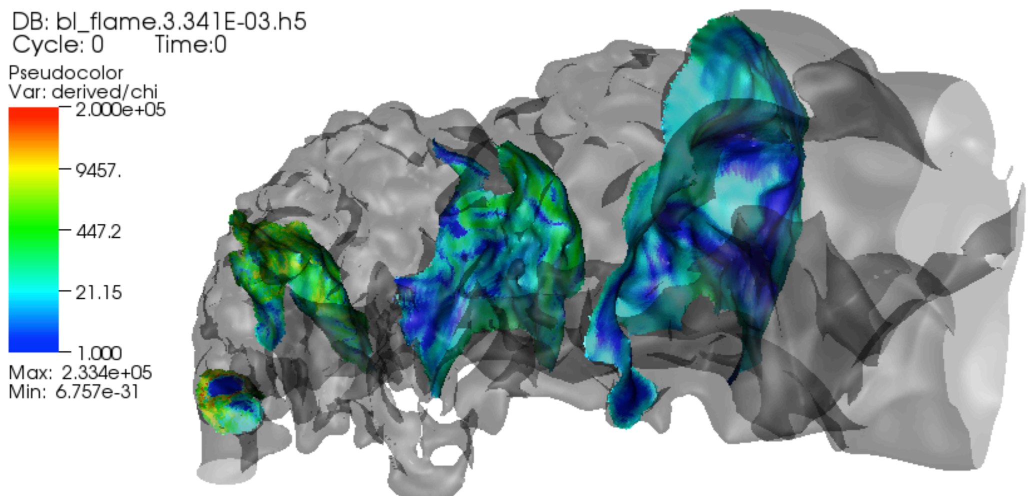
A System-Aware Optimized Data Organization for Efficient Scientific Analytics

Yuan Tian[1] Scott Klasky[2] Weikuan Yu[1] Hasan Abbasi[2] Bin Wang[1] Norbert Podhorszki[2] Ray Grout[4] Matthew Wolf[3]

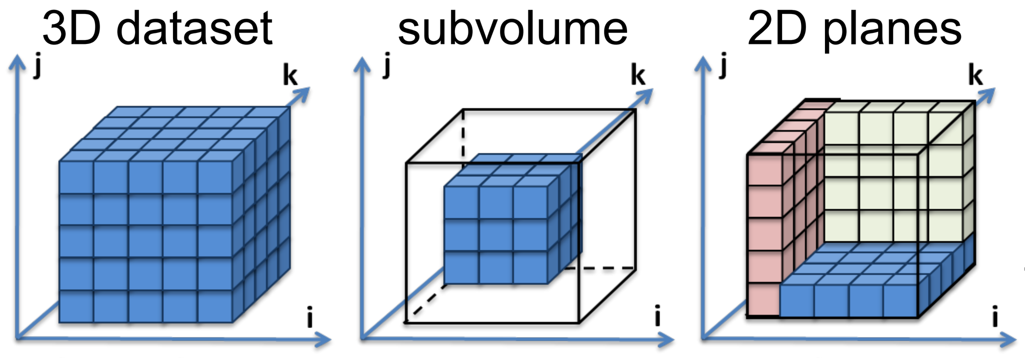
Auburn University[1] Oak Ridge National Laboratory[2]

Georgia Institute of Technology[3] National Renewable Energy Laboratory[4]

Problem Statement



An S3D Combustion simulation result. Colored regions are the points of interest



Data Representation and Access Patterns

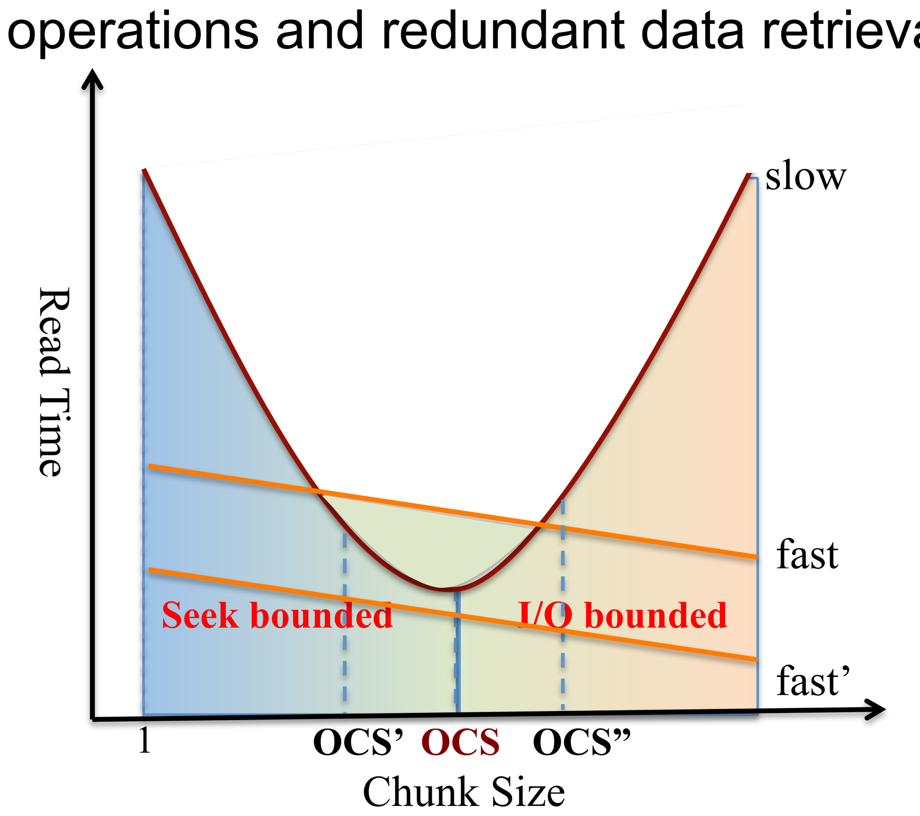
- Scientific applications generates massive amount of multi-dimensional arrays
- Read performance is crucial for application execution and data post-processing
- Existing data layouts produces imbalanced read performance for common access patterns of post-processing due to:
  - Inefficiency to alleviate the dimension dependency for common access patterns
  - Poor data concurrency on large-scale storage systems

Jaguar Supercomputer at ORNL, World #3 fast

<u>OUR GOAL</u> - A new data layout provides GOOD and BALANCED read performance for scientific data post-processing.

Optimized Chunking

- Goal: mathematically find the Optimized Chunk Size (**OCS**) that gives the balance between the overhead of seek/read operations and redundant data retrieval.



Read Time vs. The Number of Chunks

- Considering the read performance on both fast and slow dimensions, our algorithm tries to solve the Optimized Region between OCS and OCS'.

Number of Optimized Chunks:

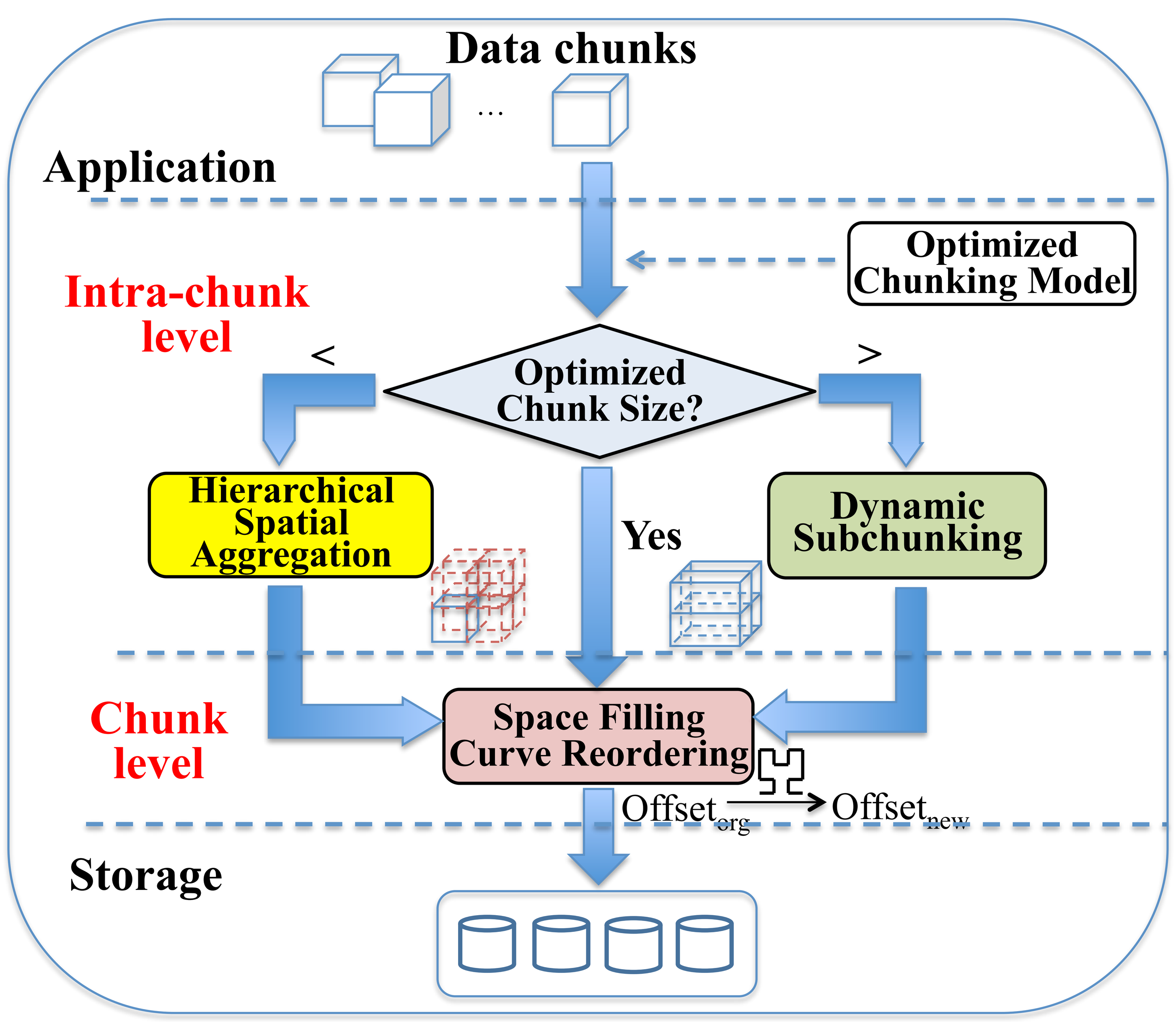
$$N_{ocs} = \begin{cases} \frac{CS}{BW_{io} \times (CC+T_s)}, & \text{when dividing large data chunks} \\ \frac{BW_{io} \times (CC+T_s)}{CS}, & \text{when aggregating small data chunks} \end{cases}$$

System-Aware Data Organization

- <u>**TWO level of data reorganization**</u>:

  Intra-chunk level: Constructing data chunks into **OCS**

  Chunk level: Reorganize data chunks using Space Filling Curve



Hiearchical Spatial Aggregation:

- <u>Method</u>: Aggregate the <u>**small**</u> data chunks into **OCS** with the spatial locality is reserved
- <u>Benefit</u>: Reduce the number of seek/read operations for reading

Dynamic Chunking:

- <u>Method</u>: An n-1 domain decomposition for <u>**large**</u> data chunks
- <u>Benefit</u>: Reduce the amount of redundant data retrieval

SFC-based Reordering:

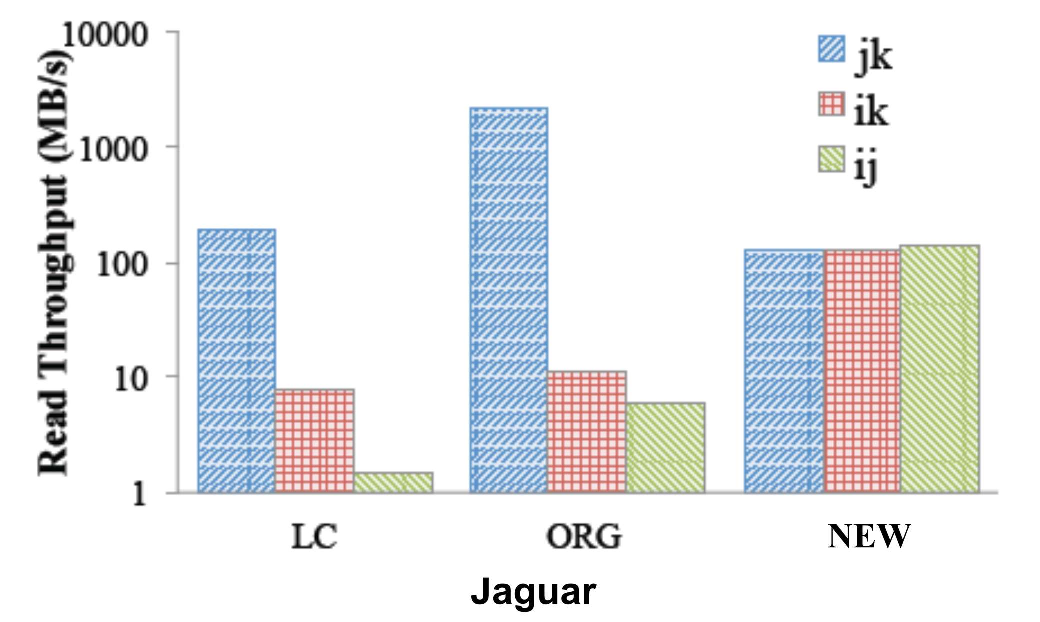
- <u>Method</u>: Data chunks with OCS are distributed along the order of Hilbert Space Filling Curve.
- <u>Benefit</u>: Guarantees the near-optimal concurrency for common access patterns

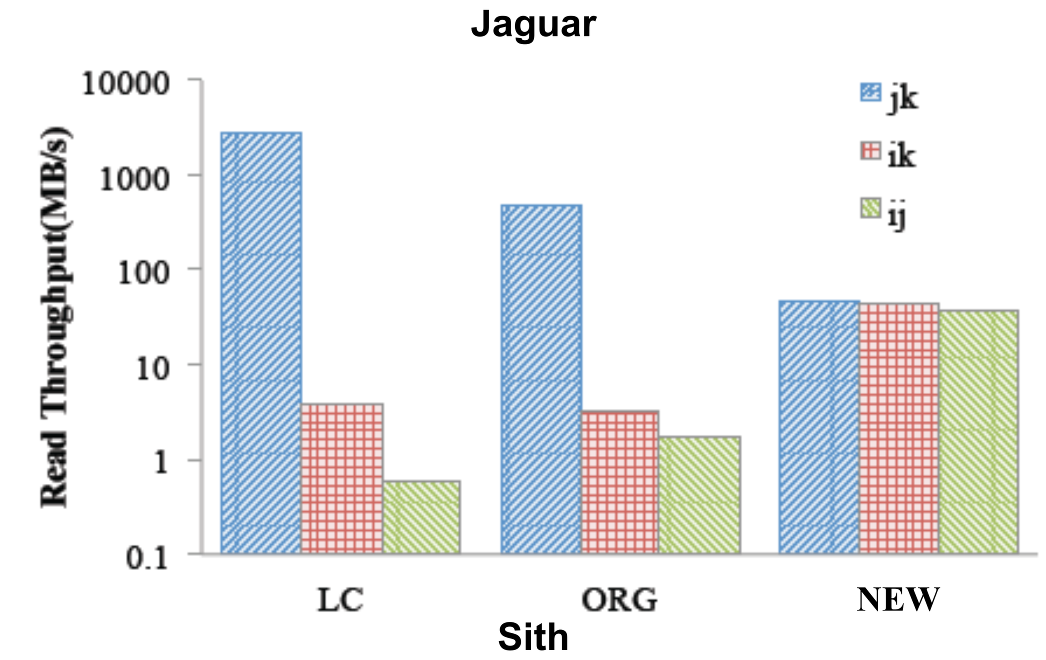
Performance Evaluation

- Experiment environment:
  - Jaguar supercomputer at ORNL
  - Combustion simulation code S3D

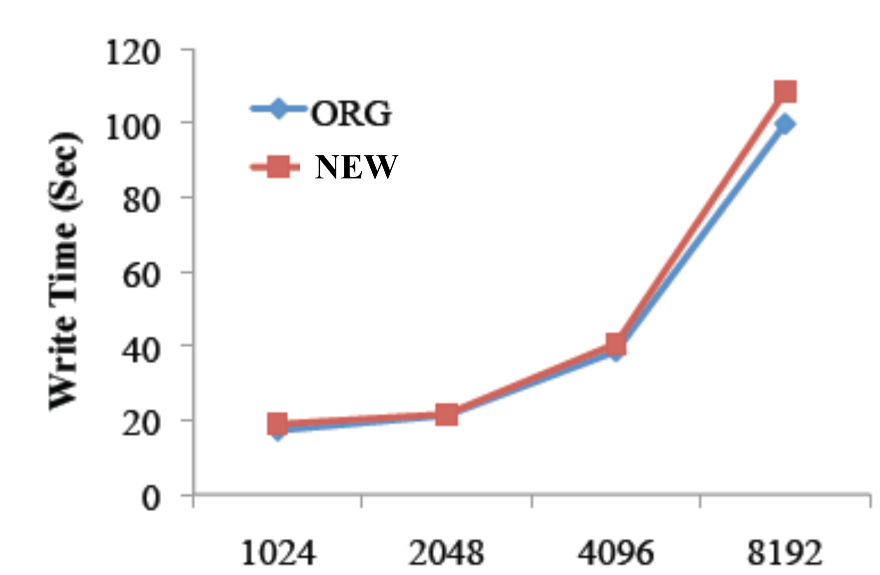
	Var1	Var2	Var3	Var4
Chunk	256^3/128MB	128^3/16MB	64^3/2MB	32^3/256KB
Variable	4096^3/512GB	2048^3/64GB	1024^3/8GB	512^3/1GB
Operations	DYS/SFC	DYS/SFC	SFC	HSA/SFC

- **Planar Readers -** ***balanced*** *and* ***improved*** *performance, maximum of* ***66*** *times speedup*
  - 4,096 writers, up to 512 readers
  - Peak performance comparison among Logically Contiguous (LC), Chunking (ORG), and our new data organization (NEW)





- **Write Overhead < 3%**



Conclusion & Future Work

- **Conclusion**
  - Our Optimized Chunking based data organization provides an improved and balanced read performance for common access patterns
  - Maximum of **66** times speedup
  - Negligible write overhead
- **Future work**
  - Fine tuning of the algorithm
  - Extension to other file system such as GPFS

Acknowledgement

This work is funded in part by a UT-Battelle grant to Auburn University, and in part by National Science Foundation award CNS- 1059376. This research is also supported by an UT-Battelle grant (UT-B-4000103043) to Auburn University. It used resources of the NCCS at ORNL, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

Reference:

[1] J. H. Chen et al. Terascale direct numerical simulations of turbulent combustion using S3D. *Comp. Sci. &Disc.*, 2(1):015001 (31pp), 2009.

[2] D. Hilbert. Ueber die stetige abbildung einer line auf ein fl¨achenst¨uck. *Math. Ann.*, 38:459–460, 1891.

[3] ADIOS. *http://www.nccs.gov/user-support/center-projects/adios/*.