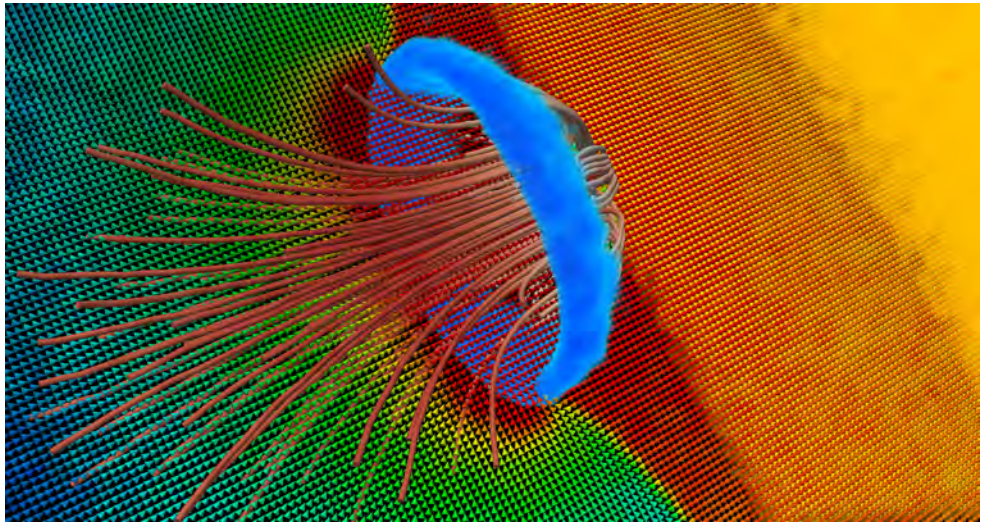
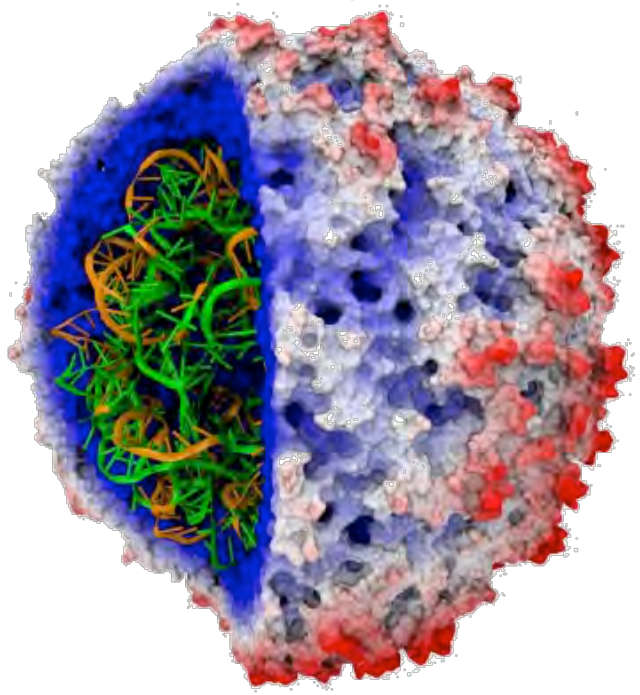
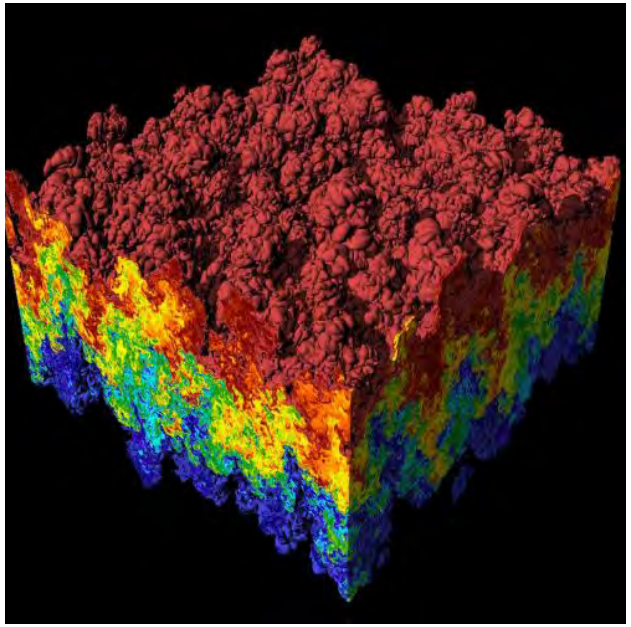
Argonne Training Program on Extreme-Scale Computing (ATPESC)

Data Analysis and Visualization







EXASCALE COMPUTING PROJECT

National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY | Office of Science

Argonne NATIONAL LABORATORY

Visualization & Data Analysis

Time	Title of presentation	Lecturer
8:30 am	**Visualization Introduction**	Mike Papka, Joe Insley, Silvio Rizzi, ANL
9:30 am	**Large Scale Visualization with ParaView (Presentation)**	Dan Lipsa, Kitware
10:30 am	*Break*	
11:00 am	**Large Scale Visualization with ParaView (Hands-on Exercises)**	Dan Lipsa, Kitware
12:00 pm	**Visualization and Analysis of Massive Data with VisIt (Presentation)**	Cyrus Harrison, LLNL
12:30 pm	*Lunch and Hands-on Exercises*	
1:30 pm	**Visualization and Analysis of Massive Data with VisIt (Hands-on Exercises)**	Cyrus Harrison, LLNL
3:00 pm	*Break*	
3:30 pm	**Scalable Molecular Visualization and Analysis Tools in VMD**	John Stone, UIUC
4:30 pm	**Exploring Visualization with Jupyter Notebooks**	Mike Papka, Joe Insley, Silvio Rizzi, ANL
5:30 pm	***Dinner Talk: Big Data Brain Maps at Argonne National Laboratory***	Bobby Kasthuri
6:30 pm	*Hands-on Exercises*	

Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

Argonne Training Program on Extreme-Scale Computing (ATPESC)

Visualization Introduction

Mike Papka
Joe Insley
Silvio Rizzi
Argonne Leadership Computing Facility
Argonne National Laboratory

Q Center, St. Charles, IL (USA)
August 10, 2017

EXASCALE COMPUTING PROJECT

National Nuclear Security Administration

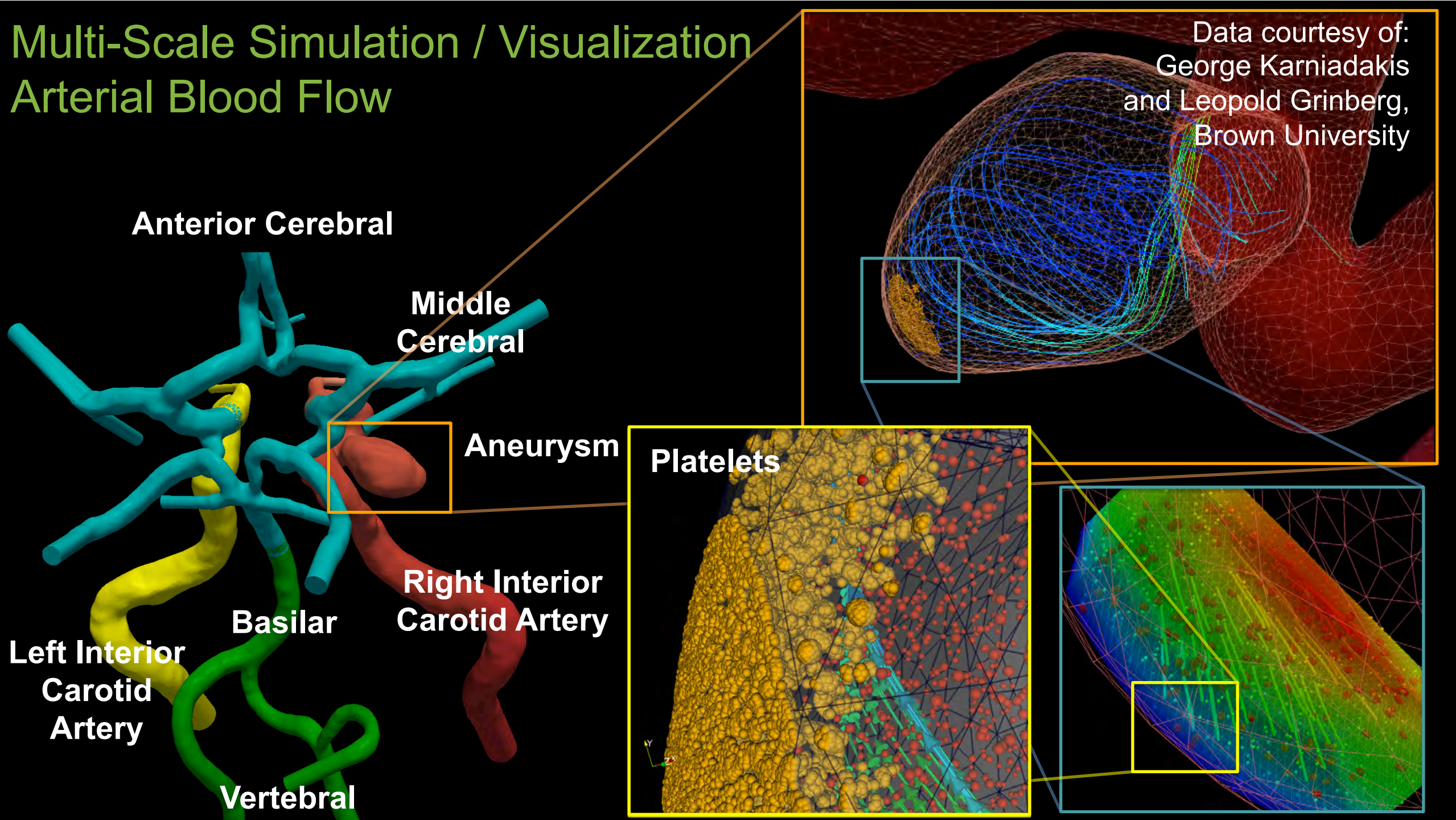
U.S. DEPARTMENT OF ENERGY | Office of Science

Argonne NATIONAL LABORATORY

Here's the plan…

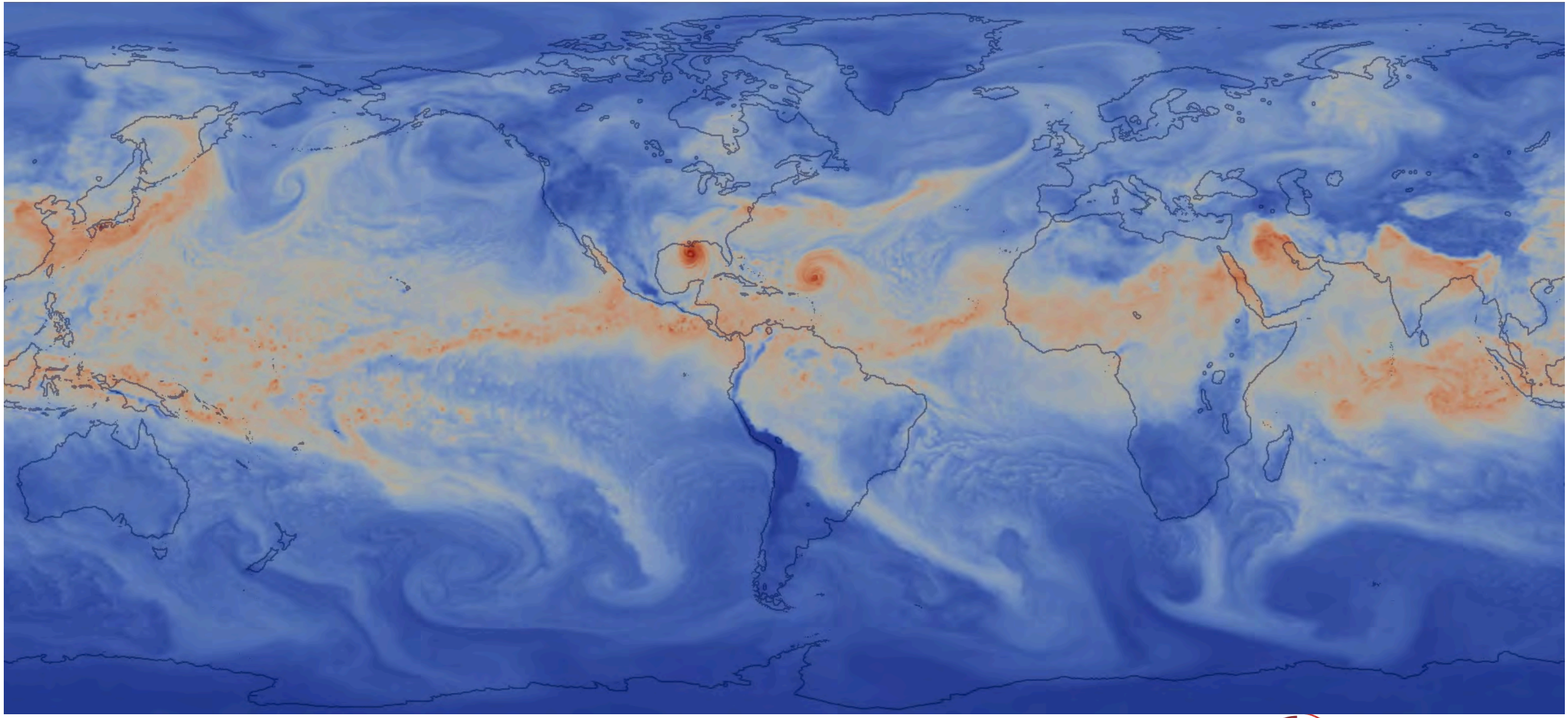
- Examples of visualizations
- Visualization resources
- Visualization tools and formats
- Data representations
- Annotation and movie creation
- Visualization for debugging
- In-Situ Visualization and Analysis

Multi-Scale Simulation / Visualization
Arterial Blood Flow



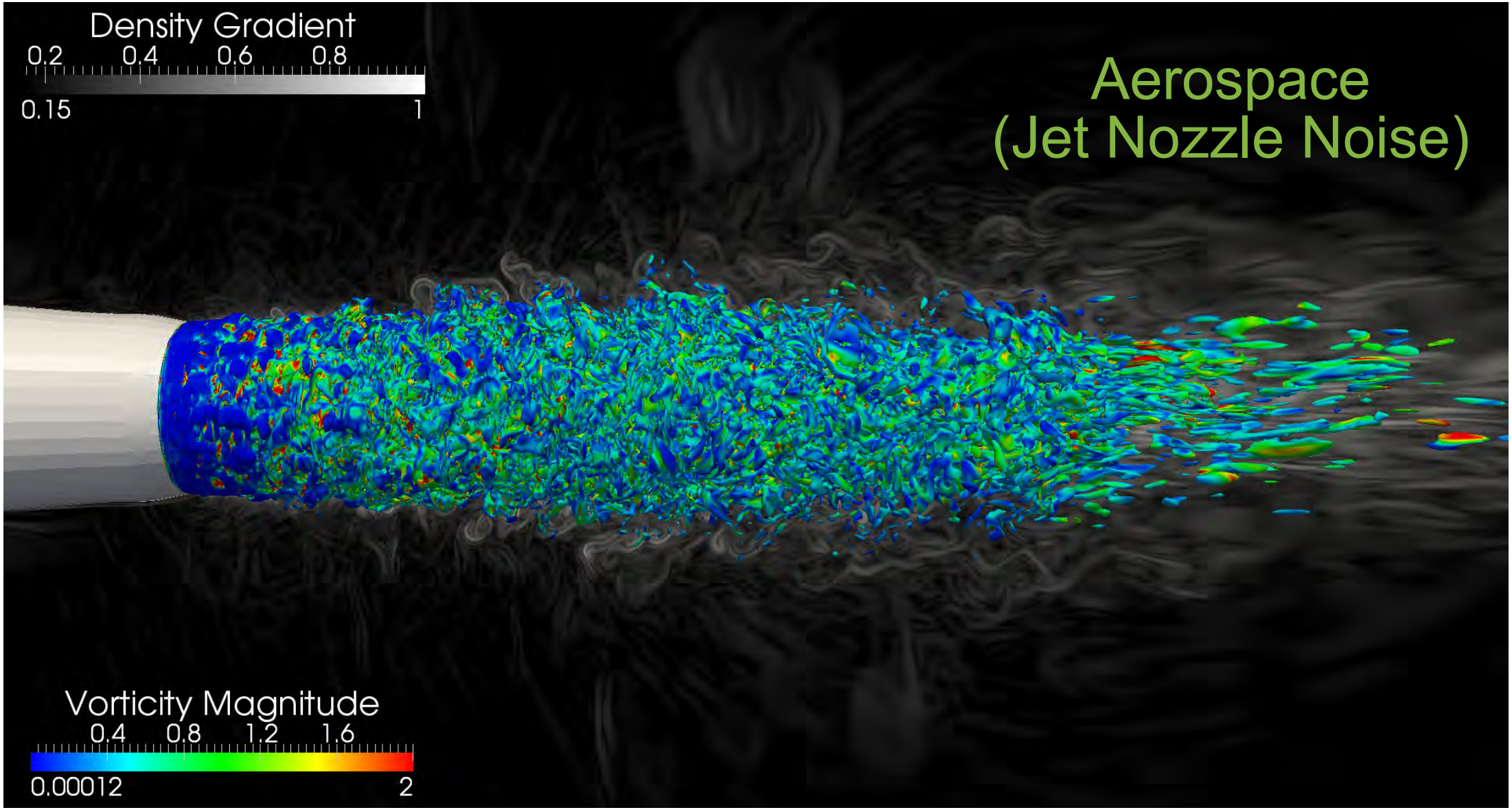
Climate

Data courtesy of: Mark Taylor, Sandia National Laboratory; Rob Jacob, Argonne National Laboratory; Warren Washington, National Center for Atmospheric Research



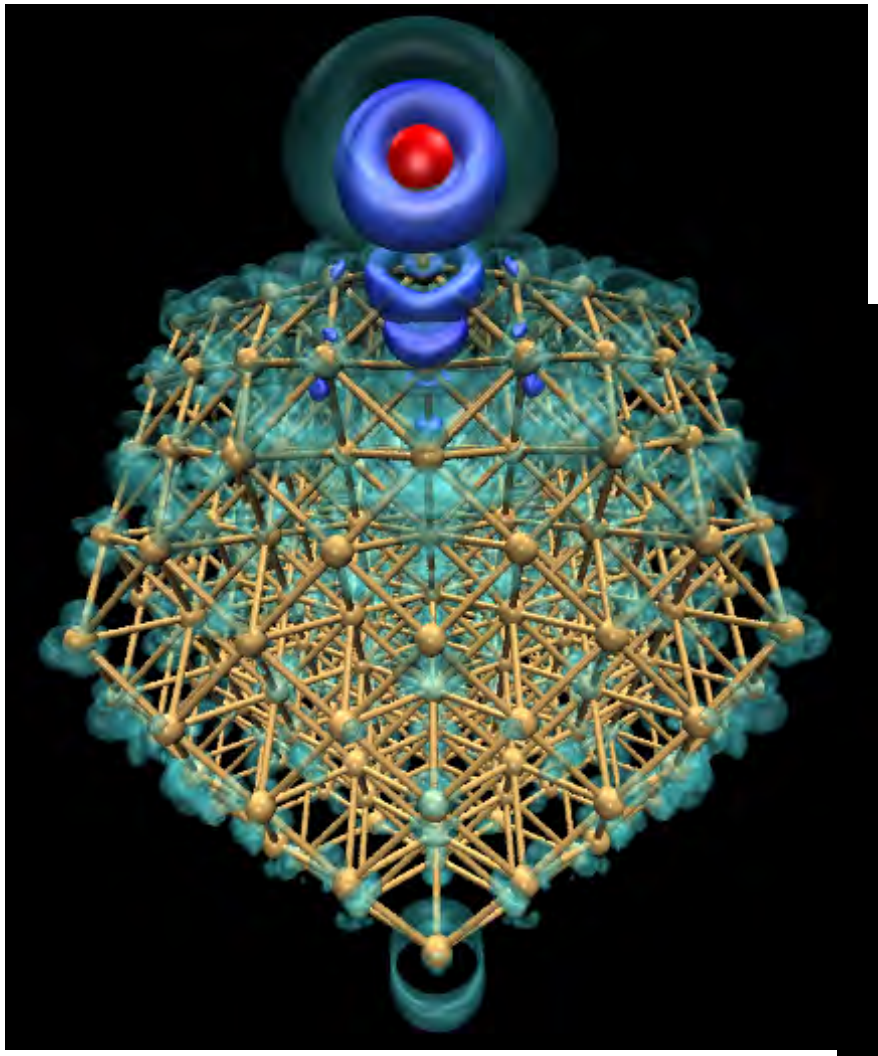
Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT



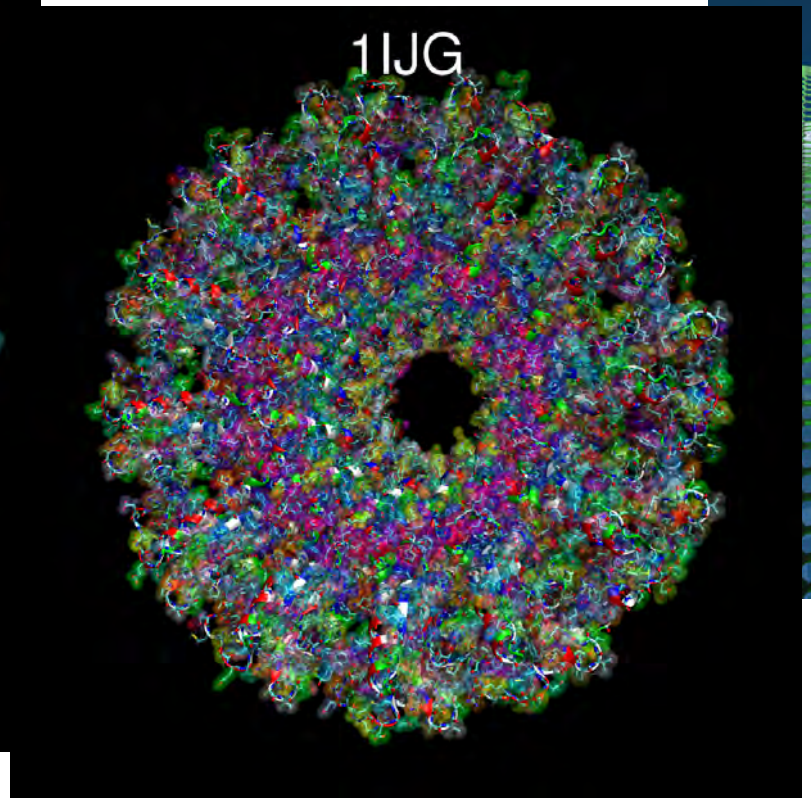
Data courtesy of: Anurag Gupta and Umesh Paliath, General Electric Global Research

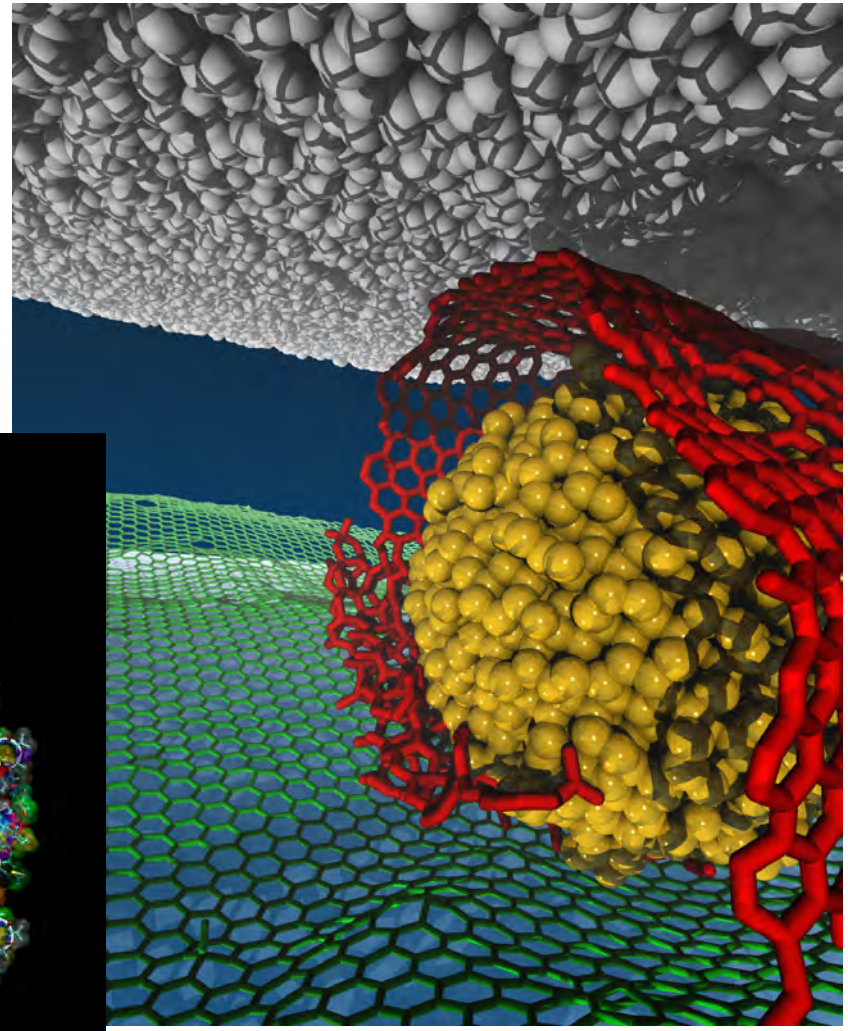
Materials Science / Molecular



Data courtesy of: Jeff Greeley, Nichols Romero, Argonne National Laboratory

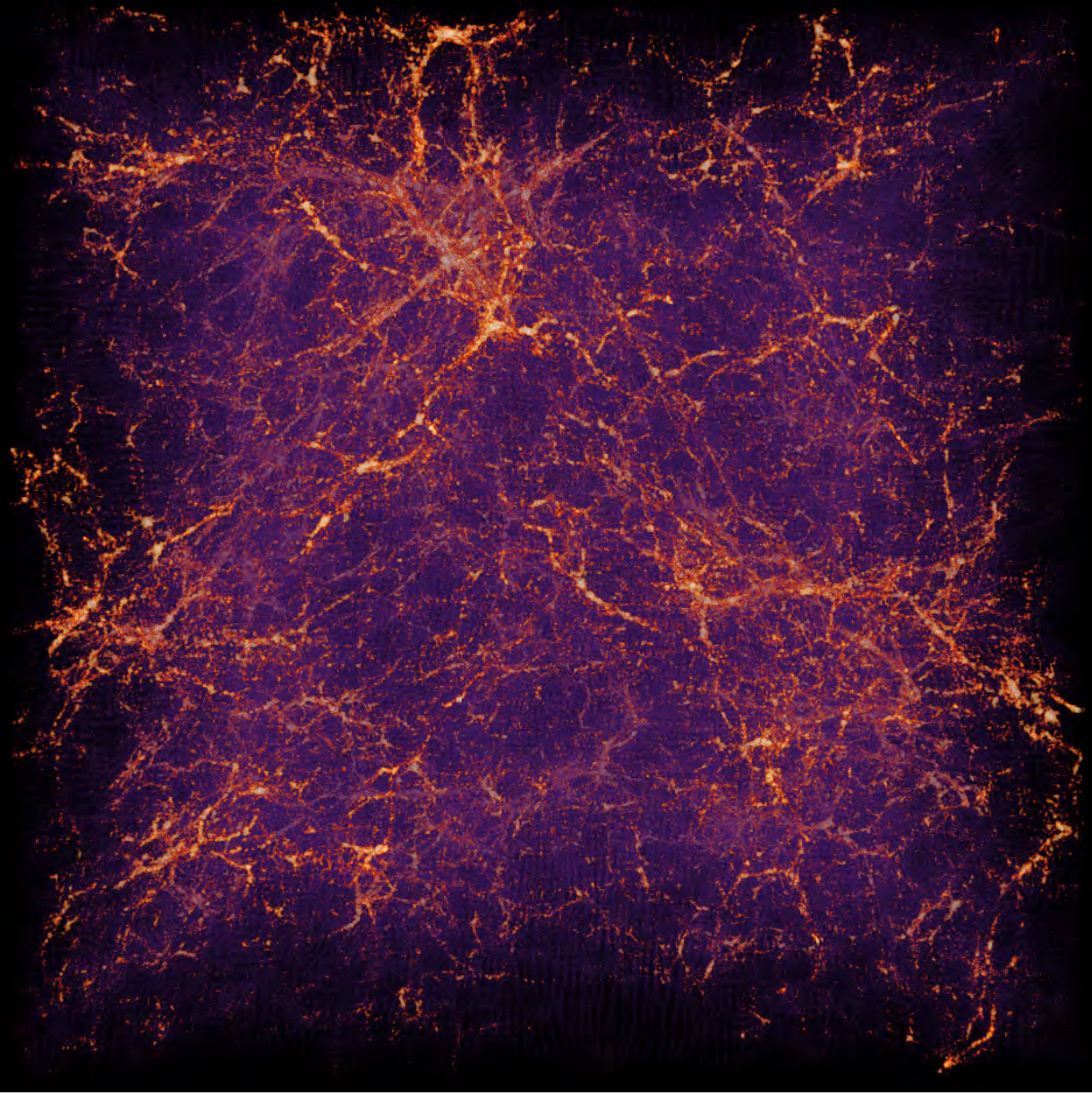
Data courtesy of: Subramanian Sankaranarayanan, Argonne National Laboratory





Data courtesy of: Advanced Photon Source, Argonne National Laboratory

Cosmology



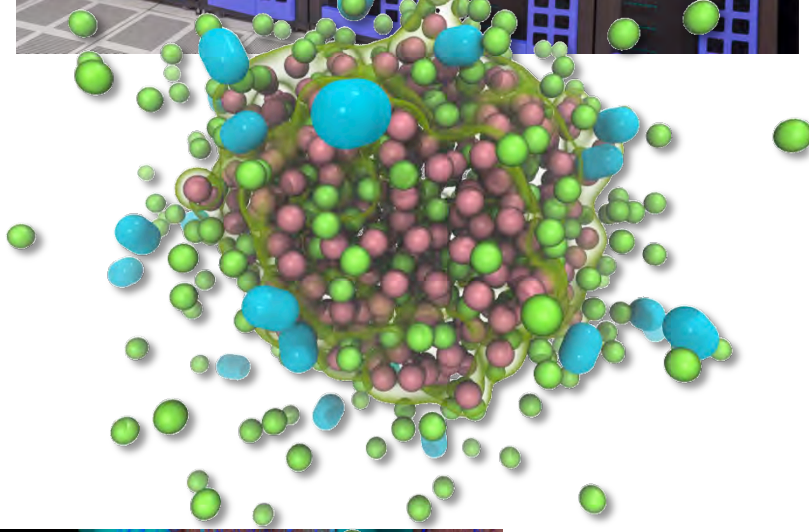
Data courtesy of: Salman Habib, Katrin Heitmann, and the HACC team, Argonne National Laboratory

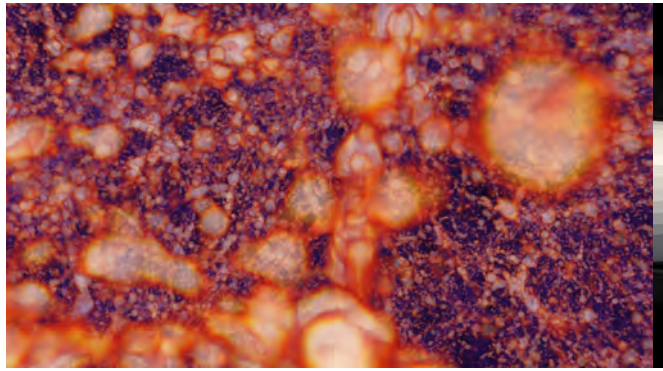
Argonne NATIONAL LABORATORY

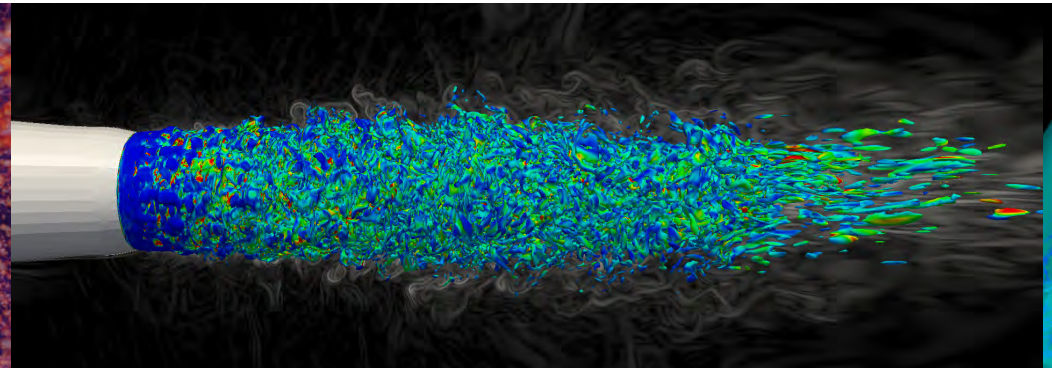
ECP EXASCALE COMPUTING PROJECT

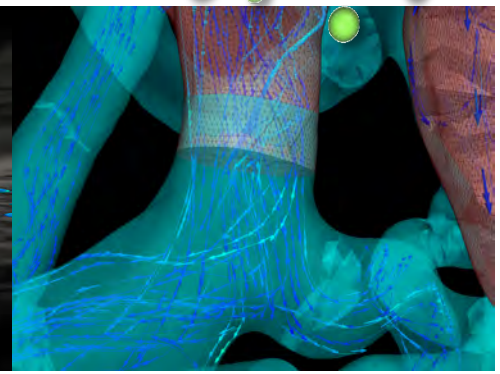
Cooley

- Analytics/Visualization cluster
- Peak 223 TF
- 126 nodes; each node has
  - Two Intel Xeon E5-2620 Haswell 2.4 GHz 6-core processors
  - NVIDIA Telsa K80 graphics processing unit (24GB)
  - **384 GB of RAM**
- Aggregate RAM of 47 TB
- Aggregate GPU memory of ~3TB
- Cray CS System
- 216 port FDR IB switch with uplinks to our QDR infrastructure
- Mounts the same GPFS file systems as Mira, Cetus









VISUALIZATION TOOLS AND DATA FORMATS

All Sorts of Tools

- Visualization Applications
  - VisIt
  - ParaView
  - EnSight
- Domain Specific
  - VMD, PyMol, Ovito
- APIs
  - VTK: visualization
  - ITK: segmentation & registration
- GPU performance
  - vl3: shader-based volume rendering
- Analysis Environments
  - Matlab
  - Parallel R
- Utilities
  - GnuPlot
  - ImageMagick

ParaView & VisIt vs. vtk

- ParaView & VisIt
  - General purpose visualization applications
  - GUI-based
  - Scriptable
  - Extendable
  - Built on top of vtk (largely)
- vtk
  - Programming environment / API
  - Additional capabilities, finer control
  - Smaller memory footprint
  - Requires more expertise (build custom applications)

Data File Formats (ParaView & VisIt)

- VTK
- Parallel (partitioned) VTK
- VTK MultiBlock (MultiGroup, Hierarchical, Hierarchical Box)
- Legacy VTK
- Parallel (partitioned) legacy VTK
- EnSight files
- EnSight Master Server
- Exodus
- BYU
- XDMF
- PLOT2D
- PLOT3D
- SpyPlot CTH
- HDF5 raw image data
- DEM
- VRML
- PLY
- Polygonal Protein Data Bank
- XMol Molecule
- Stereo Lithography
- Gaussian Cube
- Raw (binary)
- AVS
- Meta Image
- Facet
- PNG
- SAF
- LS-Dyna
- Nek5000
- OVERFLOW
- paraDIS
- PATRAN
- PFLOTRAN
- Pixie
- PuReMD
- S3D
- SAS
- Tetrad
- UNIC
- VASP
- ZeusMP
- ANALYZE
- BOV
- GMV
- Tecplot
- Vis5D
- Xmdv
- XSF

Data Wrangling

- XDMF
  - XML wrapper around HDF5 data
  - API for writing from simulation code
  - Can define
    - data sets, subsets, hyperslabs
- vtk
  - Could add to your simulation code
  - Can write small utilities to convert data
    - Use your own read routines
    - Write vtk data structures
  - C++ and Python bindings

Data Organization

- Format
  - Existing tools support many flavors
  - Use one of these formats
  - Use (or write) a format converter
  - Write a custom reader for existing tool
  - Write your own custom vis tool
- Serial vs. Parallel/Partitioned
  - Single big file vs. many small files: middle ground generally best
    - vtk data types
    - XDMF for HDF5 (VisIt and ParaView)
    - Custom

Data Organization

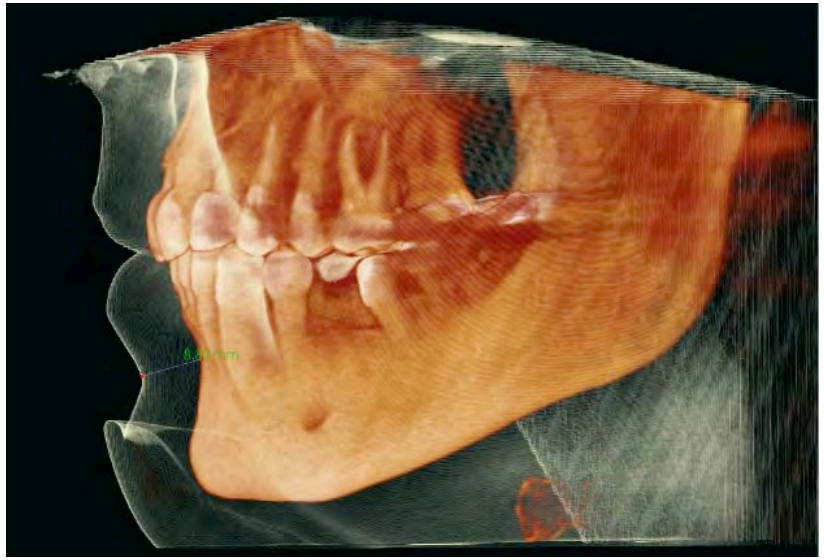
- Serial vs. Parallel/Partitioned
  - Performance trade-offs
    - vtk/paraview: serial files all data read on head node, partitioned and distributed
    - vtk/paraview: parallel files: serial files partitioned

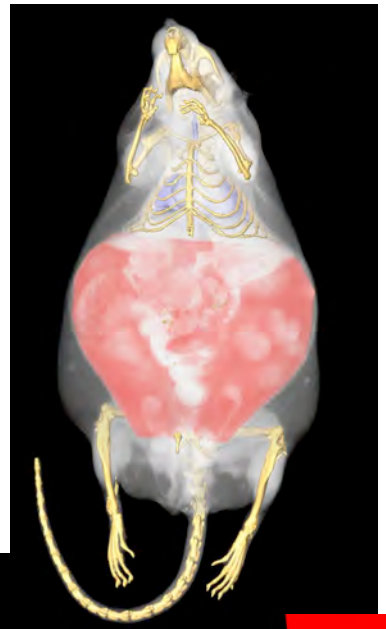
Performance example:

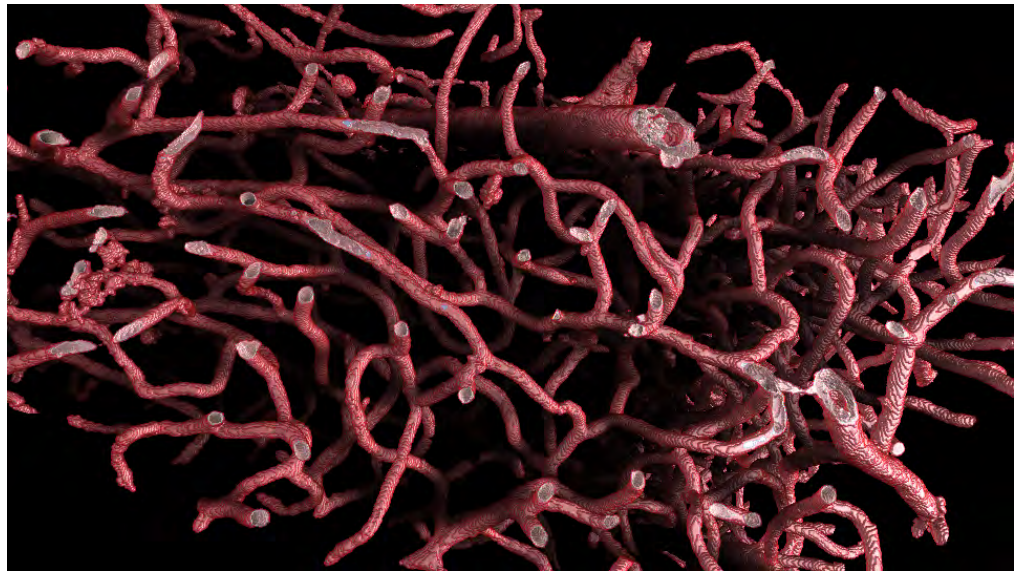
- Single serial .vtu file (unstructured grid)
  - Data size: ~3.8GB
  - Read time on 64 processes: > 15 minutes
    - most of this was spent partitioning and distributing
- Partitioned .pvtu file (unstructured grid)
  - Data size: ~8.7GB (64 partitions)
  - Read time on 64 processes: < 1 second

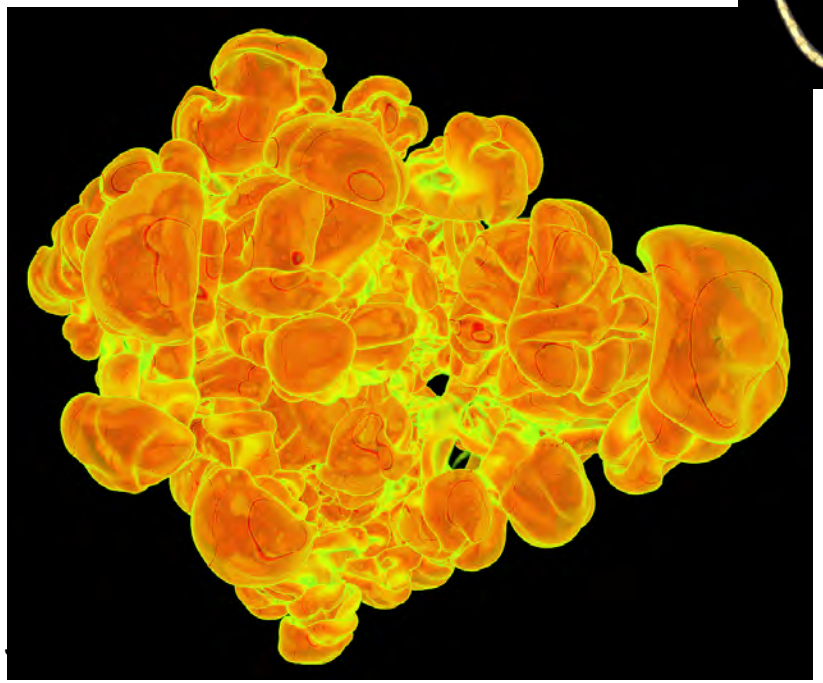
DATA REPRESENTATIONS

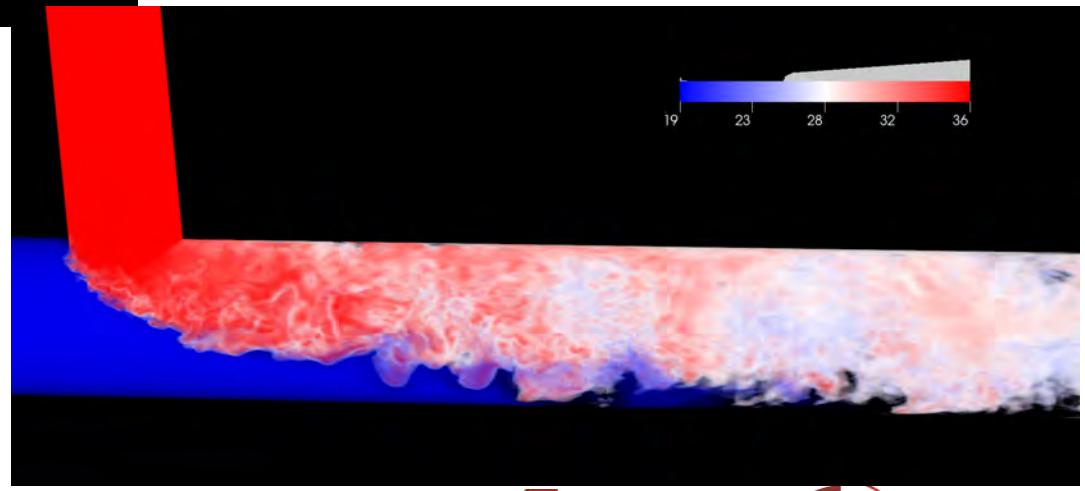
Data Representations: Volume Rendering









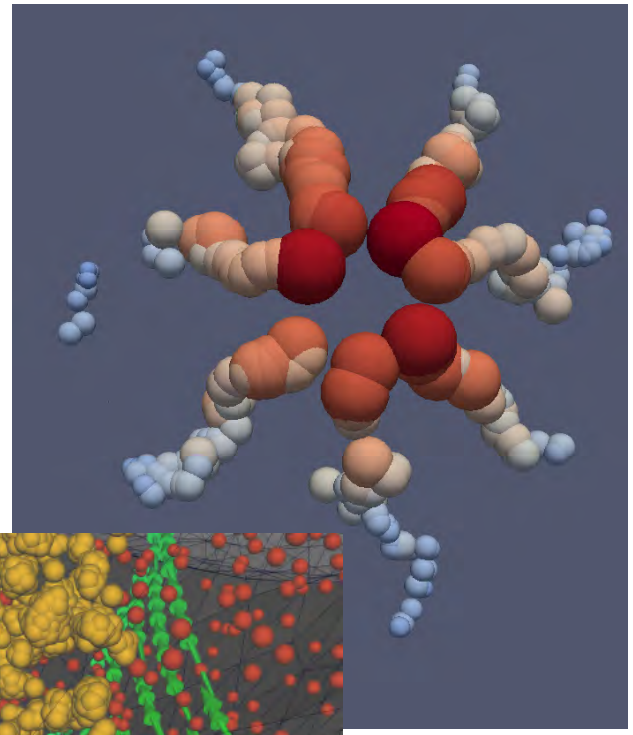


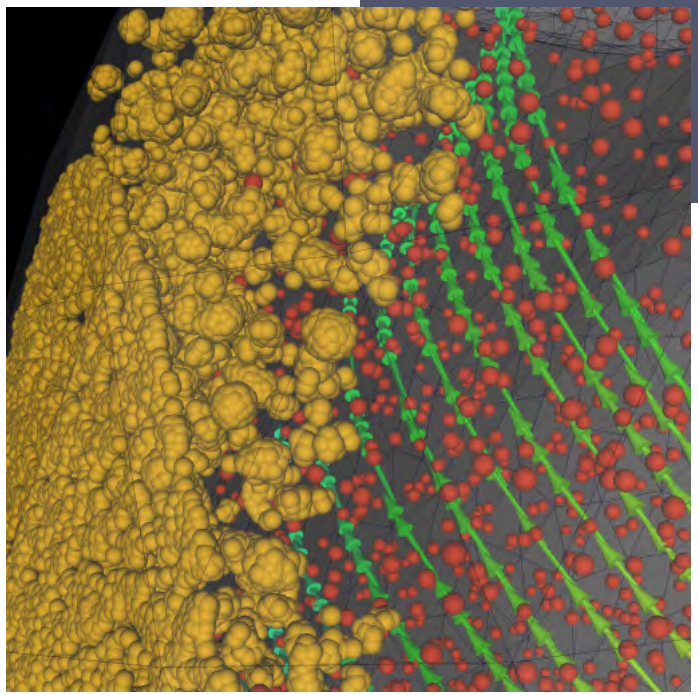
Argonne NATIONAL LABORATORY

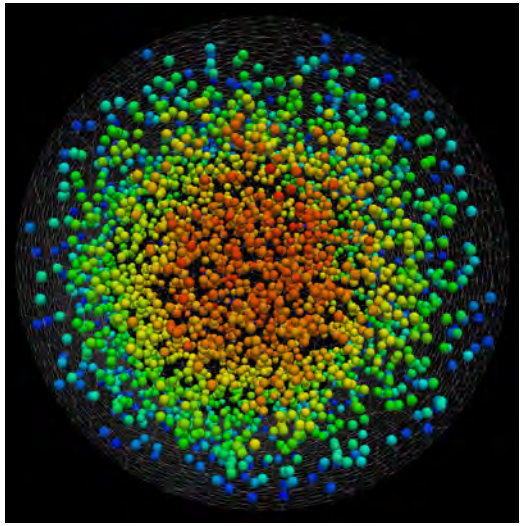
ECP EXASCALE COMPUTING PROJECT

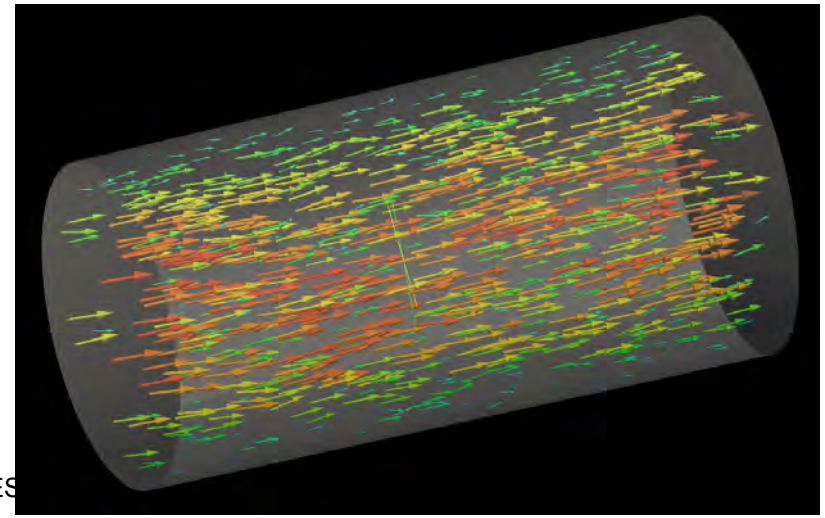
Data Representations: Glyphs

- 2D or 3D geometric object to represent point data
- Location dictated by coordinate
  - 3D location on mesh
  - 2D position in table/graph
- Attributes graphical entity dictated by attributes of a data
  - color, size, orientation







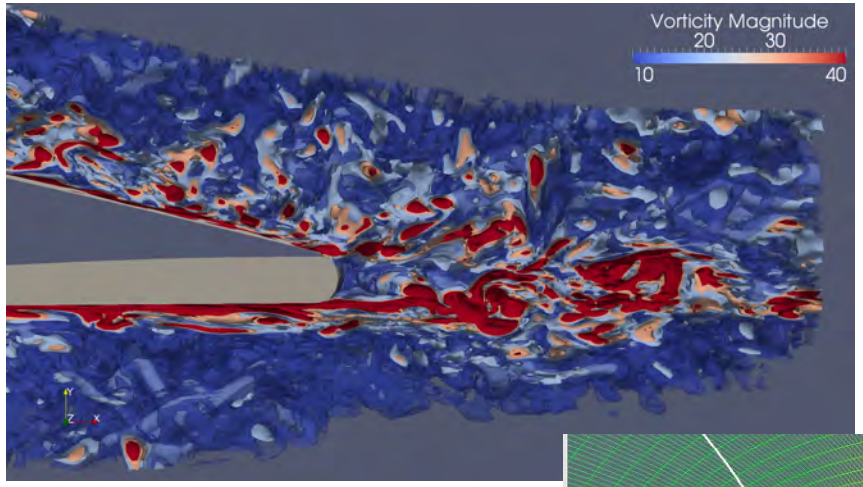


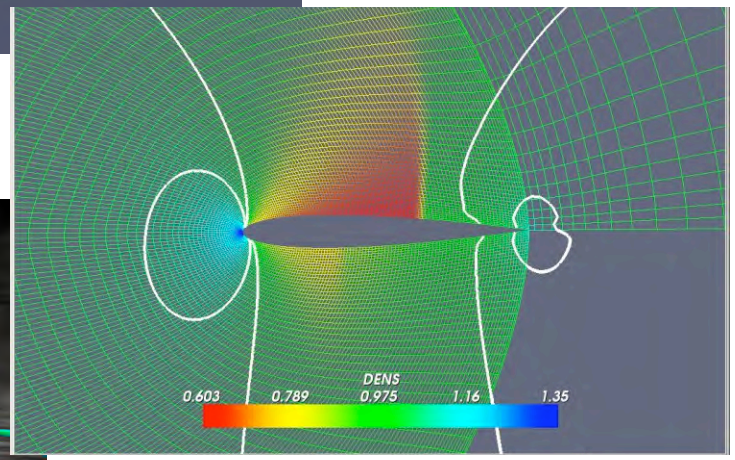
Argonne NATIONAL LABORATORY

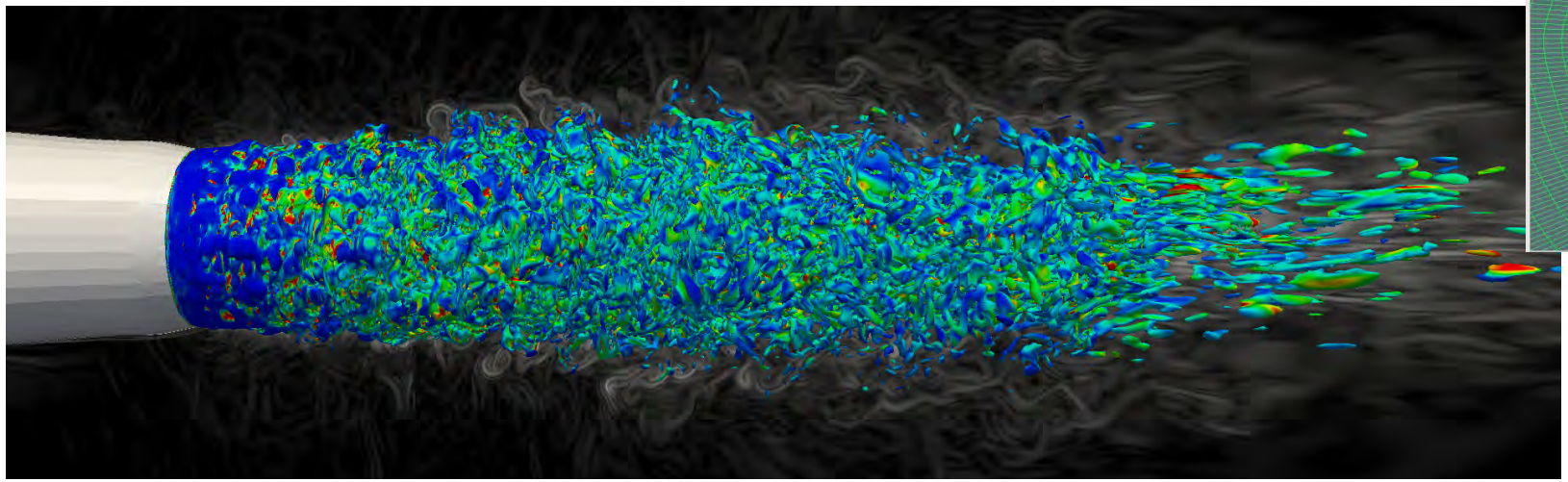
ECP EXASCALE COMPUTING PROJECT

Data Representations: Contours (Isosurfaces)

- A Line (2D) or Surface (3D), representing a constant value
- VisIt & ParaView:
  - good at this
- vtk:
  - same, but again requires more effort





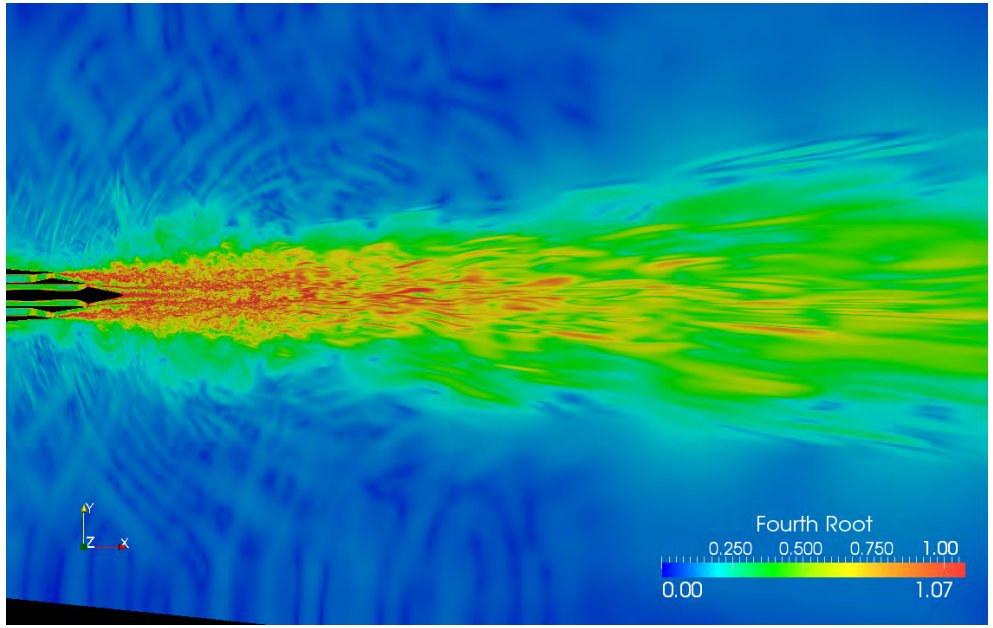


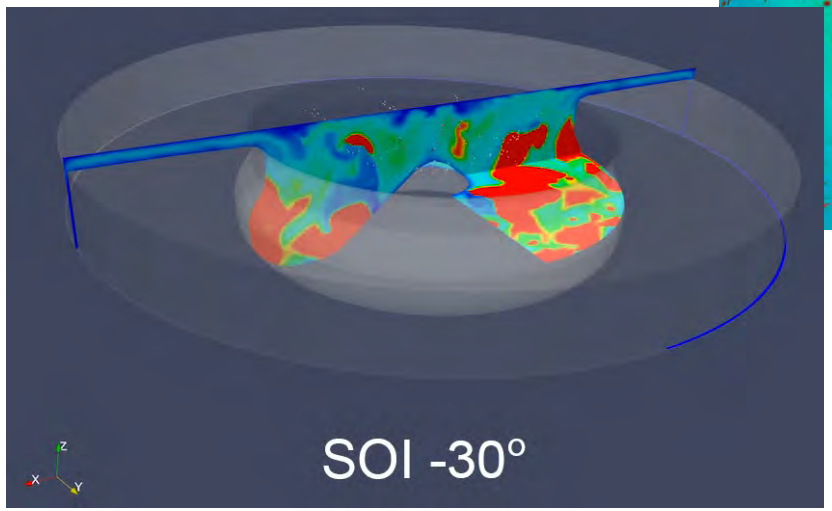
Argonne NATIONAL LABORATORY

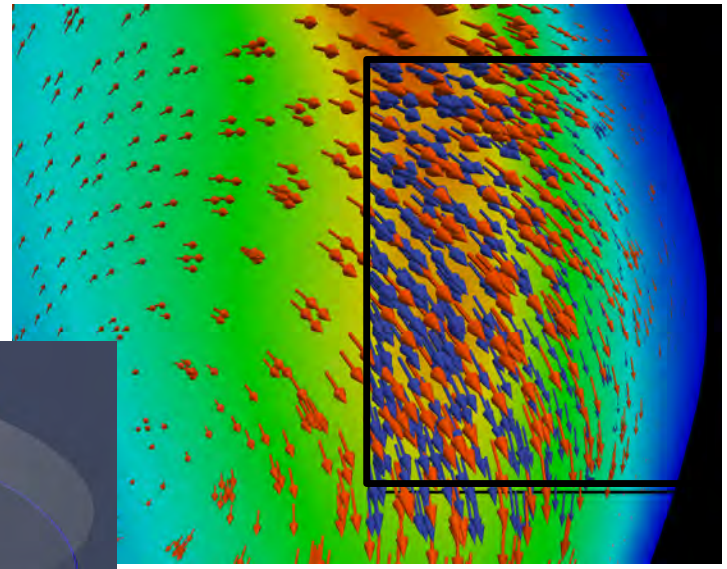
ECP EXASCALE COMPUTING PROJECT

Data Representations: Cutting Planes

- Slice a plane through the data
  - Can apply additional visualization methods to resulting plane
- VisIt & ParaView & vtk good at this
- VMD has similar capabilities for some data formats

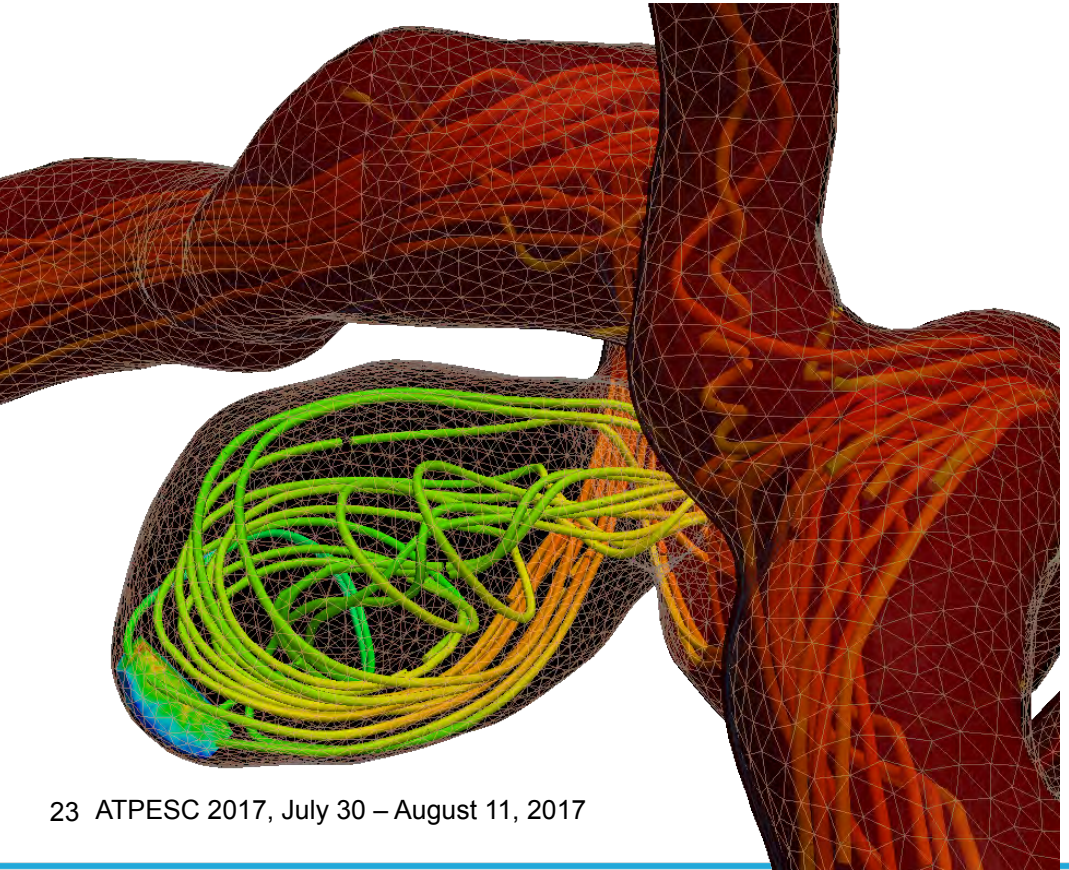


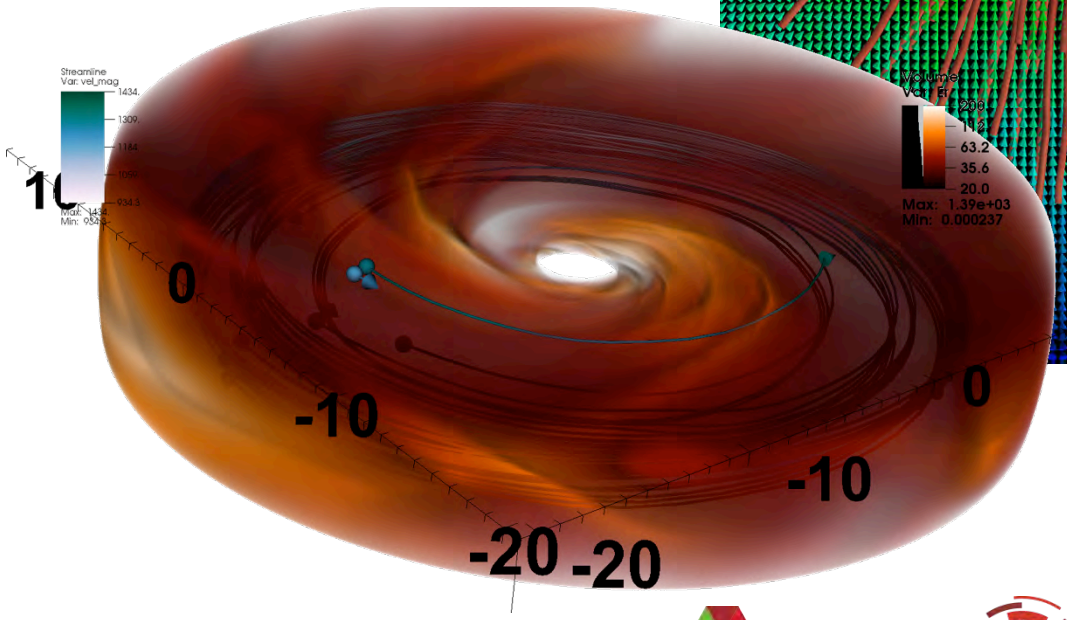


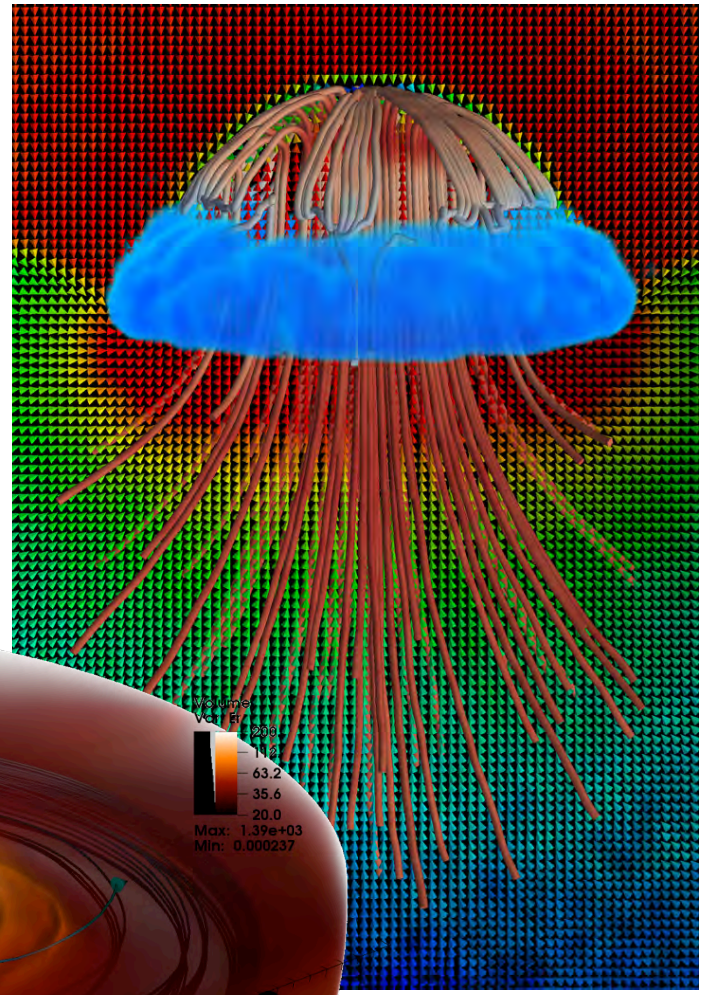


Data Representations: Streamlines

- From vector field on a mesh (needs connectivity)
  - Show the direction an element will travel in at any point in time.
- VisIt & ParaView & vtk good at this





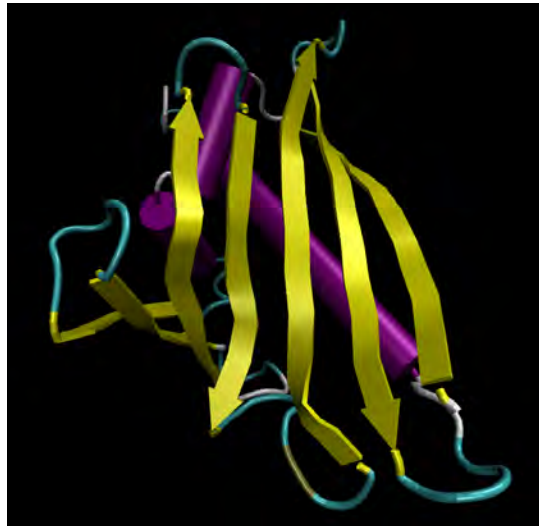


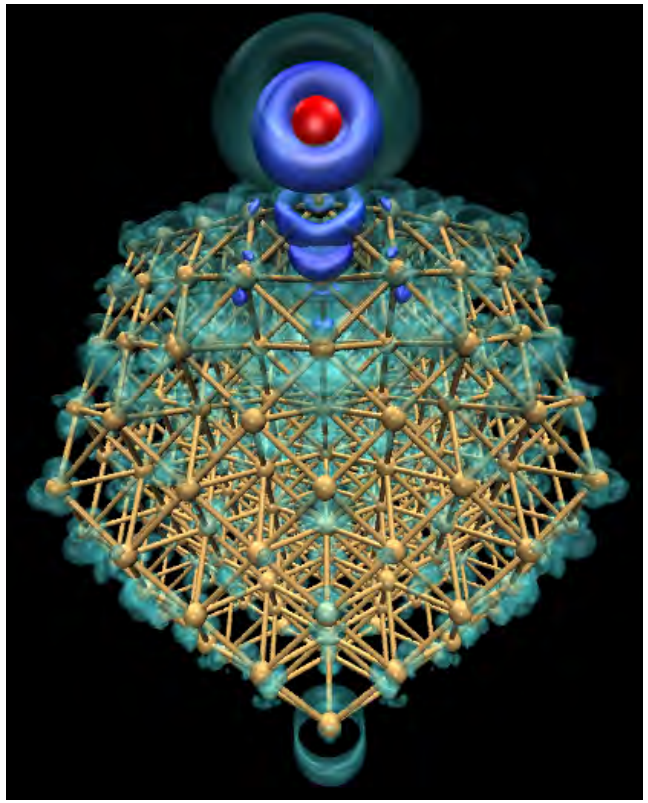
Argonne NATIONAL LABORATORY

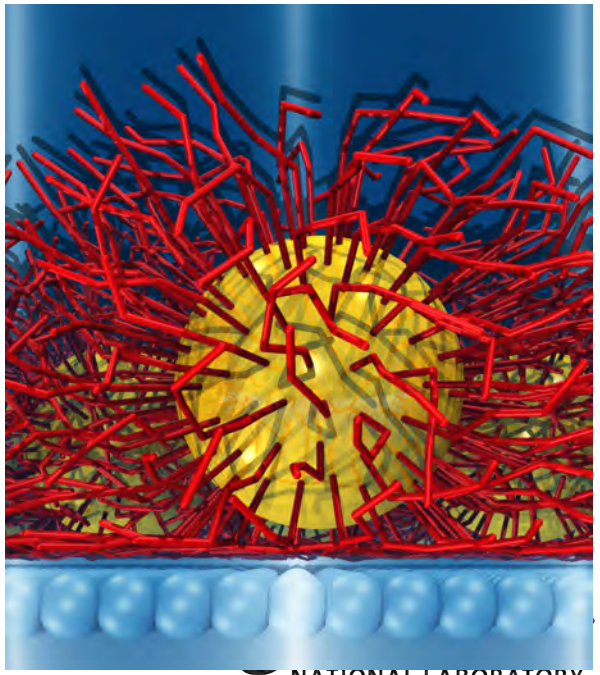
ECP EXASCALE COMPUTING PROJECT

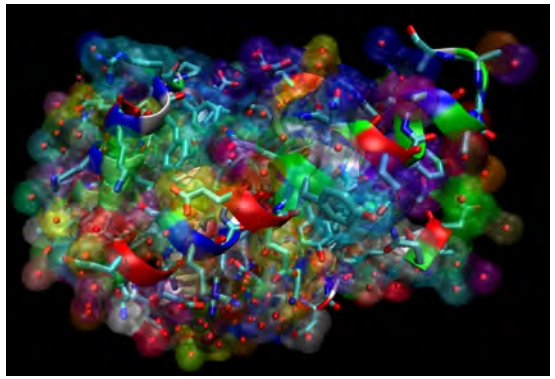
Molecular Dynamics Visualization

- VMD:
  - Lots of domain-specific representations
  - Many different file formats
  - Animation
  - Scriptable
  - Not parallel
- VisIt & ParaView:
  - Limited support for these types of representations, but improving
- VTK:
  - Anything's possible if you try hard enough





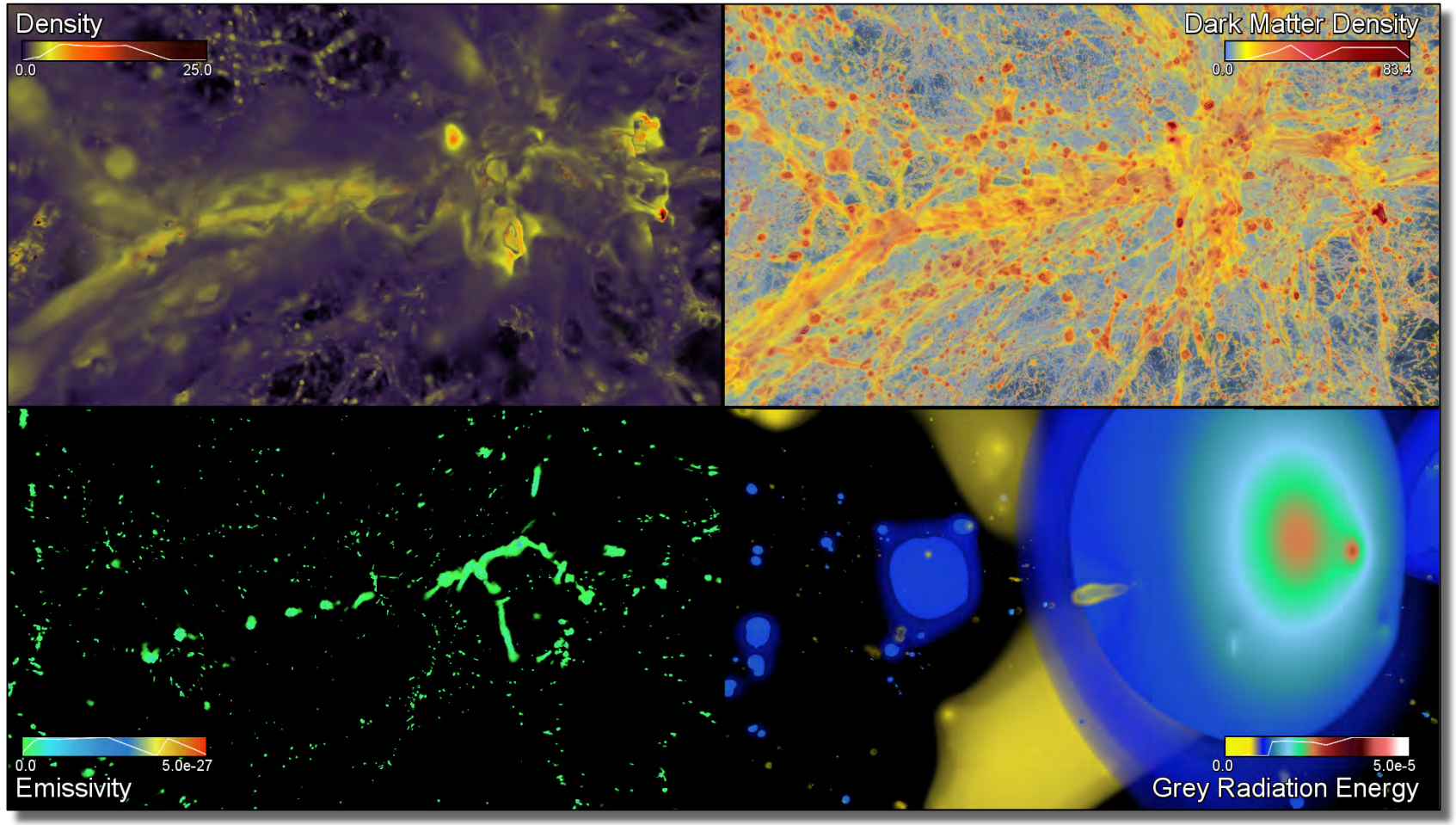




ANNOTATION AND MOVIE CREATION

Annotation, compositing, scaling…

- ImageMagick
  - convert, composite, montage, etc.

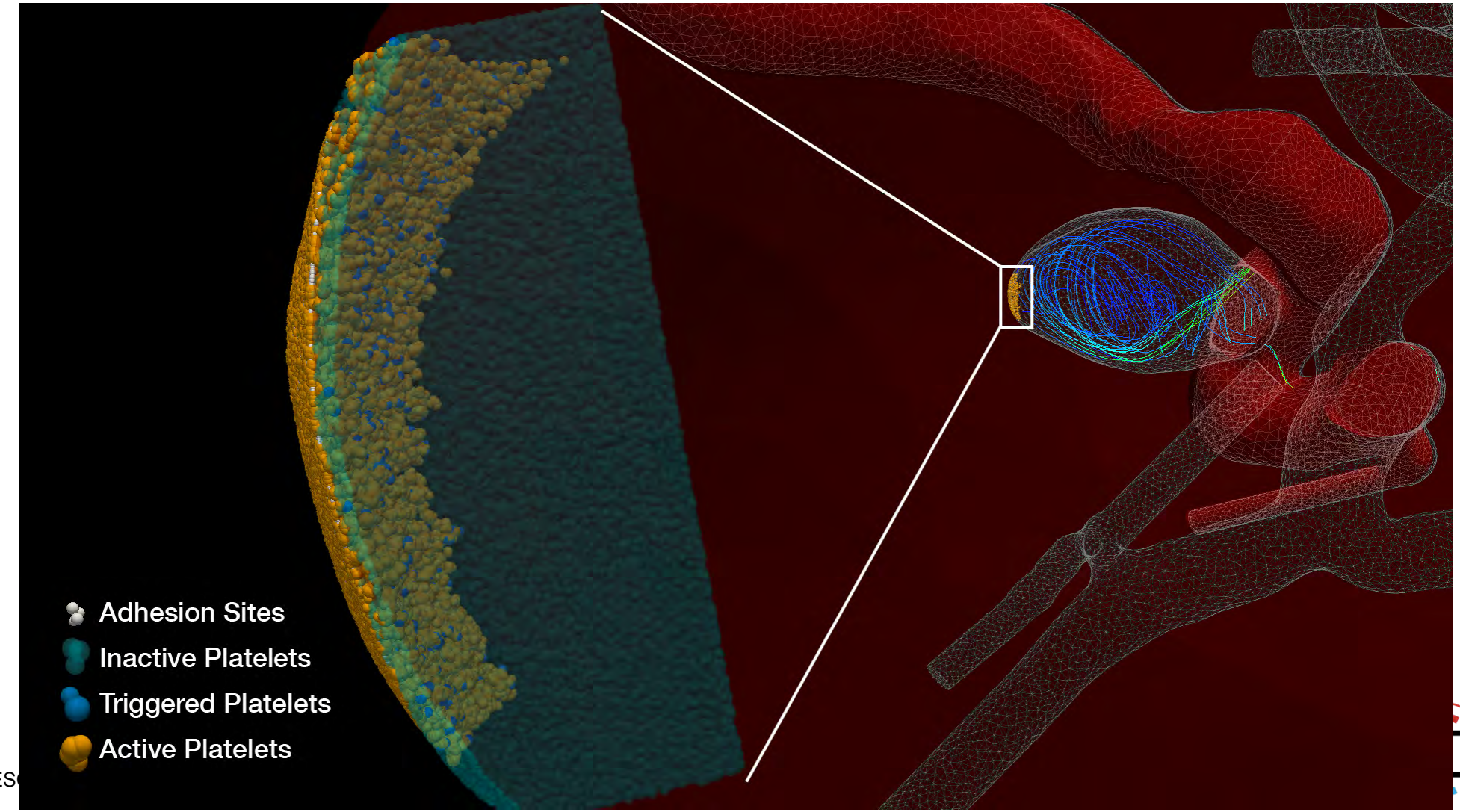


Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

Annotation, compositing, scaling…

- ImageMagick
  - scale, fade



Movie Creation

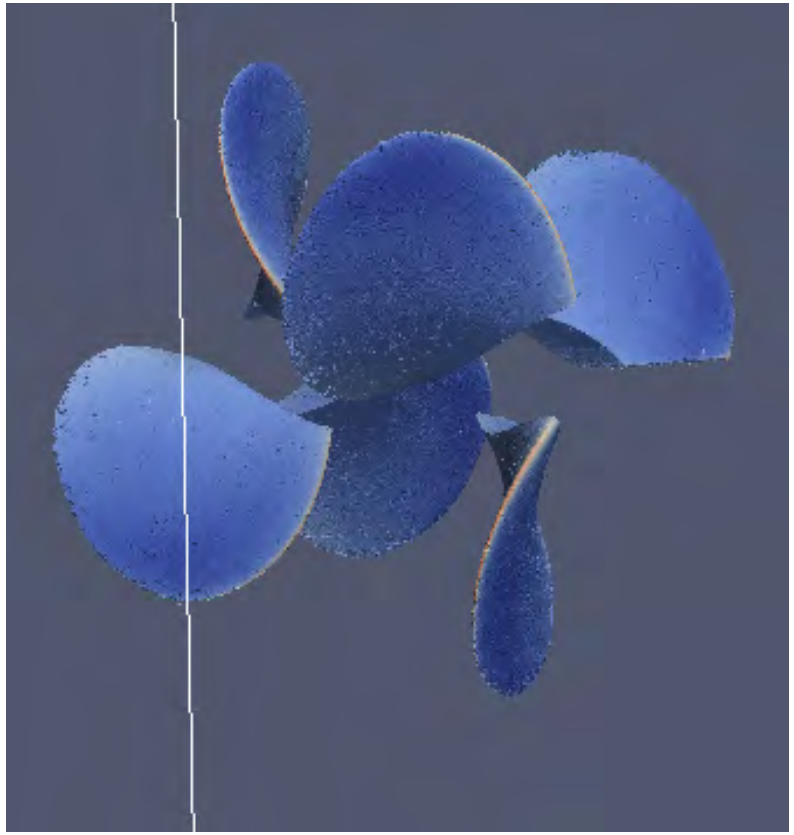
- VisIt and ParaView can spit out a movie file (.avi, etc.)
  - can also spit out individual images

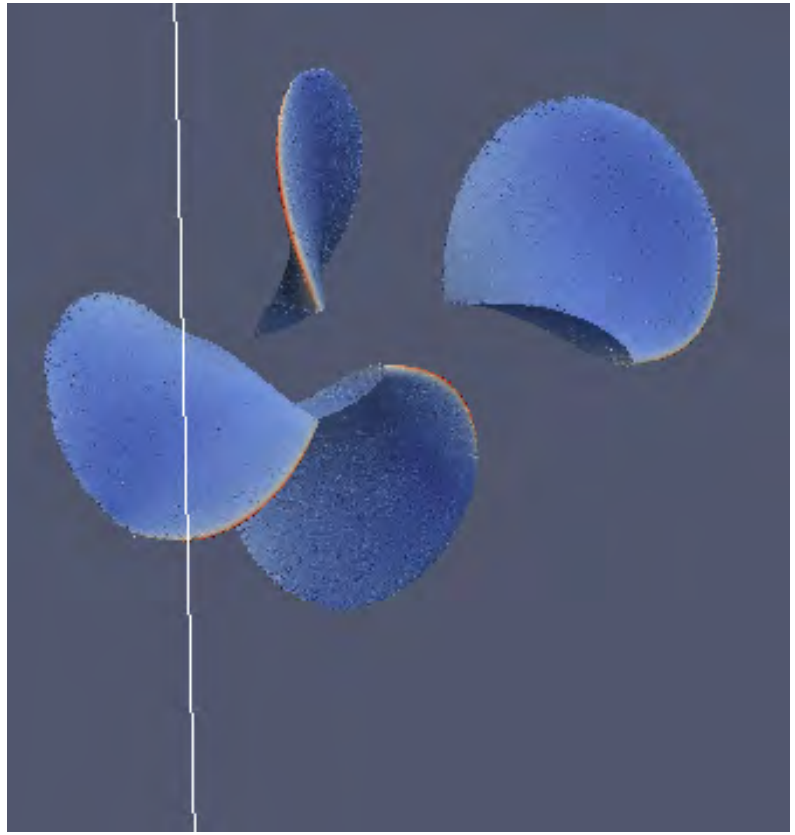
- Combine multiple segments of frames
  - Create a directory of symbolic links to all frames in order

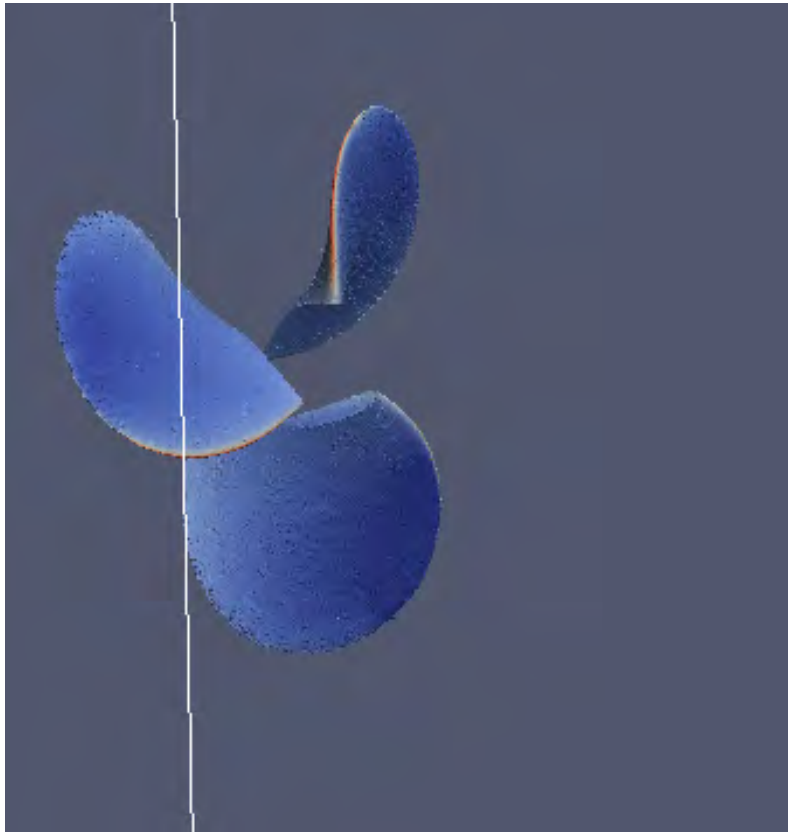
- ffmpeg: Movie encoding
  - ffmpeg –sameq –i frame.%04d.png movie.mp4

VISUALIZATION FOR DEBUGGING

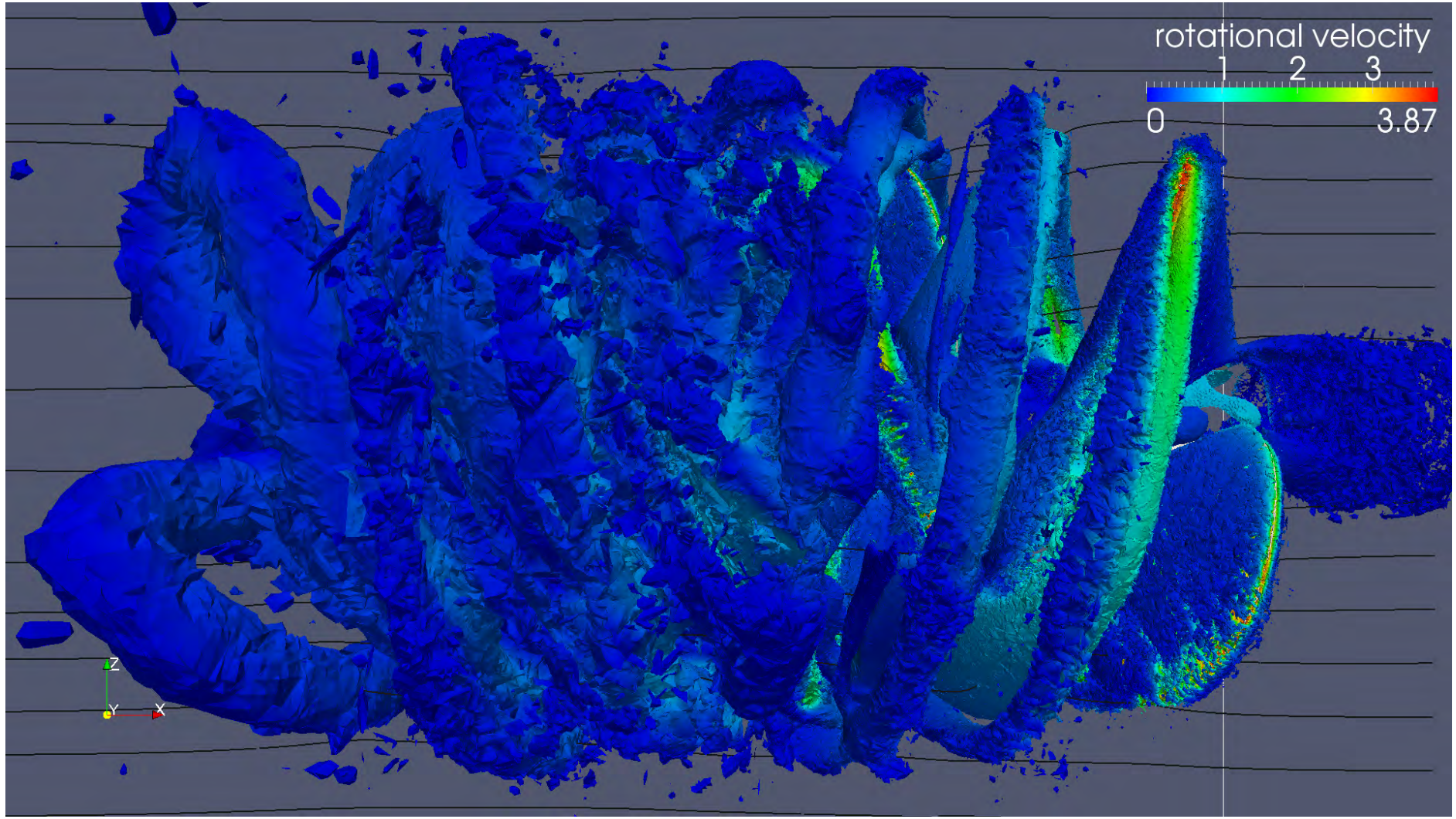
Visualization for Debugging







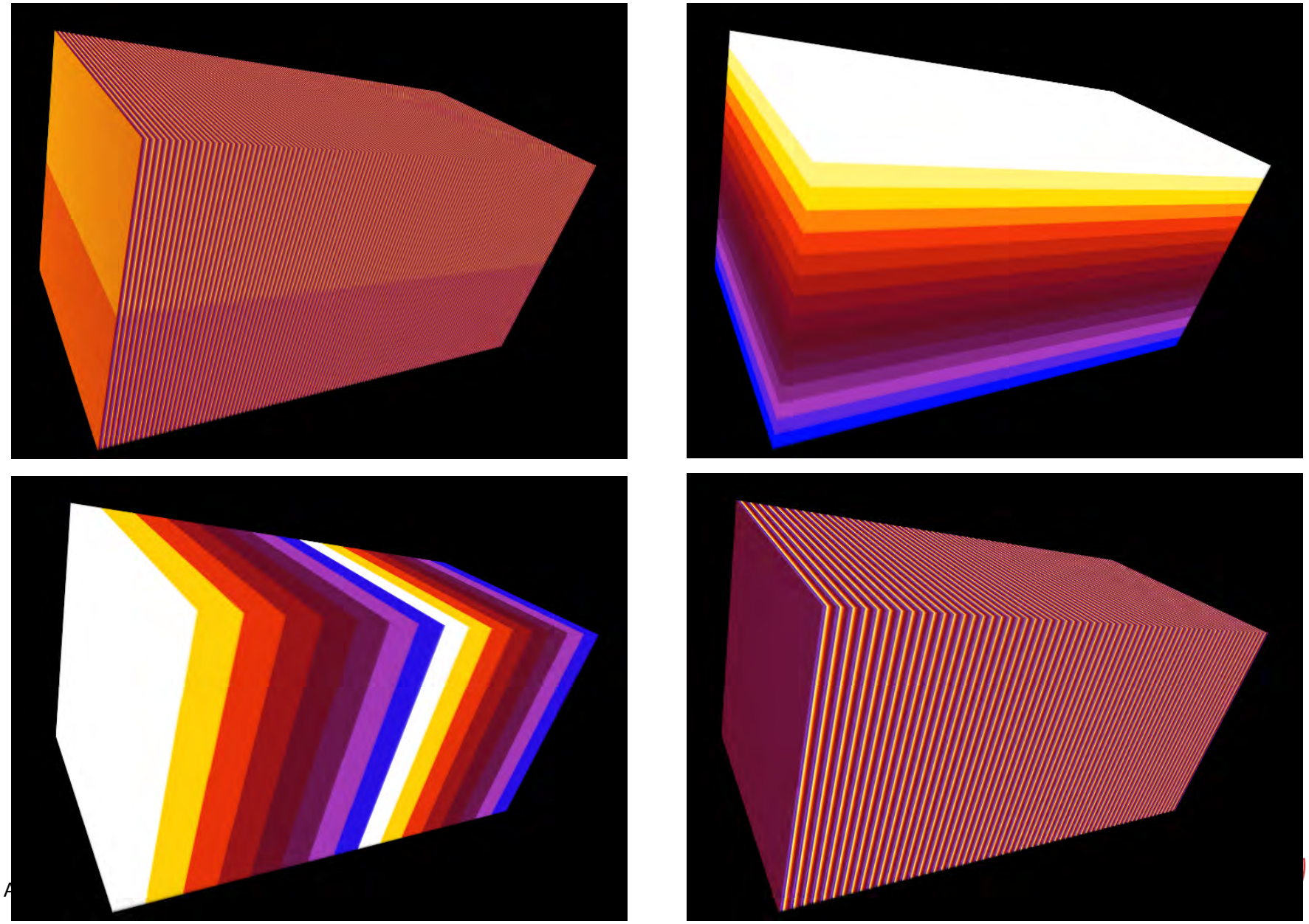
Visualization for Debugging



Argonne NATIONAL LABORATORY

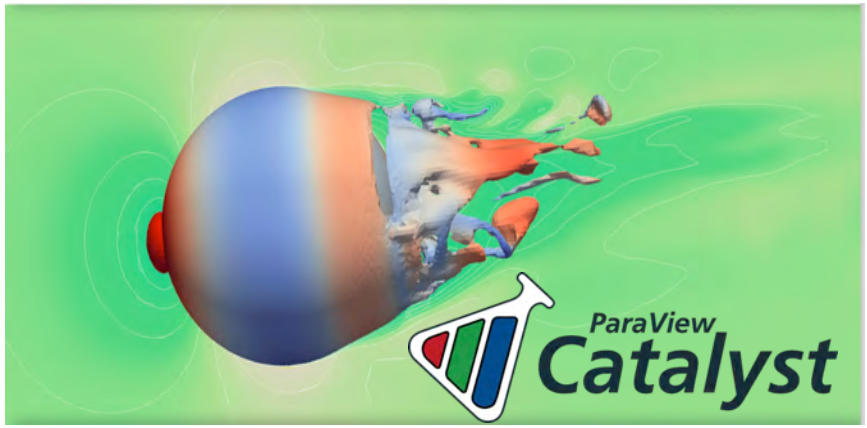
ECP EXASCALE COMPUTING PROJECT

Visualization as Diagnostics: Color by Thread ID



IN-SITU VISUALIZATION AND ANALYSIS

Multiple in-situ infrastructures



VISIT

ADIOS

LibSim

Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

Can WE….

- Enable use of any in situ framework?
- Develop analysis routines that are portable between codes?
- Make it easy to use?

OUR APPROACH

- Data model – to pass data between Simulation & Analysis
- API – for instrumenting simulation and analysis codes

SENSEI
insitu

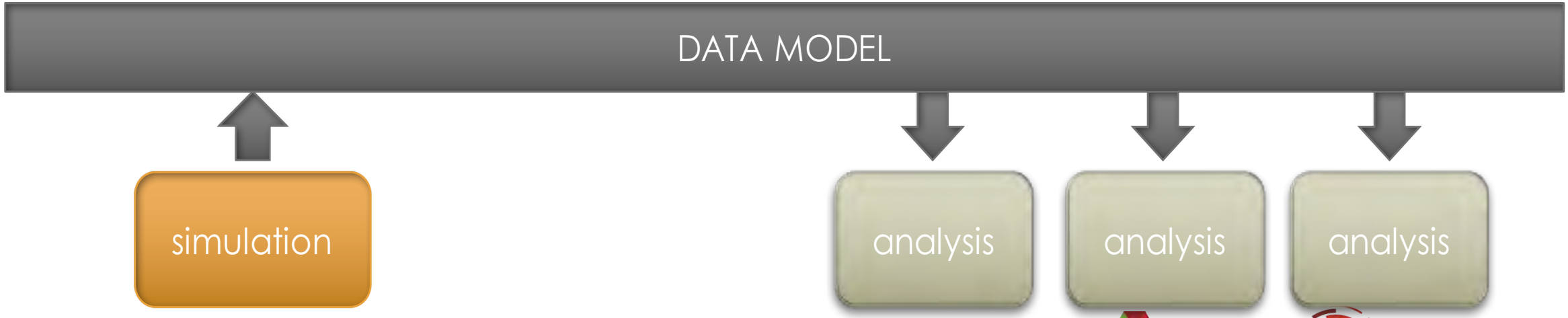
Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

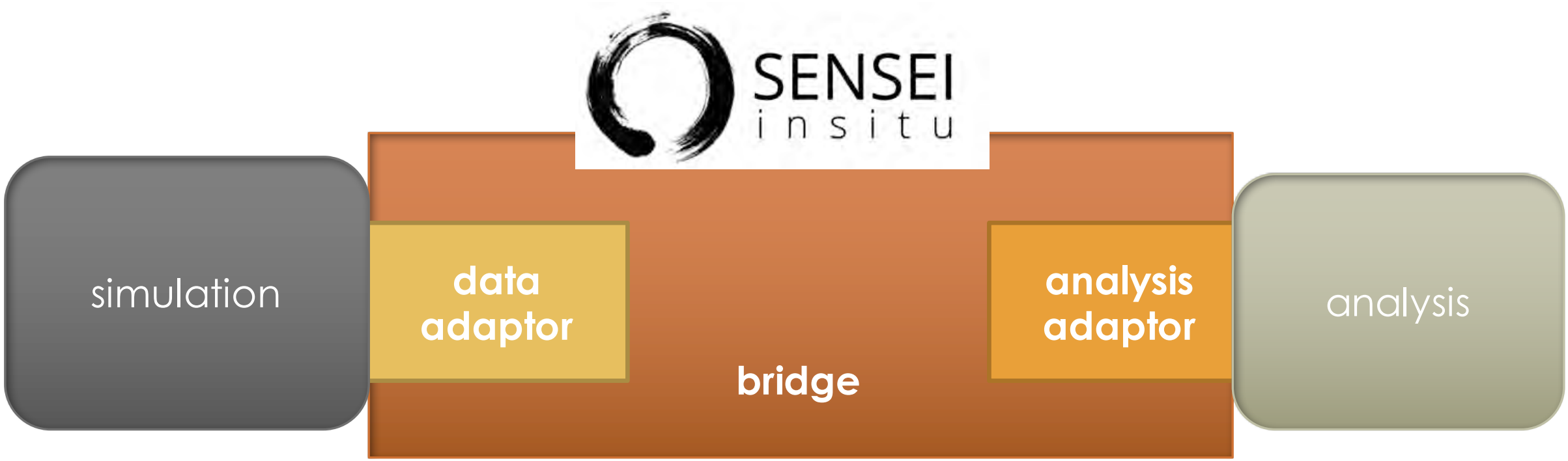
Data Model: VTK

- Used by ParaView/Catalyst and VisIt/Libsim
- Supports common scientific dataset types
- On going independent efforts to evolve for exascale
- Supports using simulation memory directly (zero-copy) for multiple memory layouts

http://www.vtk.org/



Sensei: API: Components



Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

INSTRUMENTATION TASKS

FOR SIMULATION

- Write a Data Adaptor to map simulation data to VTK data model
- Write a Bridge to define API entry points for simulation

FOR ANALYSIS

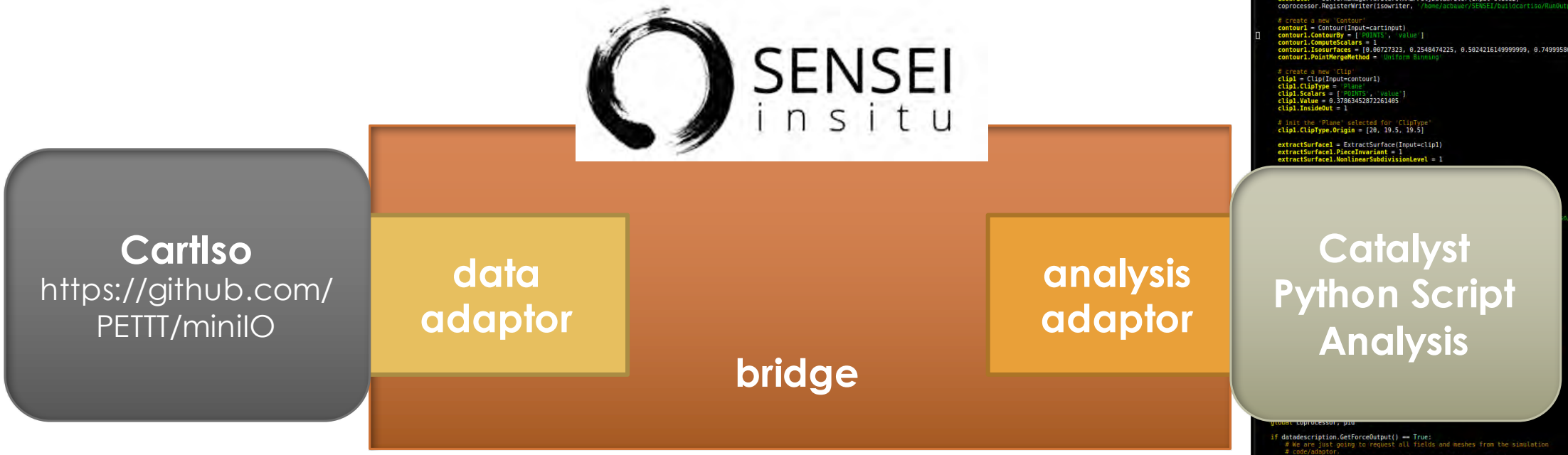
- Write analysis adaptor that uses Data Adaptor API to access Data
- Transform data, if needed and invoke analysis

Adding A catalyst Python Script analysis

- 13 lines of CMake code changes
- 18 lines of C++ code
- In situ work can be specified via SENSEI XML

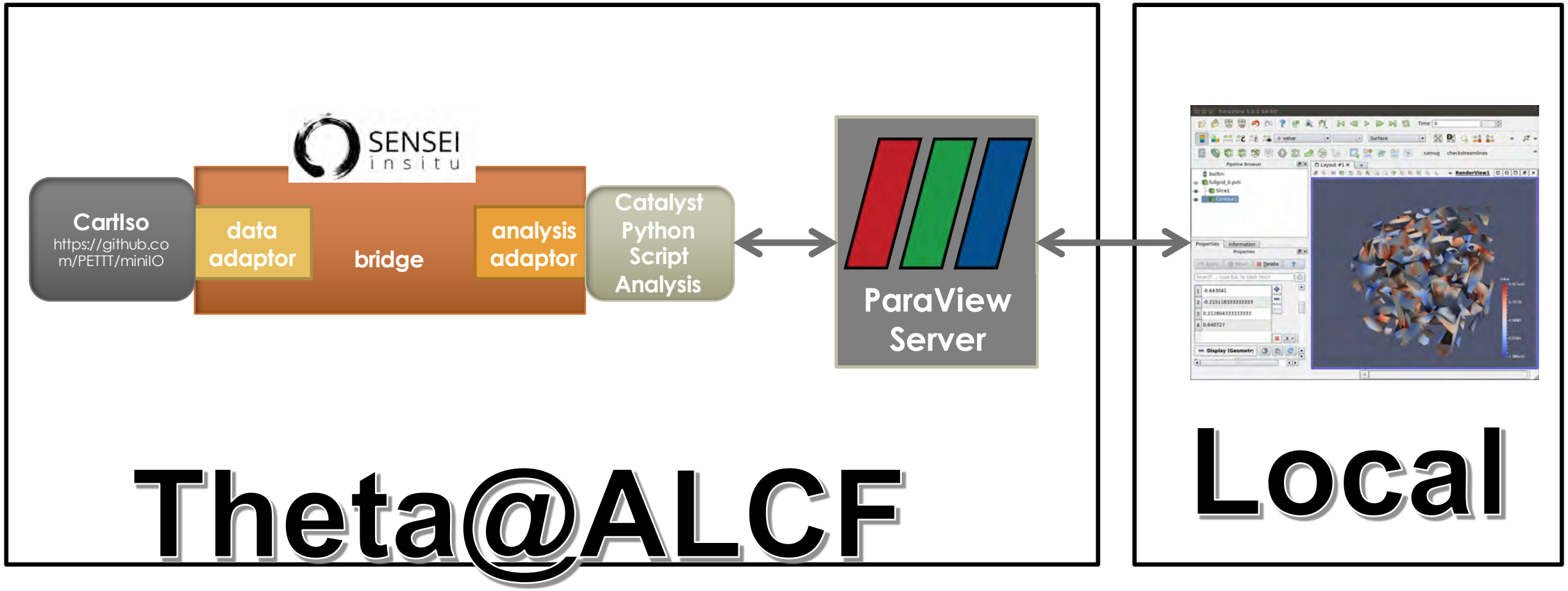
```
<sensei>
  <analysis type="catalyst" pipeline="pythonscript" filename="slice_contourcut.py"/>
  <analysis type="autocorrelation" array="data" association="cell" window="10" k-max="3"/>
  <analysis type="adios" filename="oscillators.bp" method="MPI"/>
  <analysis type="libsim" options="-no-icet" plots="Pseudocolor" plotvars="cell_data"
            visitdir="/global/homes/w/whitlocb/Development/SC16/install_static"
            slice-origin="32.5,32.5,32.5" slice-normal="0,0,1" image-filename="slice%ts"
            image-width="1600" image-height="1600" image-format="png"/>
</sensei>
```

Example with Catalyst Python script



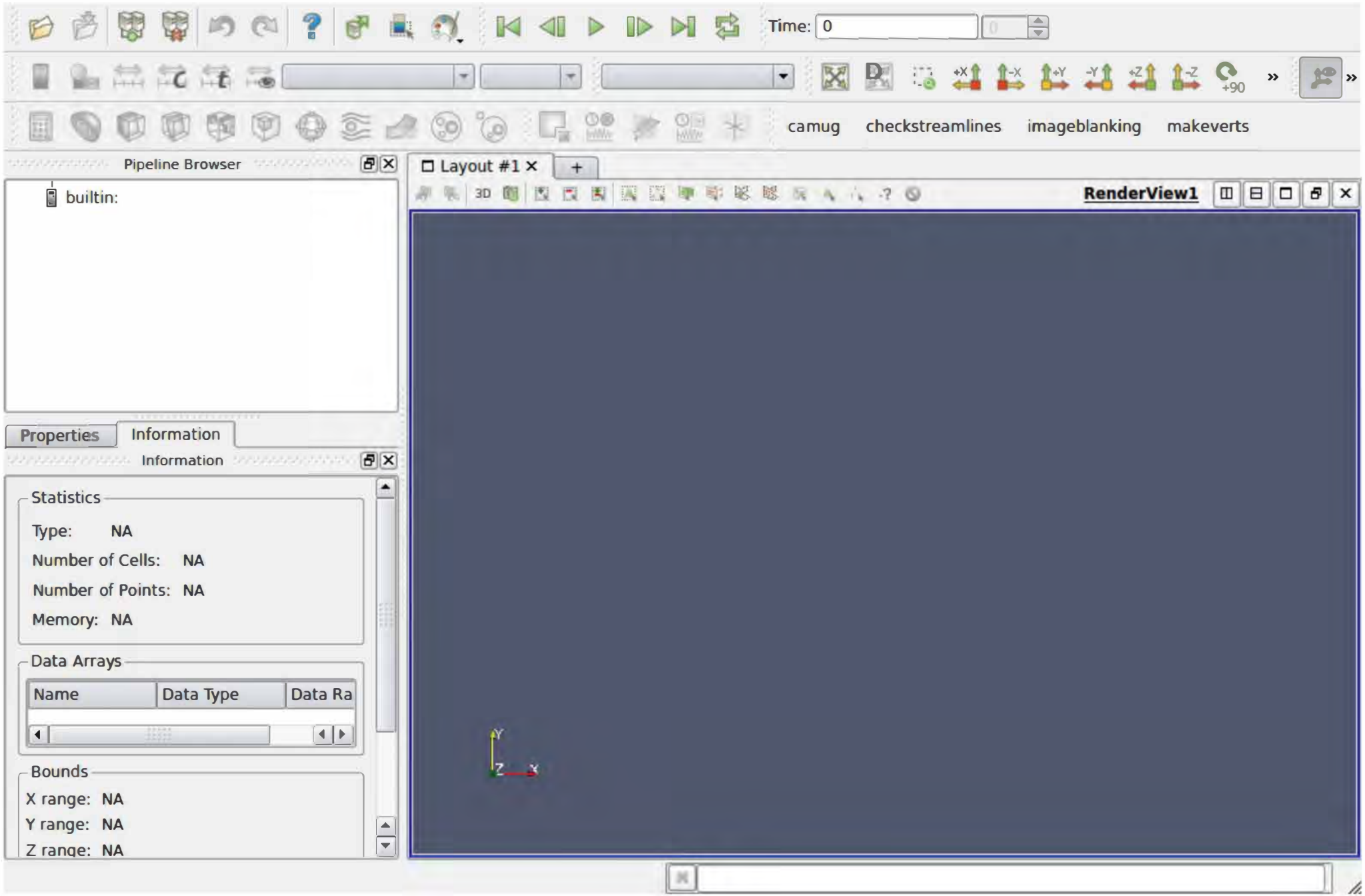

```
<sensei>
  <analysis type="catalyst" pipeline="pythonscript" filename="slice_contourcut.py" />
</sensei>
```

Catalyst Live through python script



Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT


Time: 0
camug
checkstreamlines
imageblanking
makeverts
Pipeline Browser
builtin:
Layout #1
RenderView1
Properties
Information
Information
Statistics
Type: NA
Number of Cells: NA
Number of Points: NA
Memory: NA
Data Arrays
Name
Data Type
Data Ra
Bounds
X range: NA
Y range: NA
Z range: NA

SENSEI
in situ

http://www.sensei-insitu.org/

Sensei: A Lightweight In Situ Interface for Contemporary Infrastructure Tools and Architectures

Andrew Bauer, Patrick O'Leary and Utkarsh Ayachit

QUESTIONS?

Silvio Rizzi
srizzi@anl.gov

Joe Insley
insley@anl.gov

Mike Papka
papka@anl.gov

Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT