

# Big Ideas!

from the **data, devices and interaction Laboratory**

**Thursday, February 20, 2020**

**Michael E. Papka**

Northern Illinois University

# What is a Big Ideas Class?



# What is a Big Ideas Class?



Lectures and discussions of current research and technical developments in computer science for beginning graduate research students. Topics will emphasize open problems and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.

# What is a Big Ideas Class?

Lectures and **discussions** of current **research** and technical developments in computer science for beginning graduate research students. Topics will emphasize **open problems** and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.

# What is a Big Ideas Class?

Lectures and **discussions** of current **research** and technical developments in computer science for beginning graduate research students. Topics will emphasize **open problems** and recent scientific advances. Content may vary to reflect research advances in areas such as data analytics, scientific computing, graphics and visualization.

- Who has active research projects?
- What are NIU CS faculty interests?
- Where do I get more information?
- How do I get involved?

# Bit About Me (Education)

- Northern Illinois University - Physics (BS)
- University of Illinois @ Chicago - Computer Science (MS)
- University of Chicago - Computer Science (MS, PhD)
- Not the standard education ...
  - University of Chicago - Business School (SLLP)
  - Harvard University - Business School (GMP)
  - Stanford University - Hasso Plattner Institute of Design (Design Thinking)

# Bit About Me (Professional)

- Fermi National Accelerator Laboratory (Undergraduate/Graduate)
- Argonne National Laboratory
- Northern Illinois University

Advanced Display Environments  
2000 – 2018

Scientific Visualization and Analysis  
2000 – 2018

RD2: Assistant Scientist  
1996 – 2000

RD3: Scientist  
2000 – 2001

RD5: Scientist  
2004 – 2012

Collaboration Environments  
1997 – 2008

RD4: Scientist  
2001 – 2004

Deputy Associate Laborat  
2006

2000

2005

# Bit About Me (Research - Areas/Interests)

- Advanced Display Environments
- Collaboration Technology
- High Performance Computing (Environments)
- Information Visualization
- Scientific Visualization and Analysis
- Augmented/Virtual Reality

Advanced Display Environments  
1997 – 2008

Scienza  
1997 – 2008

Scientific Visualization and Analysis  
2000 – 2018

RD2: Assistant Scientist  
1997 – 2001

RD3: Scientist  
2001 – 2004

RD5: Scientist  
2004 – 2012

Collaboration Environments  
1997 – 2008

RD4: Scientist  
2001 – 2004

Deputy Associate Laborat  
2006



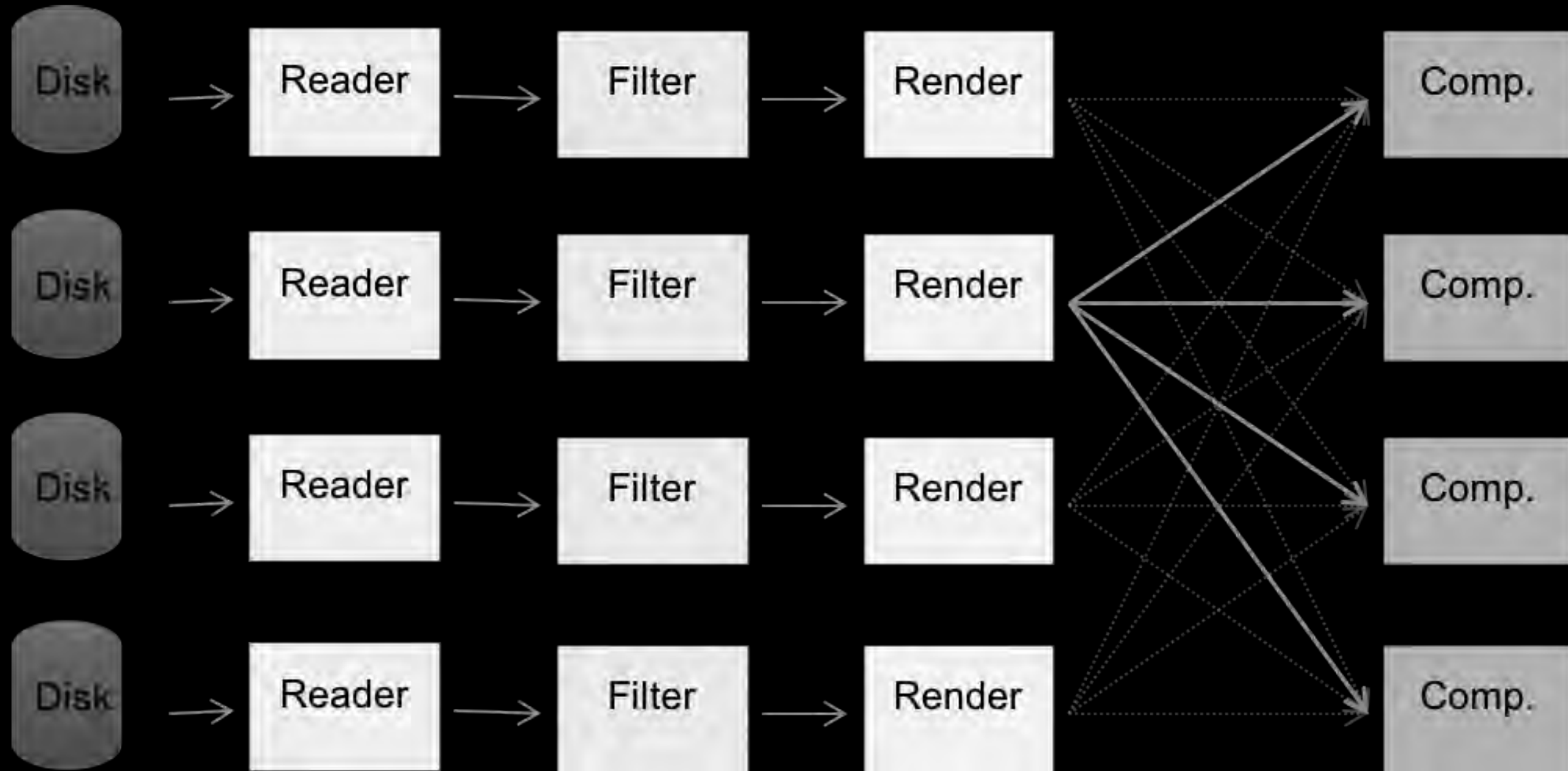
# Bit About the ddiLab

- Joint lab with **School of Art and Design**  
*Professor Joseph Insley* from Time Arts
- Emphasis on visualization and data analysis coupled to high-performance computing in the support of science
- Students
  - 1 PhD (Information Visualization)
  - 3 MS (HPC log analysis, authentication infrastructure and machine learning)
  - 3 Undergraduates (virtual reality and HPC log analysis)

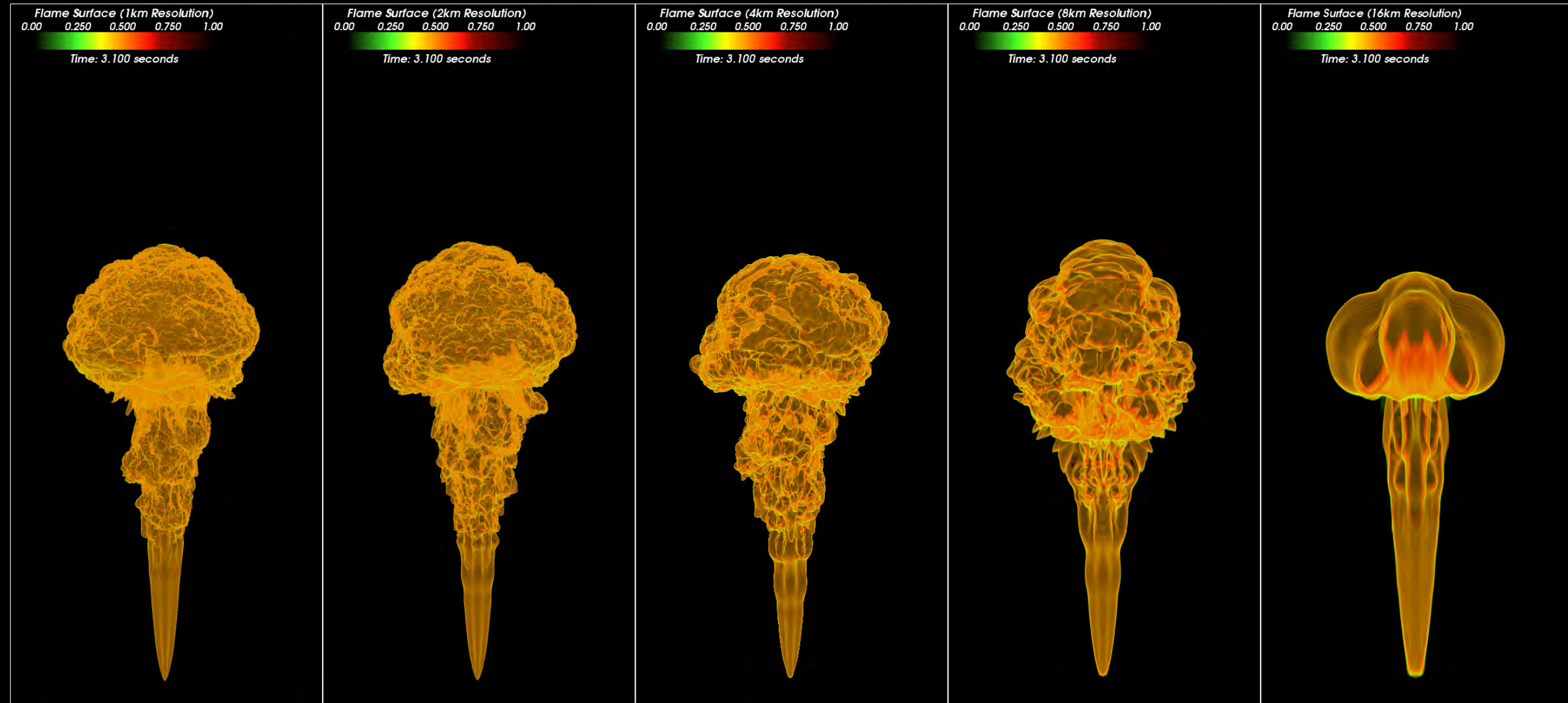
# Scientific Visualization and Analysis

- v13: volume rendering library
  - Parallel volume rendering library that exploits GPU hardware
  - Uses native data formats
- Integration with virtual and augmented reality
- Usability and collaboration
- Domain specific visualizations

# vl3: Volume Rendering Library



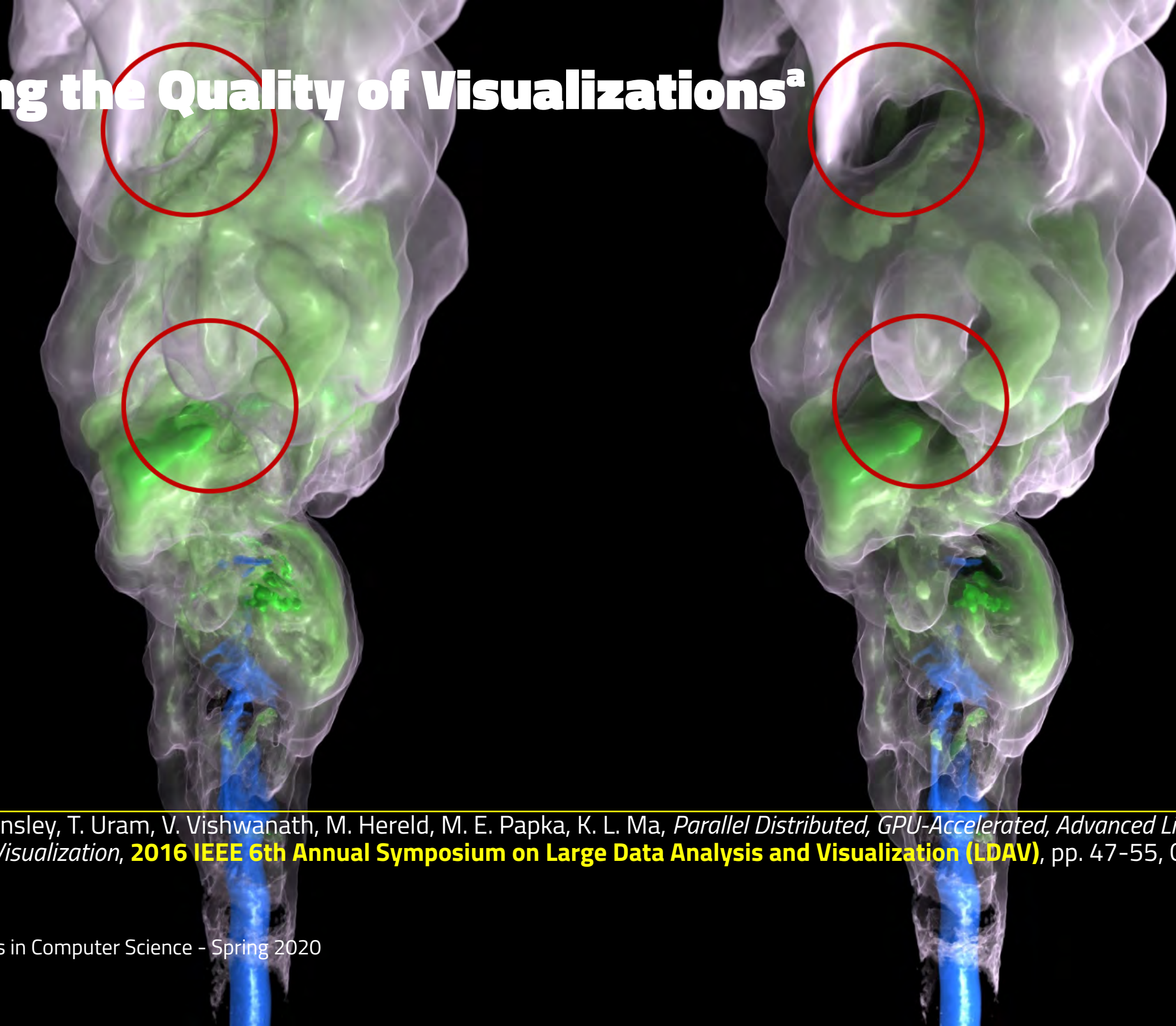
# Thermonuclear Flame Plume Rising in a Column



Local Lighting

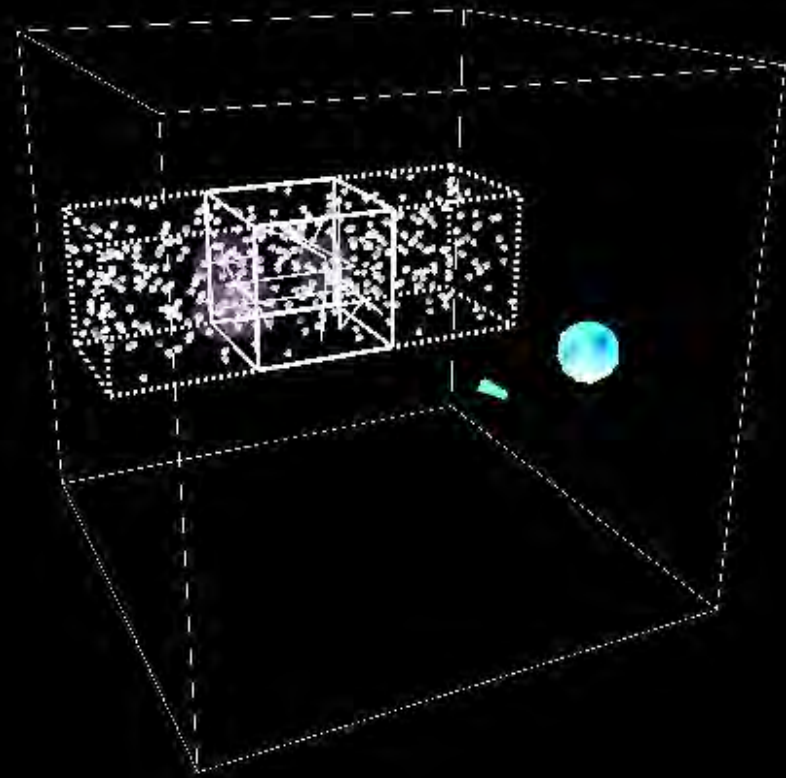
# Improving the Quality of Visualizations<sup>a</sup>

Global Lighting



<sup>a</sup>M. Shih, S. Rizzi, J. Insley, T. Uram, V. Vishwanath, M. Hereld, M. E. Papka, K. L. Ma, *Parallel Distributed, GPU-Accelerated, Advanced Lighting Calculations For Large-Scale Volume Visualization*, **2016 IEEE 6th Annual Symposium on Large Data Analysis and Visualization (LDAV)**, pp. 47-55, October 2016.]

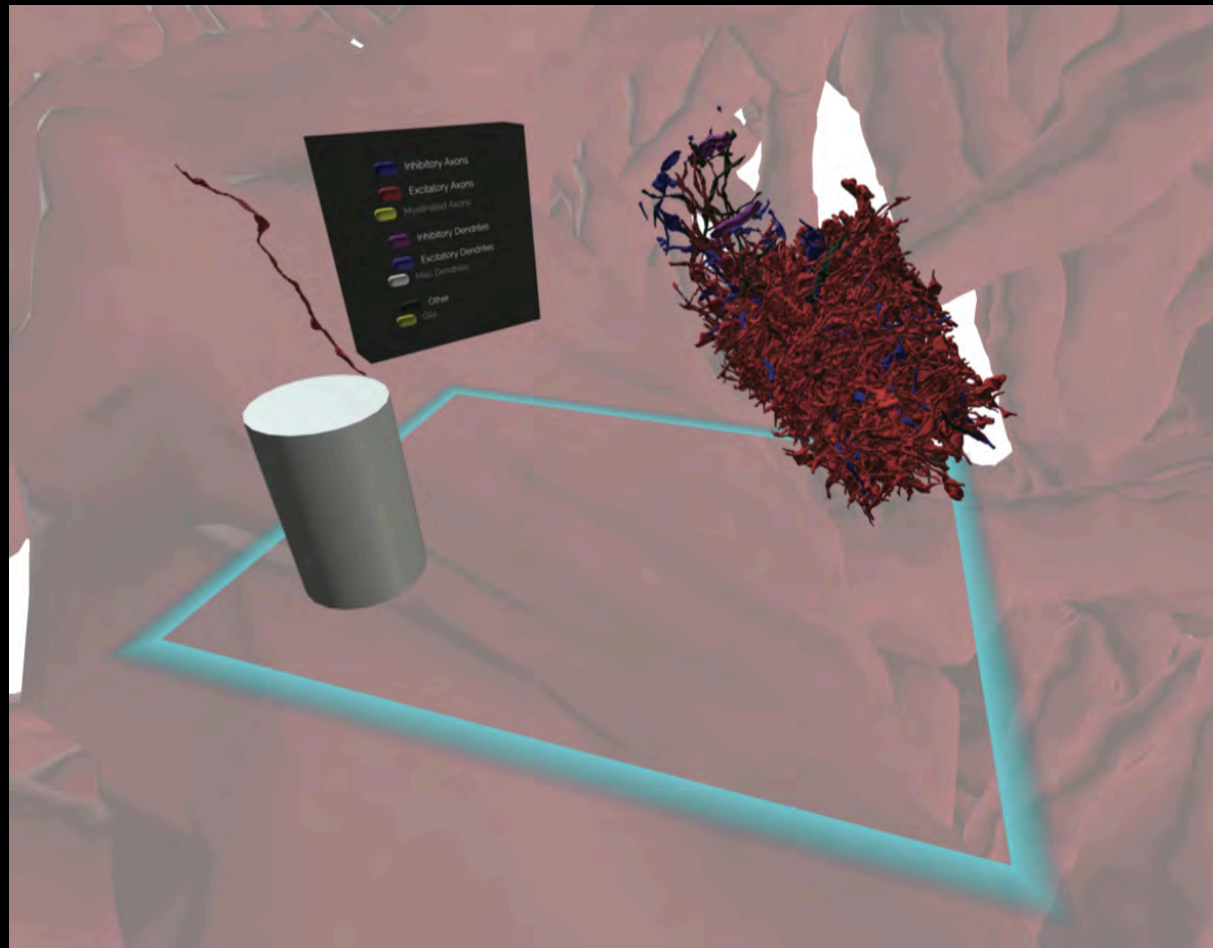
# Virtual Reality<sup>bc</sup>



<sup>b</sup>T. Disz, M. E. Papka, R. Stevens, M. Pellegrino, V. Taylor, *Virtual Reality Visualization of Parallel Molecular Dynamics Simulation*, **1995 Simulation Multiconference Symposium**, pp. 483-87, Phoenix, AZ, April 1995.

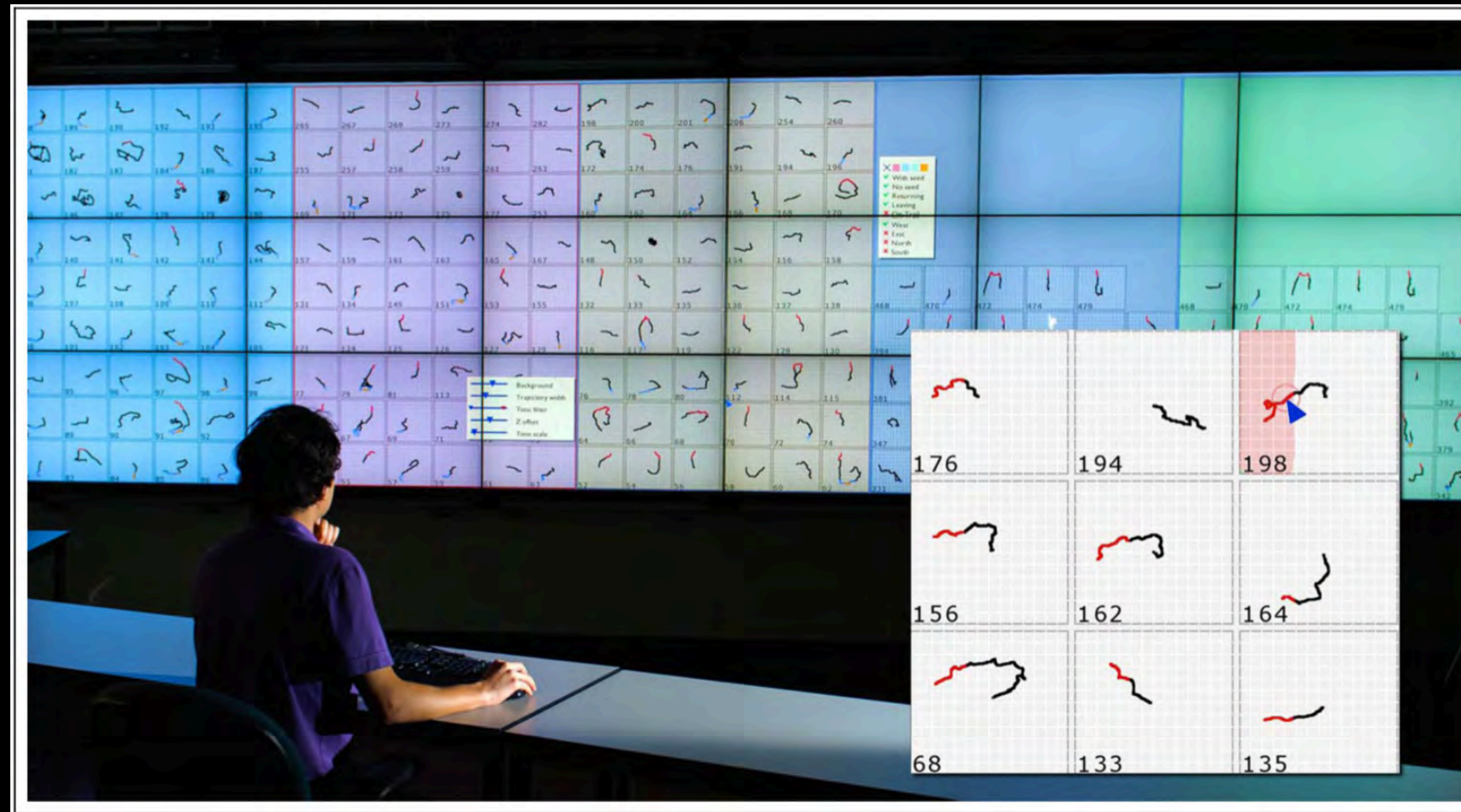
<sup>c</sup>K. Reda, A. Knoll, K. Nomura, M. E. Papka, A. E. Johnson, J. Leigh, *Visualizing Large-Scale Atomistic Simulations in Ultra-resolution Immersive Environments*, **Proceedings of the 2013 IEEE Symposium on Large Data Analysis and Visualization (LDAV 2013)**, pp. 59-66, Atlanta, GA, October 13-14, 2013.

# Virtual Reality<sup>d</sup>



<sup>d</sup>E. B. Brooks, J. A. Insley, M. E. Papka, S. Rizzi, *Virtual reality tools for the correction of automated volume segmentation errors using dense surface reconstructions*, **2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV)**, pp. 92-93, October 2, 2017. [POSTER]

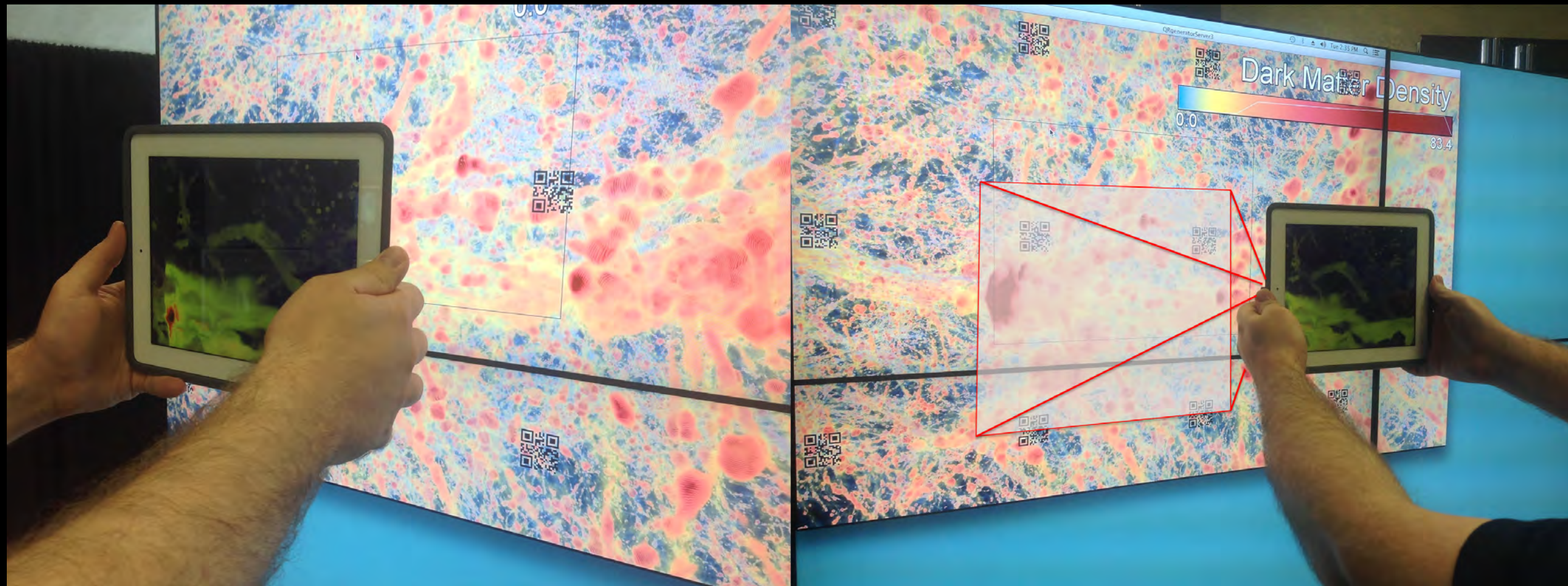
# Usability and Collaboration<sup>e</sup>



<sup>e</sup>K. Reda, A. E. Johnson, M. E. Papka, J. Leigh, *Modeling and Evaluating User Behavior in Exploratory Visual Analysis*, **Information Visualization** 15(4), pp. 325-339, October 2016.



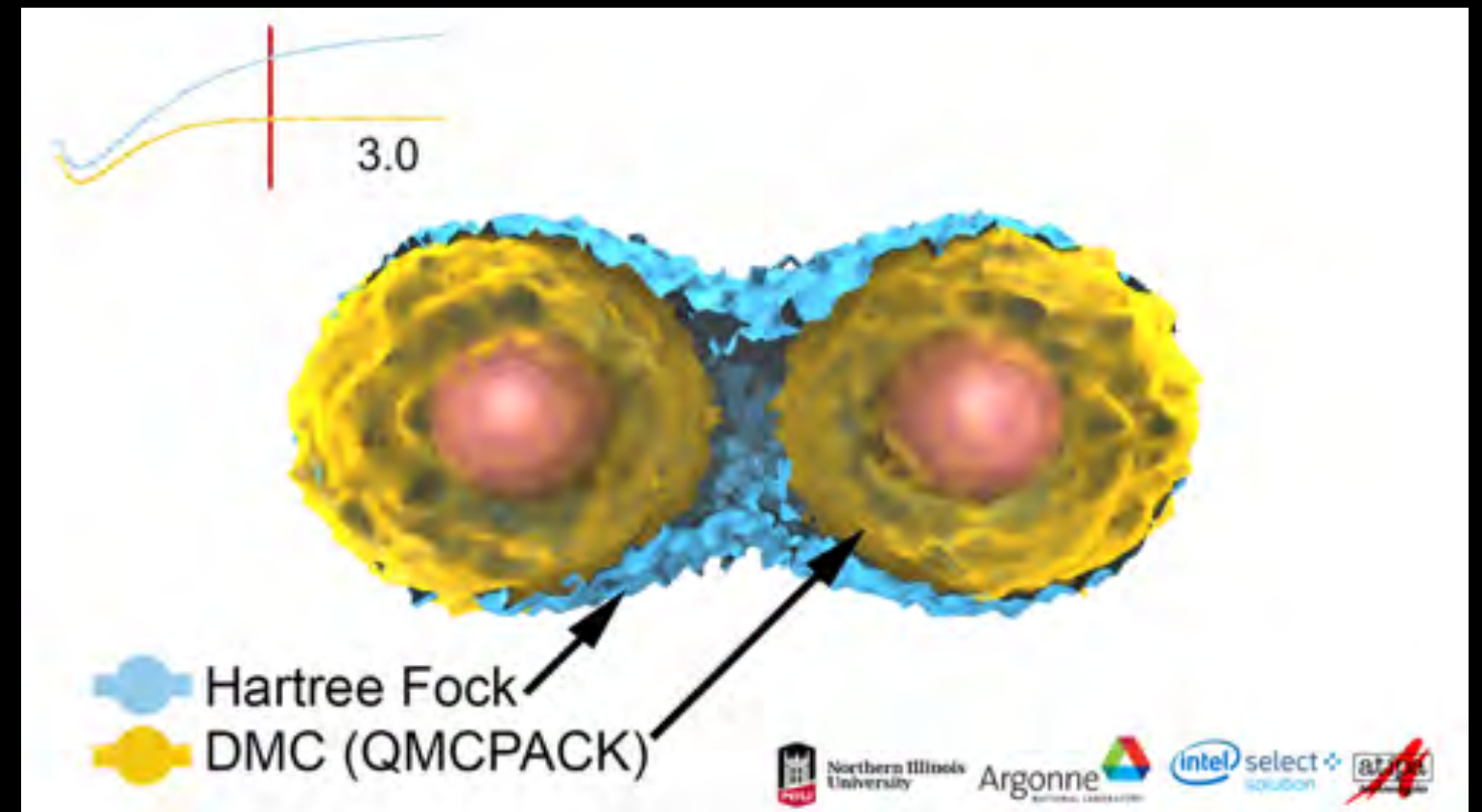
# Usability and Collaboration<sup>f</sup>



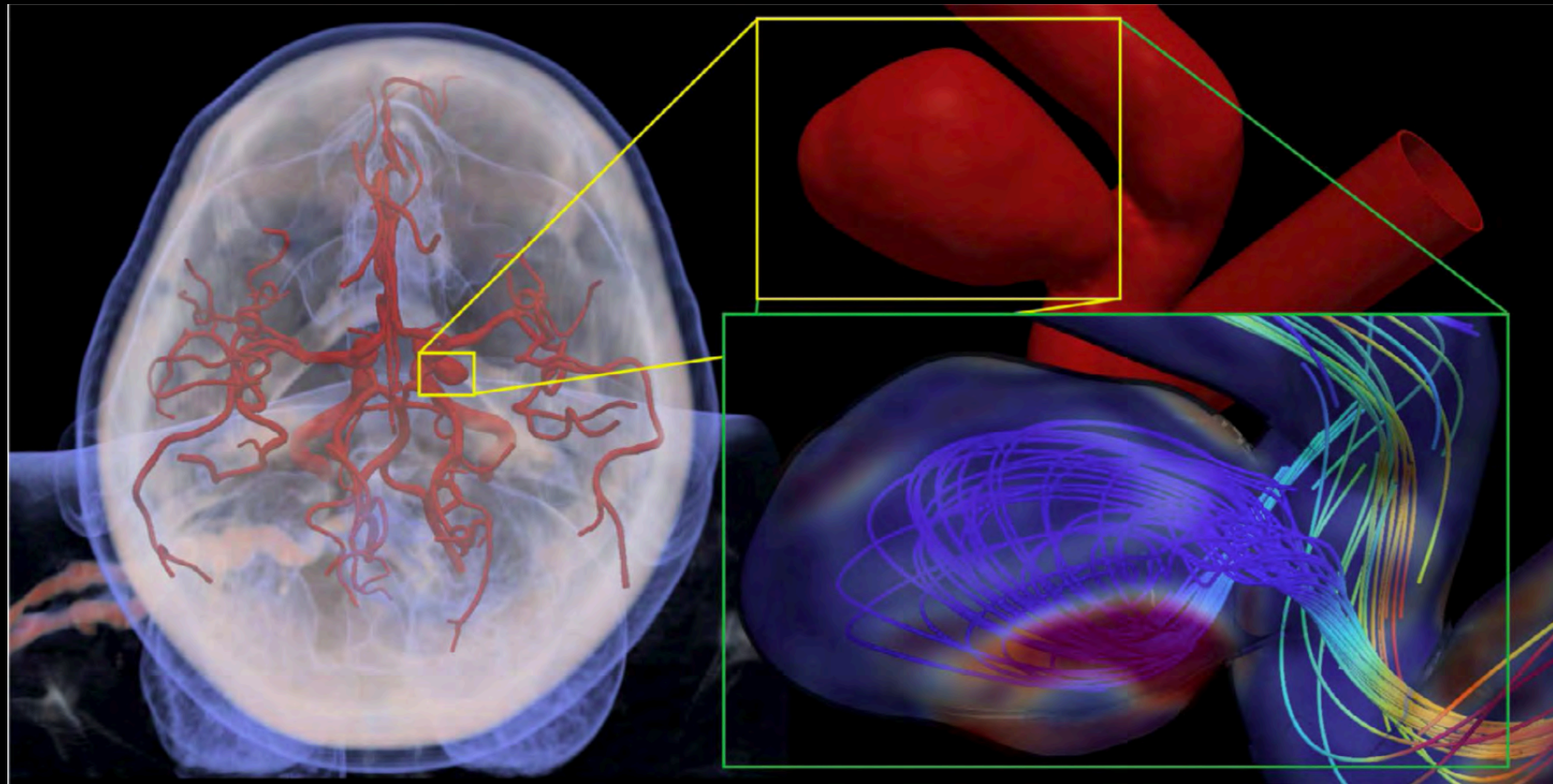
<sup>f</sup> **P. Lindner, A. Rodriguez**, T. Uram, M. E. Papka, *Augmenting Views on Large Format Displays with Tablets*, **Proceedings of the 2nd ACM Symposium on Spatial User Interaction (SUI 2014)**, Honolulu, HI, October 4-5, 2014. [Poster]

# Domain Specific Visualizations

- Applied solutions to specific problems within domain
- Deep partnership with domain experts
- Current effort with NIU Chemistry
  - *Visualizing and Quantifying Structural Ordering Underlying Static Structure Factor Peaks from Molecular Dynamics Simulations* Travis Mackoy, Bharat Kale, Ralph Wheeler
  - *Comparison Visualizations of Electron Density Approximation Methods* Anouar Benali (ANL), Joe Insley, Ralph Wheeler

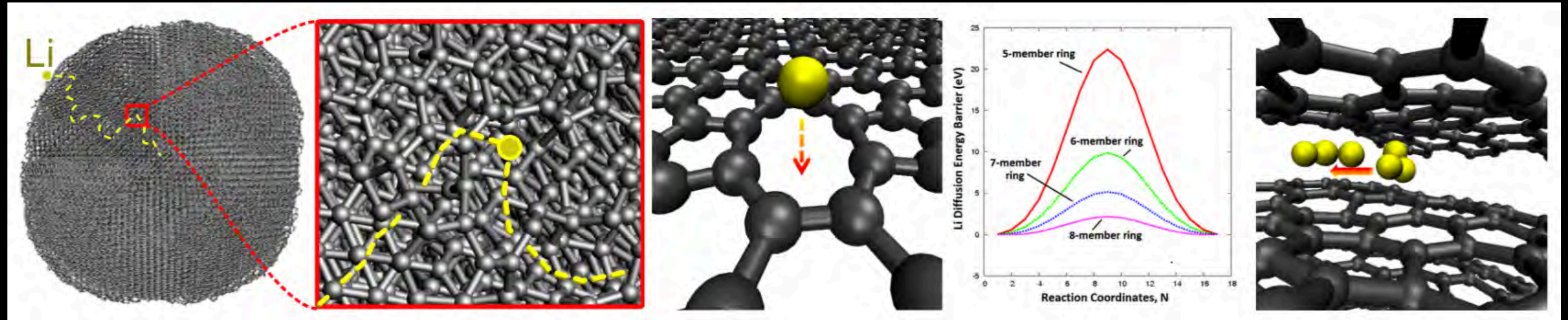


# Domain Specific Visualizations<sup>8</sup>



<sup>8</sup>P. Perdikaris, J.A. Insley, L. Grinberg, Y. Yu, M. E. Papka, G. E. Karniadakis, *Visualizing Multiphysics, Fluid-Structure Interaction Phenomena in Intracranial Aneurysms*, **Parallel Computing**, 55, pp. 9-16, July 2016.

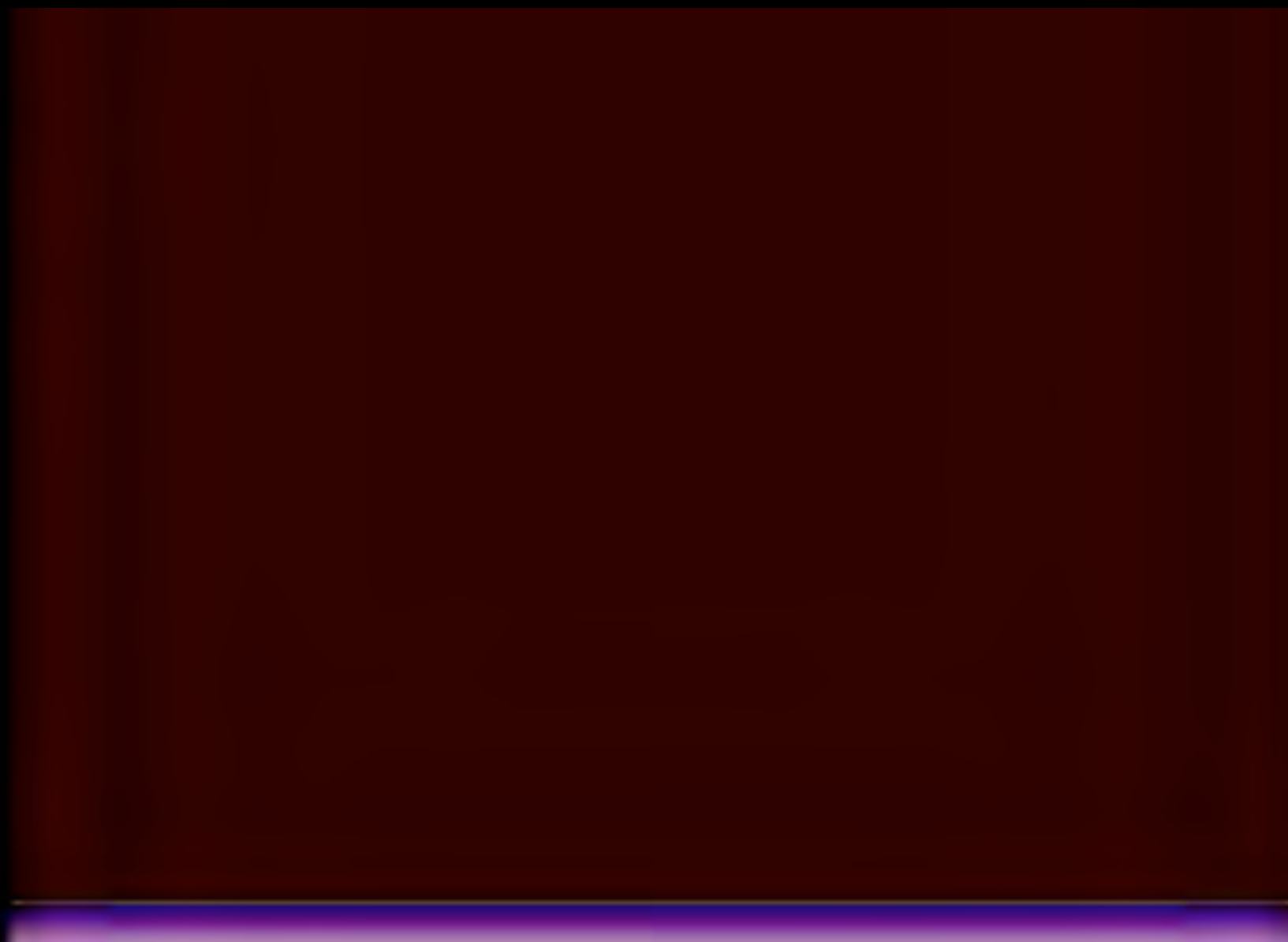
# Domain Specific Visualizations<sup>h</sup>



<sup>h</sup>A. Gyulassy, A. Knoll, K. C. Lau, B. Wang, P.-T. Bremer, M. E. Papka, L. Curtiss, V. Pascucci, *Interstitial and Interlayer Ion Diffusion Geometry Extraction in Graphitic Nanosphere Battery Materials*, **IEEE Transactions on Visualization and Computer Graphics**, 22(1):916-925, January 2016.

# Domain Specific Visualizations

Vertical Bar Chart



# Domain Specific Visualizations

# High Performance Computing

- Applications<sup>i</sup>
- Communication<sup>j</sup>
- Operations<sup>k</sup>

<sup>i</sup>R. Fisher, L. Kadanoff, D. Lamb, A. Dubey, T. Plewa, A. Calder, F. Cattaneo, P. Constantin, I. Foster, M. E. Papka, S. I. Abarzhi, S. M. Asida, P. M. Rich, C. C. Glendenin, K. Antypas, D. J. Sheeler, L. B. Reid B. Gallagher, and S. G. Needham, *Terascale Turbulence Computation Using the FLASH3 Application Framework on the IBM Blue Gene/L System*, **IBM Journal of Research and Development**, 52(1.2):127-36, 2008.

<sup>j</sup>V. Vishwanath, M. Hereld, V. Morozov, M. E. Papka, *Topology-Aware Data Movement and Staging for I/O Acceleration on Blue Gene/P Supercomputing Systems*, **SC'11 Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis**, Article No. 19, Seattle, WA, November 2011.

<sup>k</sup>S. Read, M. E. Papka, *Operational Metrics Reporting Processes at Scientific User Facilities: Comparing A High-Energy X-Ray Synchrotron Facility to a Supercomputing Facility*, **2017 IEEE International Professional Communication Conference (ProComm)**, pp. 1-6, Madison, WI, July 23, 2017.

# High Performance Computing

- Power<sup>l</sup>
- Scheduling<sup>m</sup>
- Workflows/Workloads<sup>n, o</sup>

<sup>l</sup>S. Wallace, Z. Zhou, V. Vishwanath, S. Coghlan, J. Tramm, Z. Lan, M. E. Papka, *Application Power Profiling on IBM Blue Gene/Q*, **Parallel Computing**, 57, pp. 73-86, September 2016.

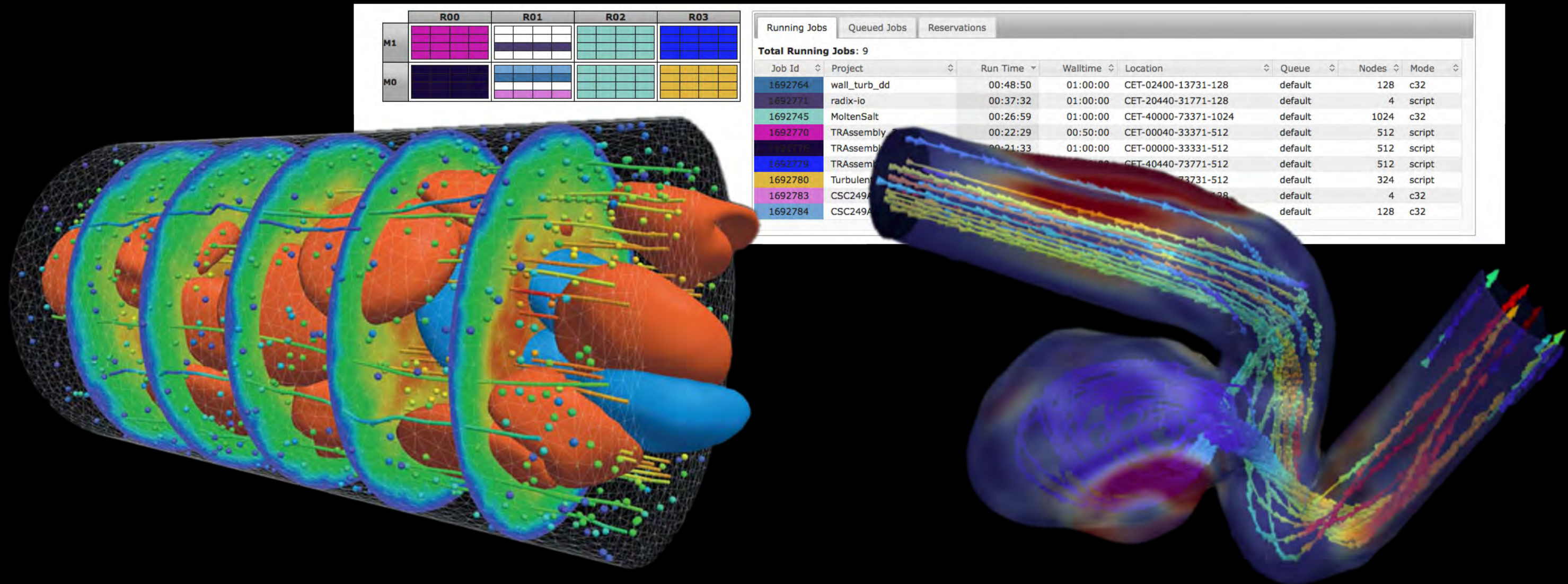
<sup>m</sup>Y. Fan, Z. Lan, P. Rich, W. E. Allcock, M. E. Papka, B. Austin, D. Paul, *Scheduling Beyond CPUs for HPC*, **Proceedings of the 28th International Symposium on High-Performance Parallel and Distributed Computing**, pp. 97-108, June 2019.

<sup>n</sup>W. E. Allcock, B. S. Allen, R. Ananthakrishnan, B. Blaiszik, K. Chard, R. Chard, I. Foster, L. Lacinski, M. E. Papka, R. Wagner, *Petrel: A Programmatically Accessible Research Data Service*, **Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines**, pp. 49, July 2019.

<sup>o</sup>M. A. Salim, T. D. Uram, J. T. Childers, P. Balaprakash, V. Vishwanath, M. E. Papka, *Balsam: Automated Scheduling and Execution of Dynamic, Data-Intensive HPC Workflows*, **arXiv preprint arXiv:1909.08704**, September 2019.

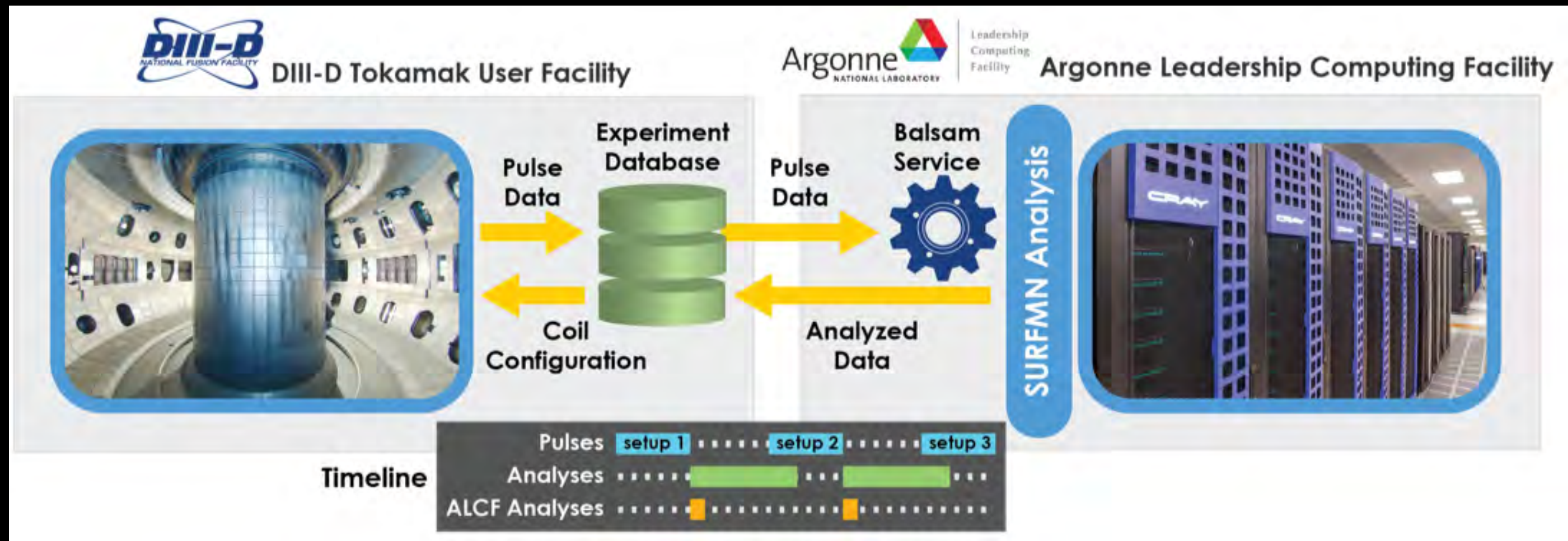


# Traditional<sup>P</sup>



<sup>P</sup>L. Grinberg, J. A. Insley, D. Fedosov, V. A. Morozov, M. E. Papka, G. E. Karniadakis, *Tightly Coupled Atomistic-Continuum Simulations of Brain Blood Flow on Petaflop Supercomputers*, **Computing in Science and Engineering**, 14(6):58-67, 2012.]

# Evolving (scheduling constraints)<sup>9</sup>

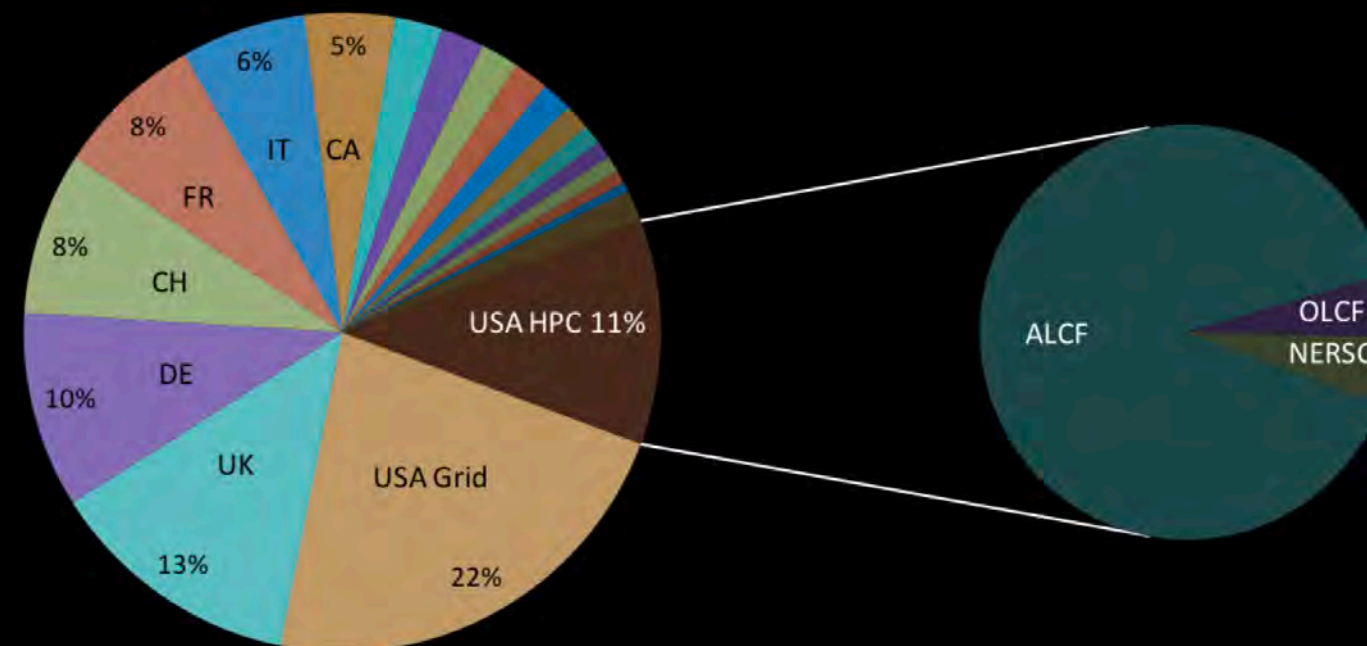


<sup>9</sup>M. Kostuk, T. D. Uram, T. Evans, D. M. Orlov, M. E. Papka, D. Schissel, *Automatic Between-Pulse Analysis of DIII-D Experimental Data Performed Remotely on a Supercomputer at Argonne Leadership Computing Facility*, **Fusion Science and Technology**, September 2017.]

# Evolving (complex workflows)<sup>r</sup>

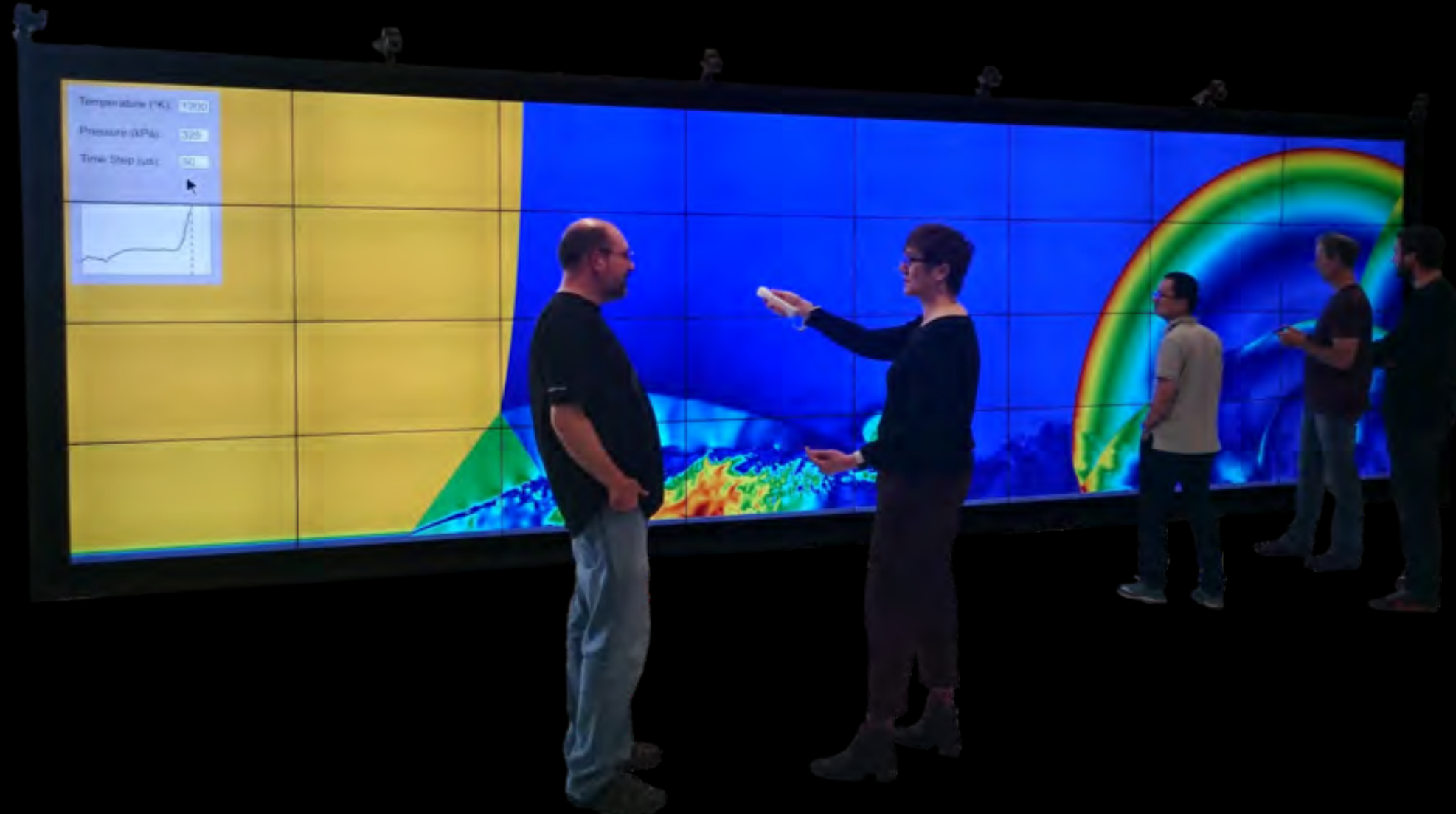
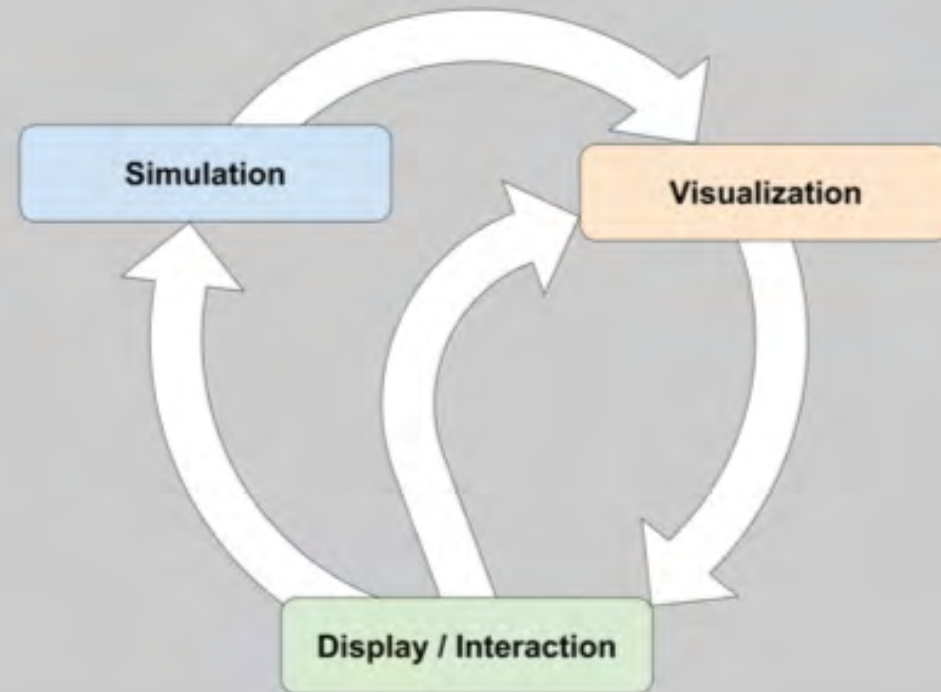
50% of the ATLAS papers based on 2015 data use the HPC-produced computing in a demonstrable manner

- These would still eventually be written without the US HPC effort, but they probably would not exist today: the **time-to-science has been dramatically shortened.**



[T. LeCompte(HEP){06/02/2016} and J. T. Childers, T. D. Uram, D. Benjamin, T. J. LeCompte, M. E. Papka, *An Edge Service for Managing HPC Workflows*, **Proceedings of the Fourth International Workshop on HPC User Support Tools (HUST'17)**, Denver, CO, November 12, 2017.]

# Evolving (increased engagement)<sup>5</sup>



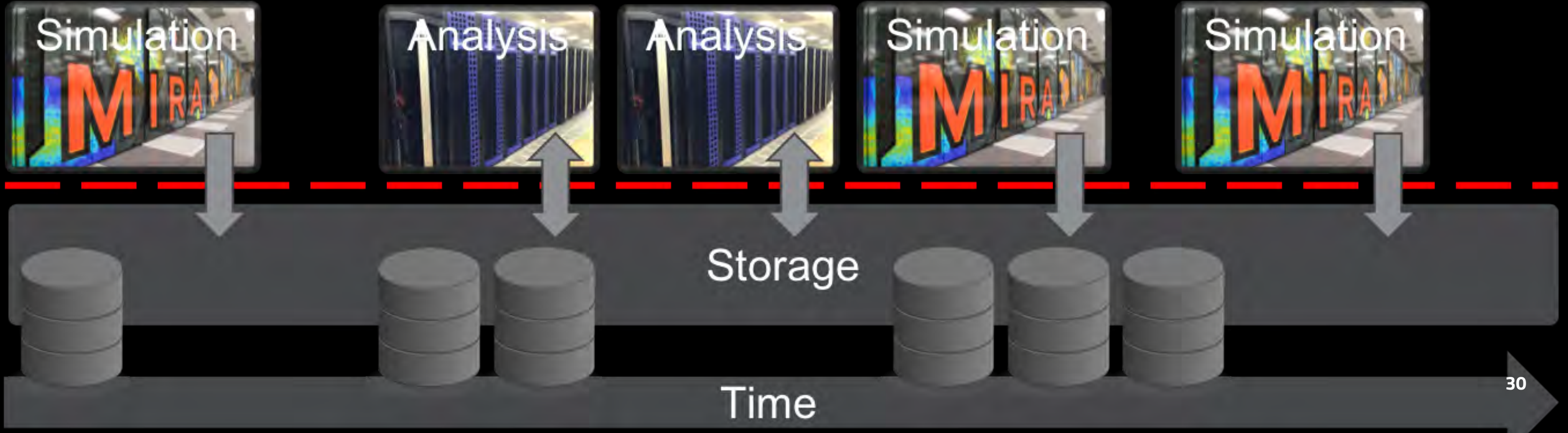
<sup>5</sup>T. Marrinan, A. Nishimoto, J. A. Insley, S. Rizzi, A. Johnson, M. E. Papka, *Interactive Multi-Modal Display Spaces for Visual Analysis*, **Proceedings of the 2016 ACM on Interactive Surfaces and Spaces**, pp. 421-426, Niagara Falls, Canada, November 6, 2016.]

# HPC Environments

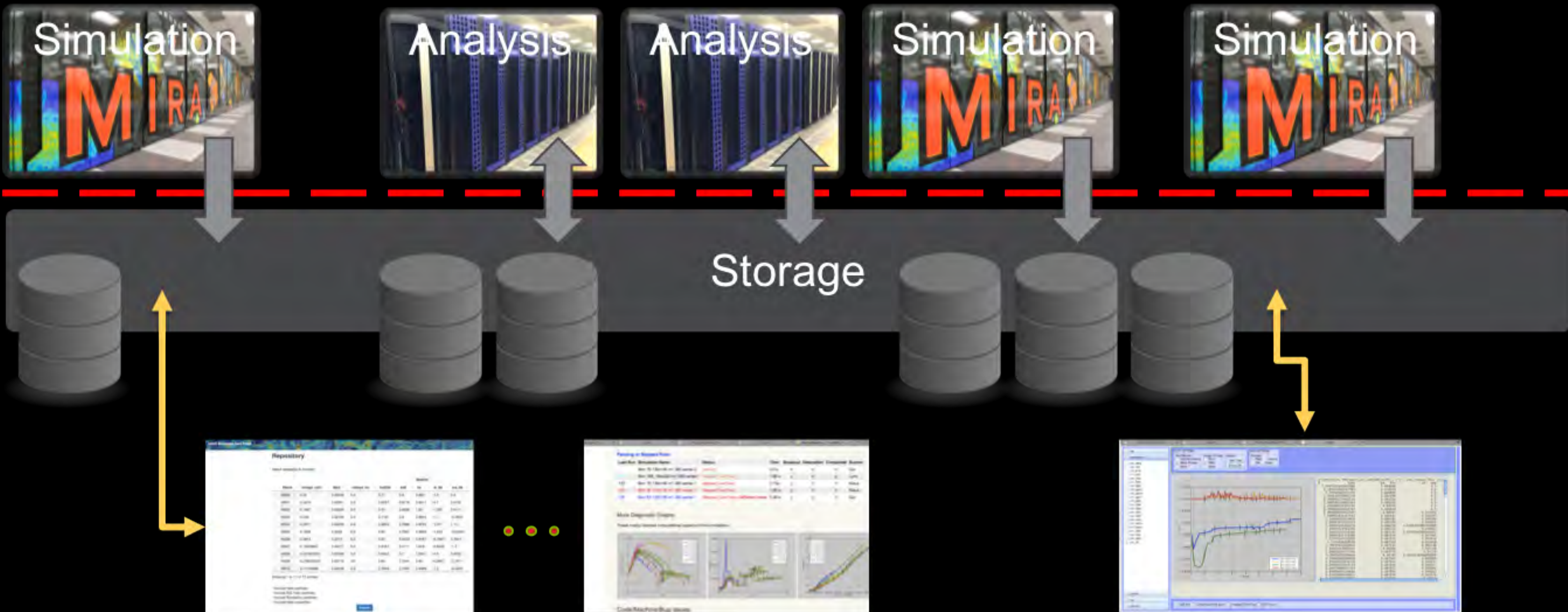
How do we **enable** scientists (users) to be the most **productive** from *start to finish*?

- How do we improve *usability*?
- How do we *enable* users of all levels?
- How do *simplify* supercomputing?

# Workflow of Today

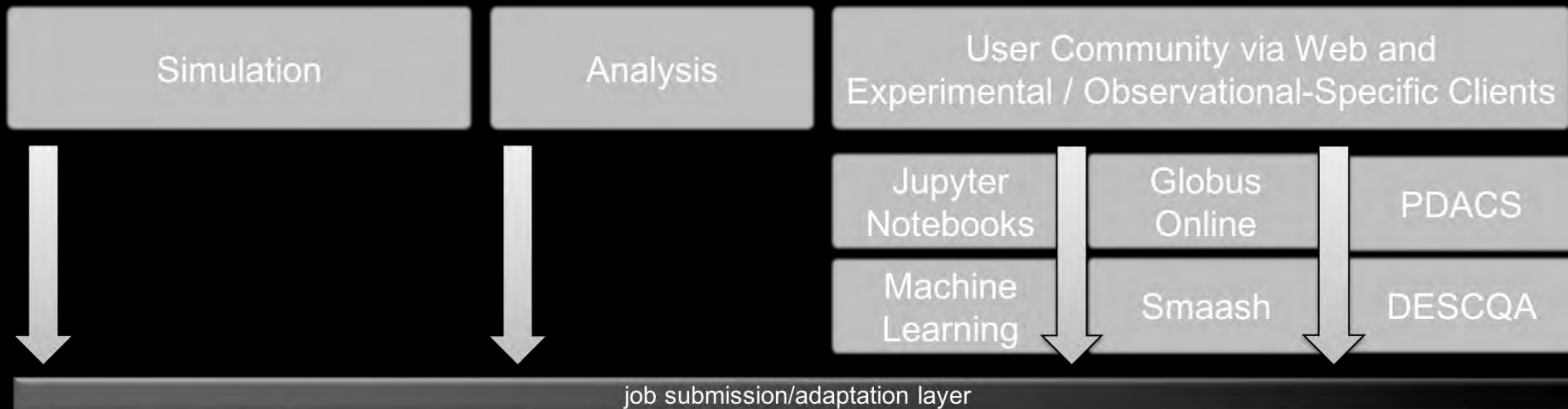


# Workflow of Tomorrow (Today)



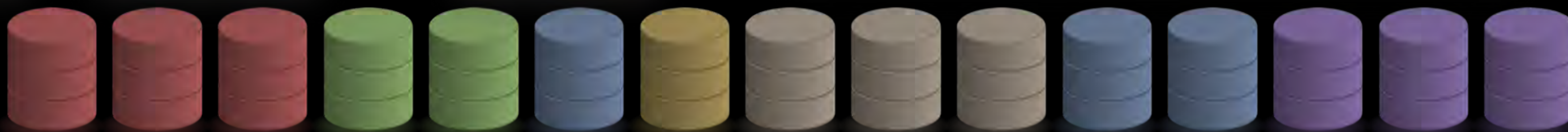
Time

# Facility of Tomorrow



**ALCF-hosted**  
Experimental / Observational-Specific Resources  
(servers like JupyterHub, databases, web, indexed search, visualization)

Facility Storage



Experimental / Observation Datasets

Cosmology

APS

SNS

LHC

Climate

Brain Imaging



# Observations *(Science Management)*

- Data-intensive science (simulations and experiments) requires **capture, curation** and **analysis**
- Data comes from many sources, in many formats and multiple sizes

# Observations (*Science Management*)

- Problem with science management:
  - Tracking simulations and output **[difficult]**
  - Finding and reproducing old simulations: **[difficult]**
  - Monitoring live simulations: **[inconvenient, idiosyncratic]**
  - Post-processing, analysis and archival of results: **[haphazard]**
  - Assessing simulation behavior/performance: **[difficult]**

# Increased Access to Scientific Communities

## Support for Application Teams

Simulation **m**anagement and **a**nalysis system for **Flash** (Smaash)<sup>†</sup>

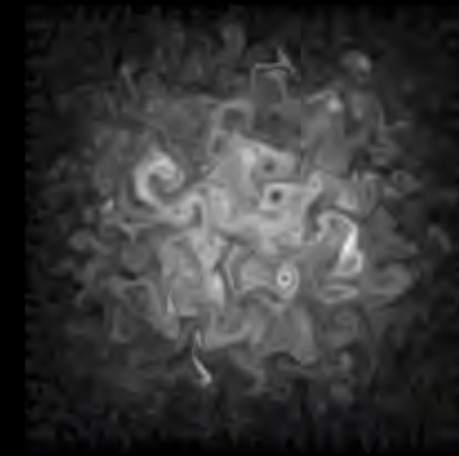
- Tracking and coordination of data (simulation and meta)
- Run-time monitoring of simulations and automated analysis of simulation output
- Method for managing / executing common workflows

---

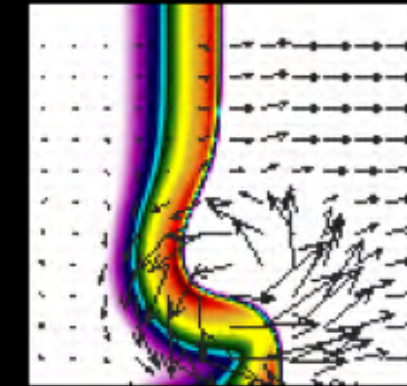
<sup>†</sup>R. Hudson, J. Norris, L. B. Reid, K. Weide, G. C. Jordan, and M. E. Papka, *Experiences Using Smaash to Manage Data-Intensive Simulations*, **Proceedings of the 20th International Symposium on High-Performance Parallel and Distributed Computing**, pp. 205-15, San Jose, CA, June 2011.

# Prototype Partner - Flash

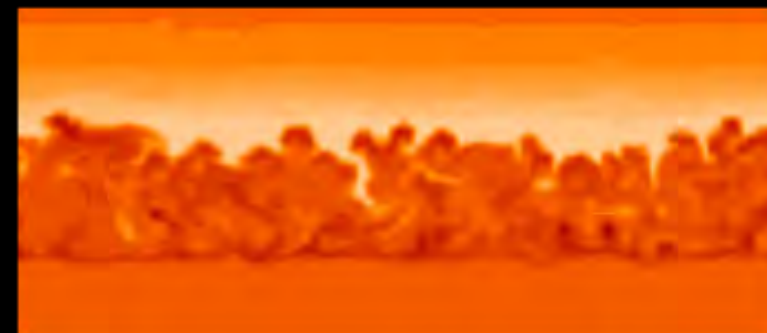
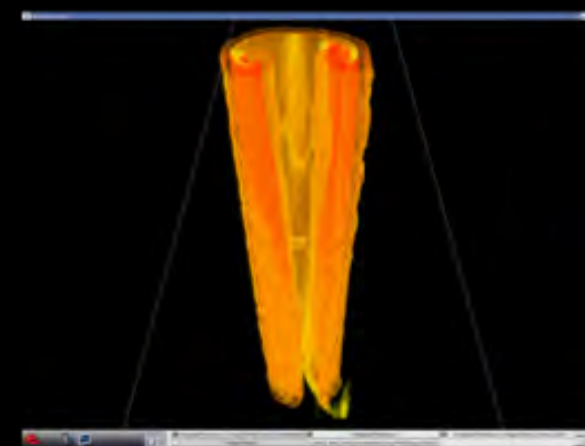
- Multi-physics
- Adaptive-mesh



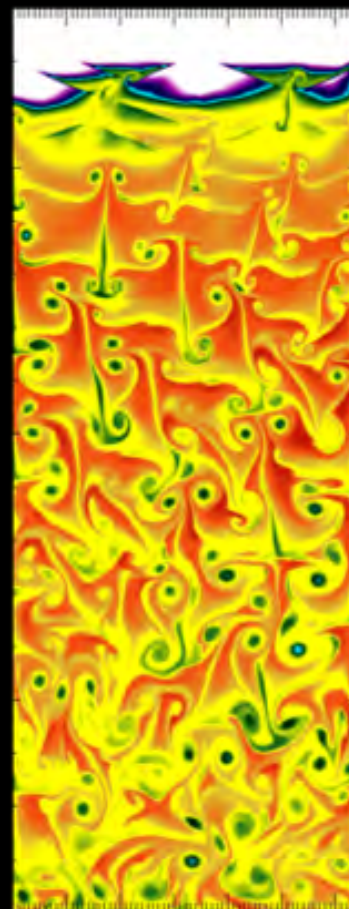
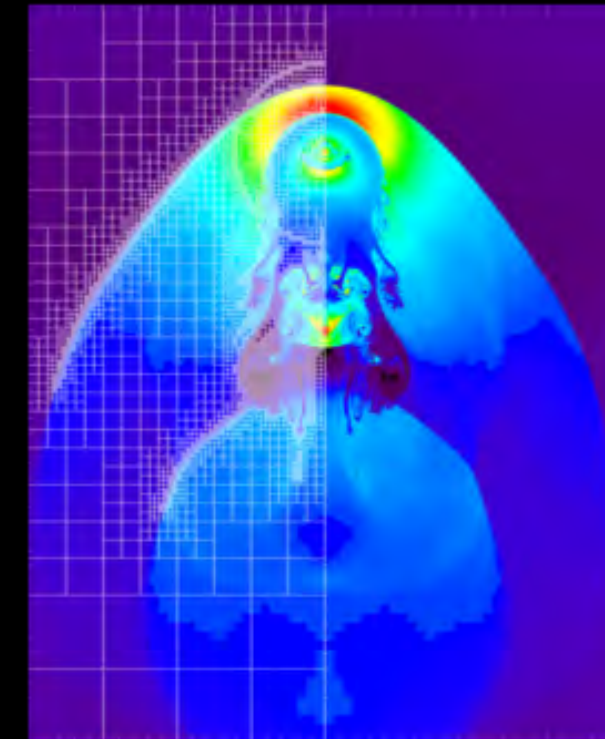
Compressible turbulence



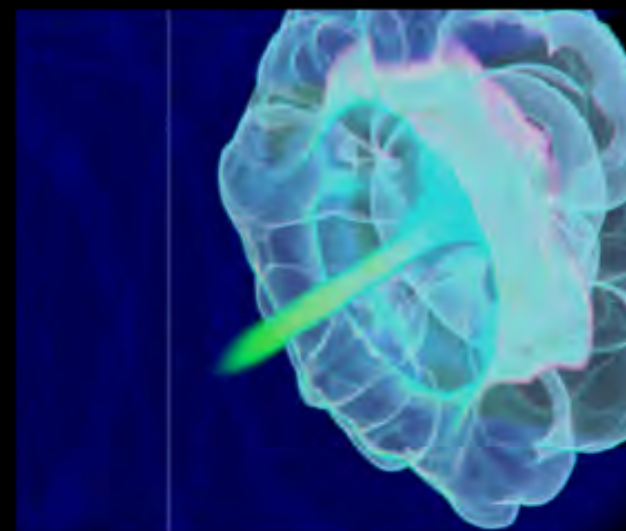
Flame-vortex interactions



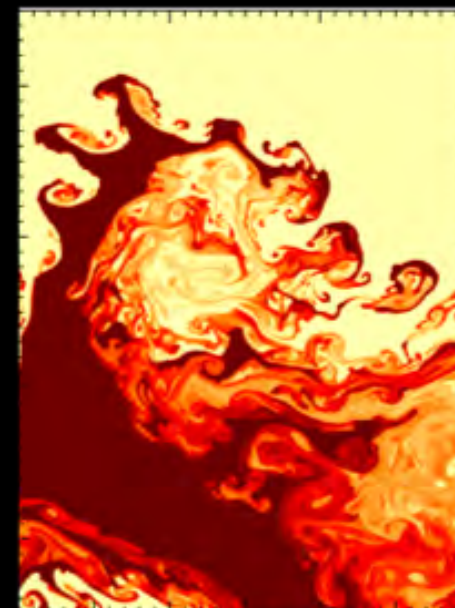
Nova outbursts on white dwarfs



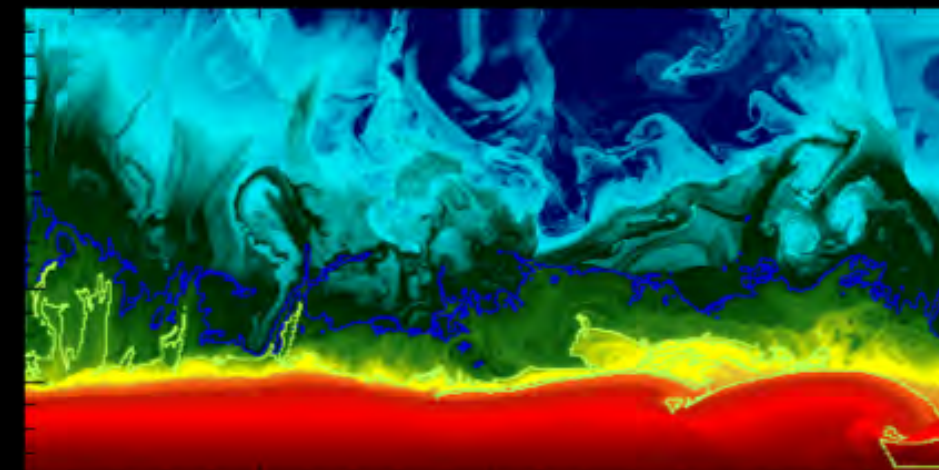
Cellular detonations



White



Rayleigh-Taylor instability



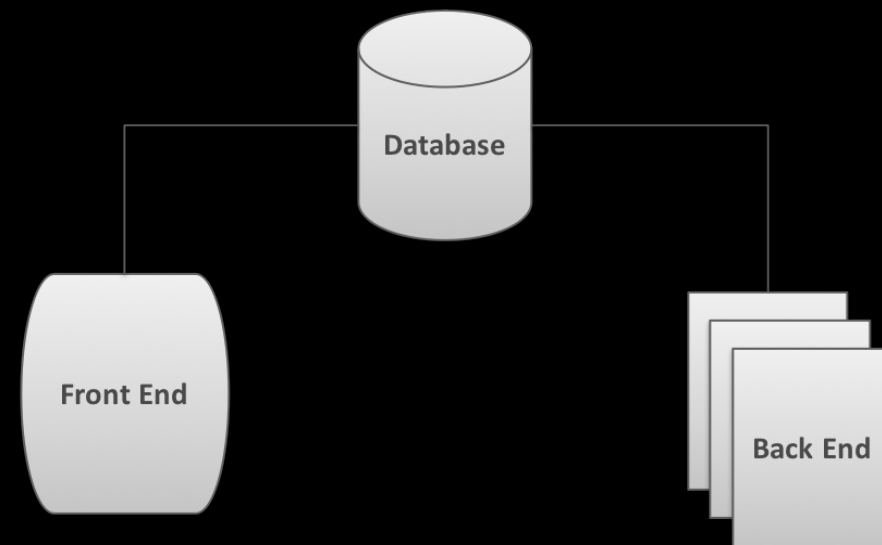
Helium burning on neutron stars

# Prototype Partner - Flash

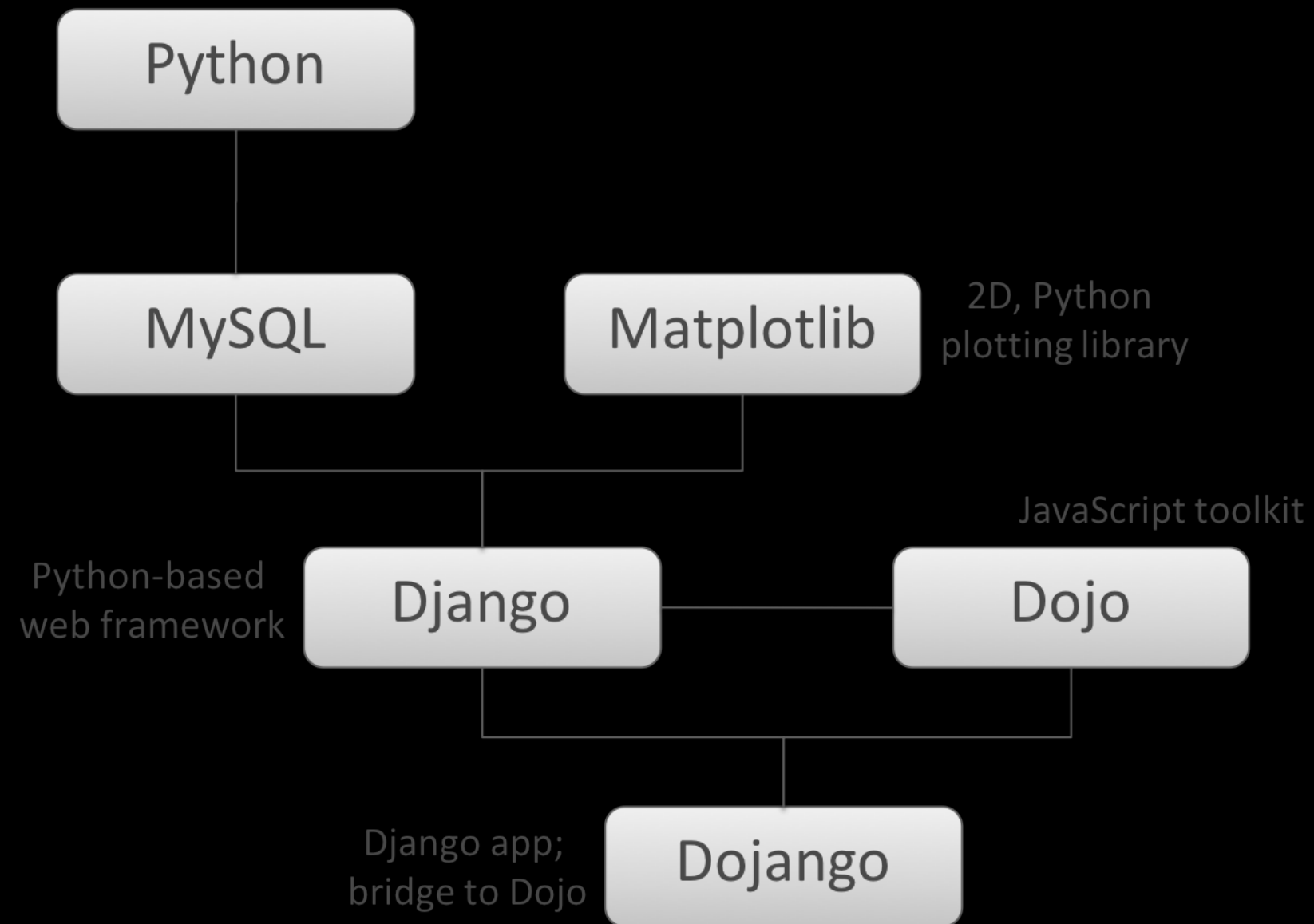
- Meta-data output
  - .log: simulation progress, warnings, errors, resource use
  - .dat: integrated grid quantities
- Scientific data output (HDF5)
  - Checkpoint: complete information needed to restart simulation
  - Plotfile: data values of interest for analysis
  - Particle files: tracer particles of interest during analysis

# Smaash Components

- Database (manages meta-data)
- Back end services (co-located with compute resources and scientific data)
- Front end interfaces (user facing)



# Smaash Implementation



# Smaash Back End Services

- Collector - captures and stores meta-data in database about simulation
- Archiver - automates the archiving of data
- Verifier - cross checks output and database entries
- Associator - connects a current simulation with campaign
- Observer - responsible for updates to user (email)
- Visualizer - automatic running of user specified visualization scripts



# Smaash Front End Interfaces (Views)

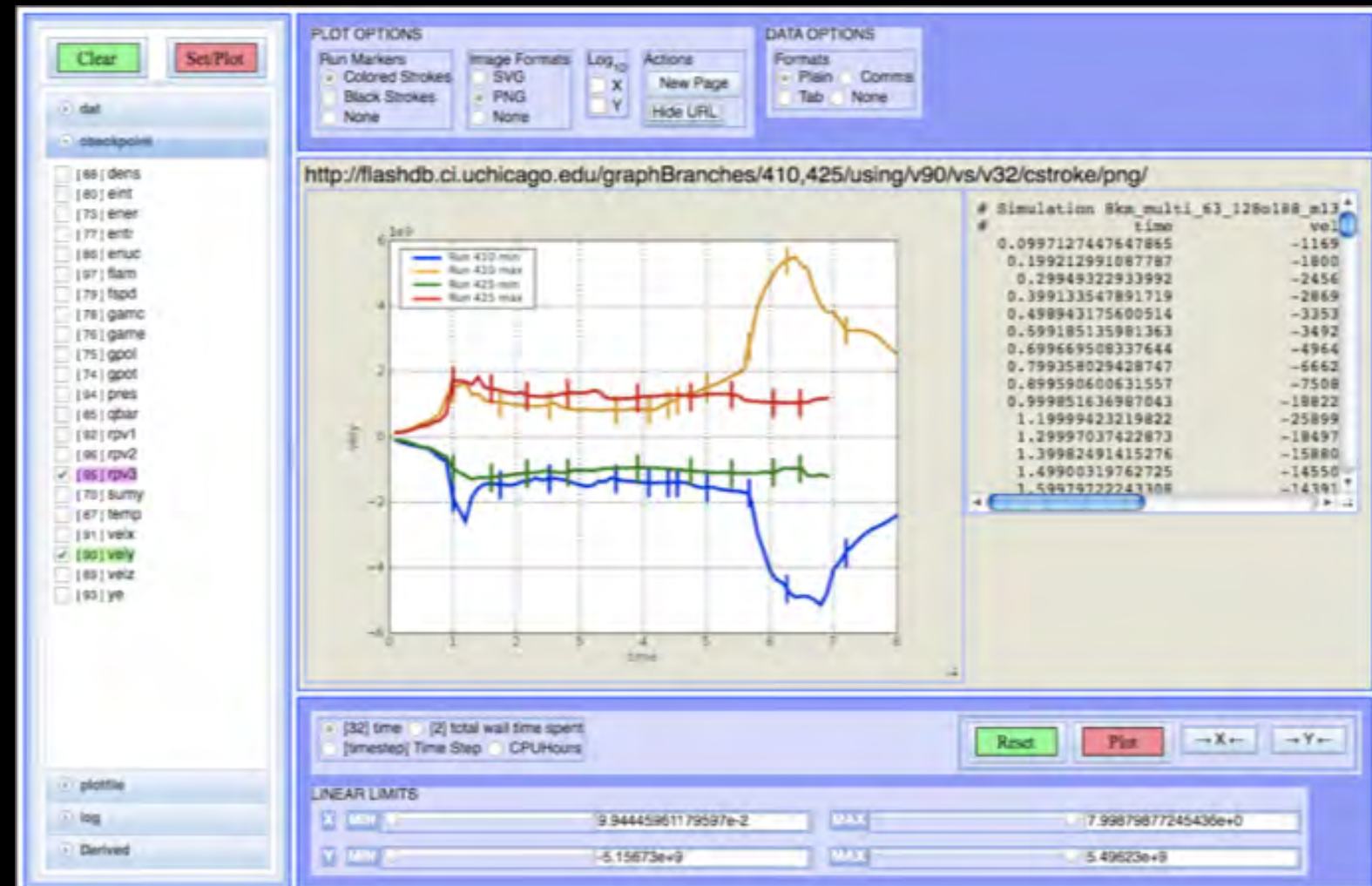
- Tree - collection of campaigns, simulations and runs
- Graph - quick graphs of results
- Monitor - automated visualizations
- Summary - details and notes

# Tree View

The screenshot displays a web interface for viewing simulation results. At the top, there are four filter panels: 'Filter by Date' (Before: 2/24/2011, After: 5/1/2010), 'Filter by Tag' (Flame Speed Study, FlameBubble, RTFlame, ResolutionStudy, WD\_def), 'Filter by Site' (ellipse.uchicago.edu, franklin.nersc.gov, intrepid.alcf.anl.gov), and 'Filter by Owner' (Cal Jordan, Carlo Graziani, Chad Glendenin, Chris Daley, Dean Townsley, Eva Wuyts). Below the filters is a 'Show Hidden' checkbox and radio buttons for 'All' and 'Any' tags. The main area is a table with columns: Name, Date, Tags, Description, Dim, and Graph. The table lists simulation results, including 'FlameSpeed [55]', 'flameBubble [54]', and a series of '1km', '2km', '4km', '8km', and '16km' resolution studies. Each row includes a tree view icon, a date, tags, a description, dimensions (16x16x16), and a graph icon.

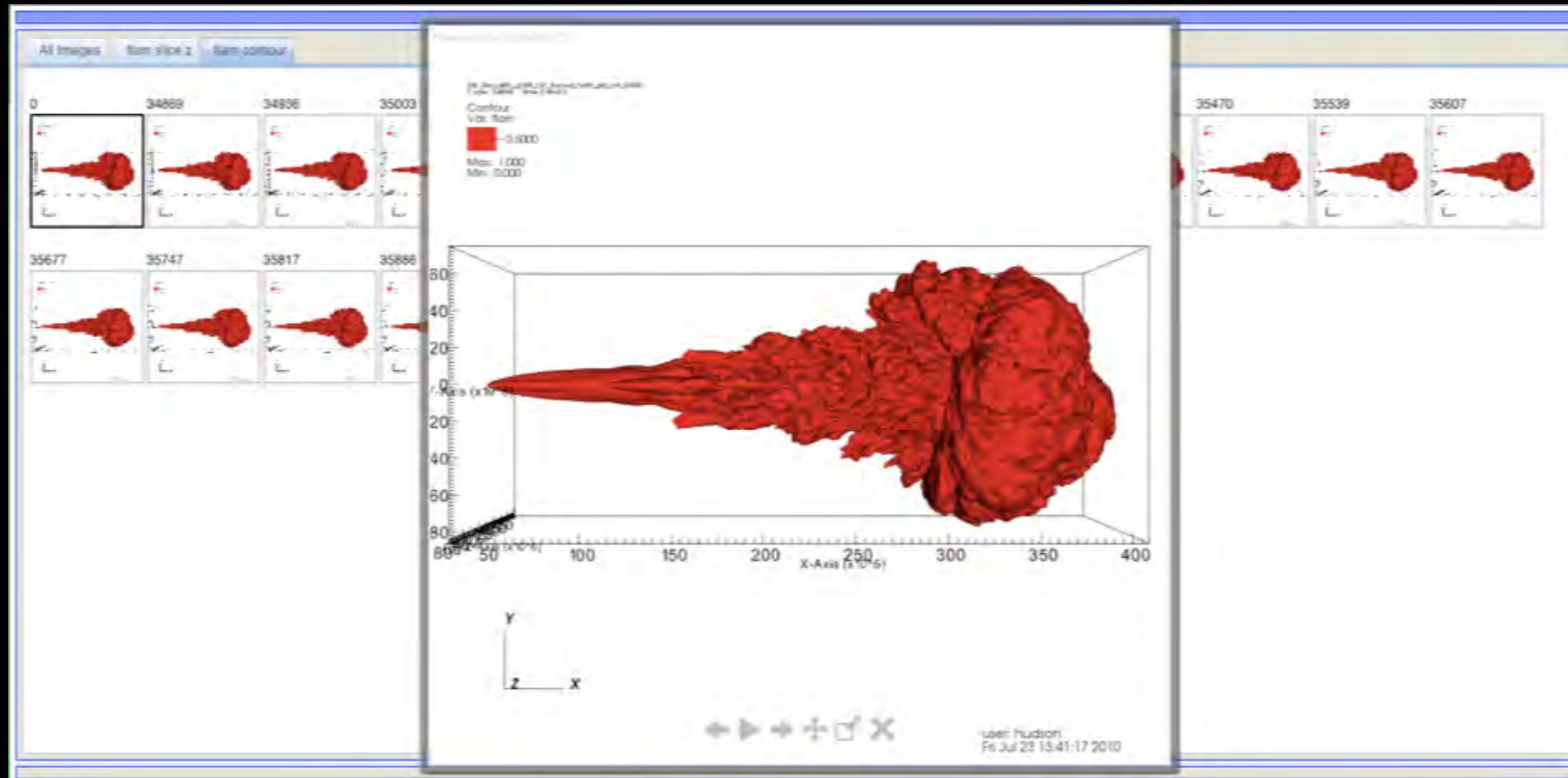
Name	Date	Tags	Description	Dim	Graph
> <a href="#">FlameSpeed [55]</a>	2010-11-16		/intrepid-fs0/users/jnorris/pe...	i 16x16x16	
> <a href="#">flameBubble [54]</a>	2010-06-12		/intrepid-fs0/users/hudson/per...	i 16x16x16	
> <a href="#">1km_sl85_q3E9_r32 [104]</a>	2010-06-15	FlameBubble	Flame bubble resolution study,...	i 16x16x16	
> <a href="#">rundir_0001 [683]</a>	2010-06-15	FlameBubble	/intrepid-fs0/users/jnorris/pe...	i 16x16x16	<input type="checkbox"/>
> <a href="#">rundir_0002 [685]</a>	2010-06-13	FlameBubble	/intrepid-fs0/users/jnorris/pe...	i 16x16x16	<input type="checkbox"/>
> <a href="#">rundir_0003 [688]</a>	2010-06-20	FlameBubble	/intrepid-fs0/users/jnorris/pe...	i 16x16x16	<input type="checkbox"/>
> <a href="#">rundir_0004 [689]</a>	2010-06-22	FlameBubble	/intrepid-fs0/users/jnorris/pe...	i 16x16x16	<input type="checkbox"/>
> <a href="#">2km_sl85_q3E9_r32 [101]</a>	2010-06-12	FlameBubble	/intrepid-fs0/users/hudson/per...	i 16x16x16	
> <a href="#">4km_sl85_q3E9_r32 [102]</a>	2010-06-12	FlameBubble	flame bubble simulation at _4 ...	i 16x16x16	
> <a href="#">8km_sl85_q3E9_r32 [103]</a>	2010-06-12	FlameBubble	Flame bubble resolution study,...	i 16x16x16	
> <a href="#">16km_sl85_q3E9_r32 [100]</a>	2010-06-12	FlameBubble	/intrepid-fs0/users/hudson/per...	i 16x16x16	

# Graph View



<http://flashdb.ci.uchicago.edu/graphBranches/410,425/using/v90/vs/v32/cstroke/png>

# Monitor View




# Summary View

**FlameBubble problem on 2048 processors**

Run completed

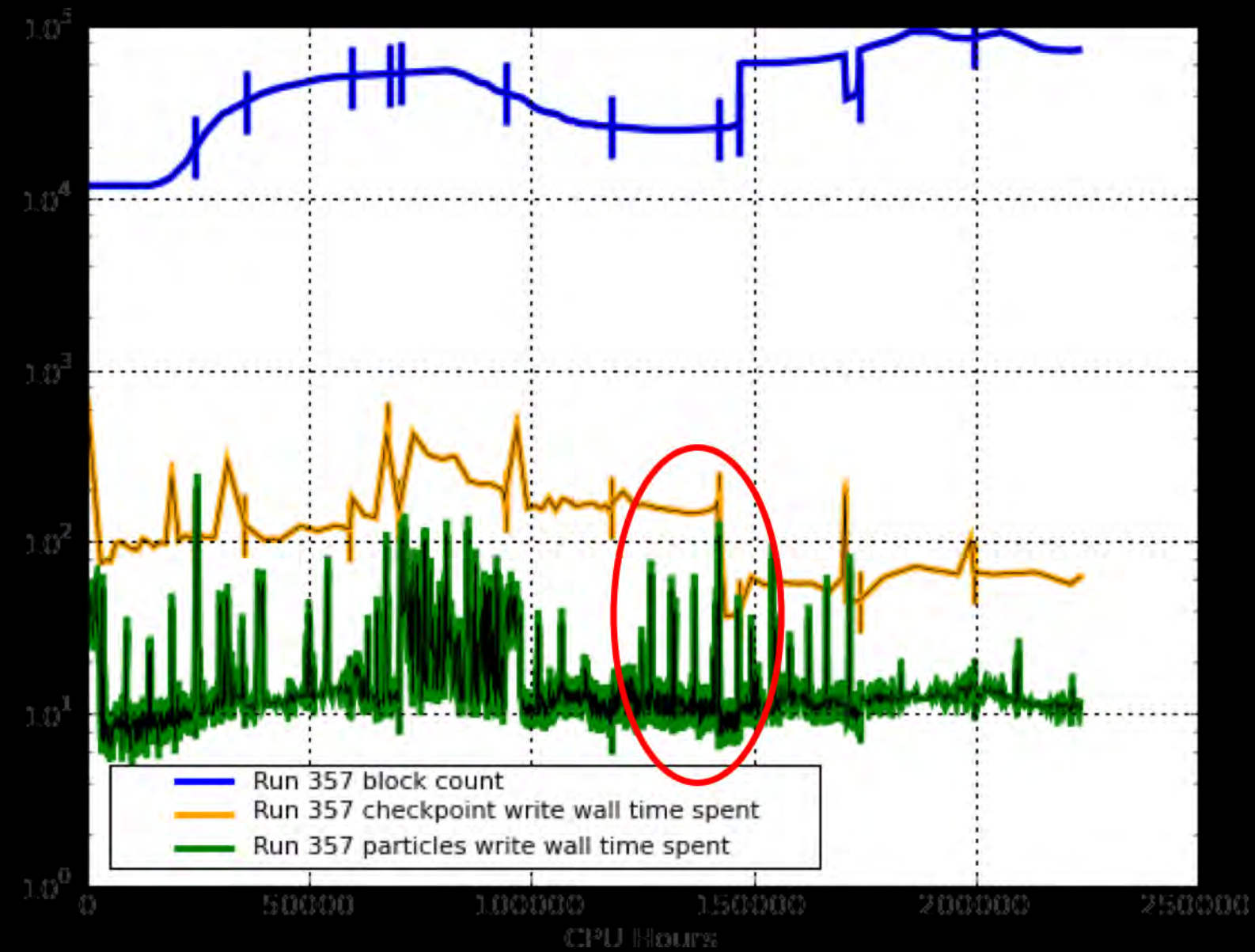
Parent: [rudir\\_0001 \[683\]](#)

Details Files Images 

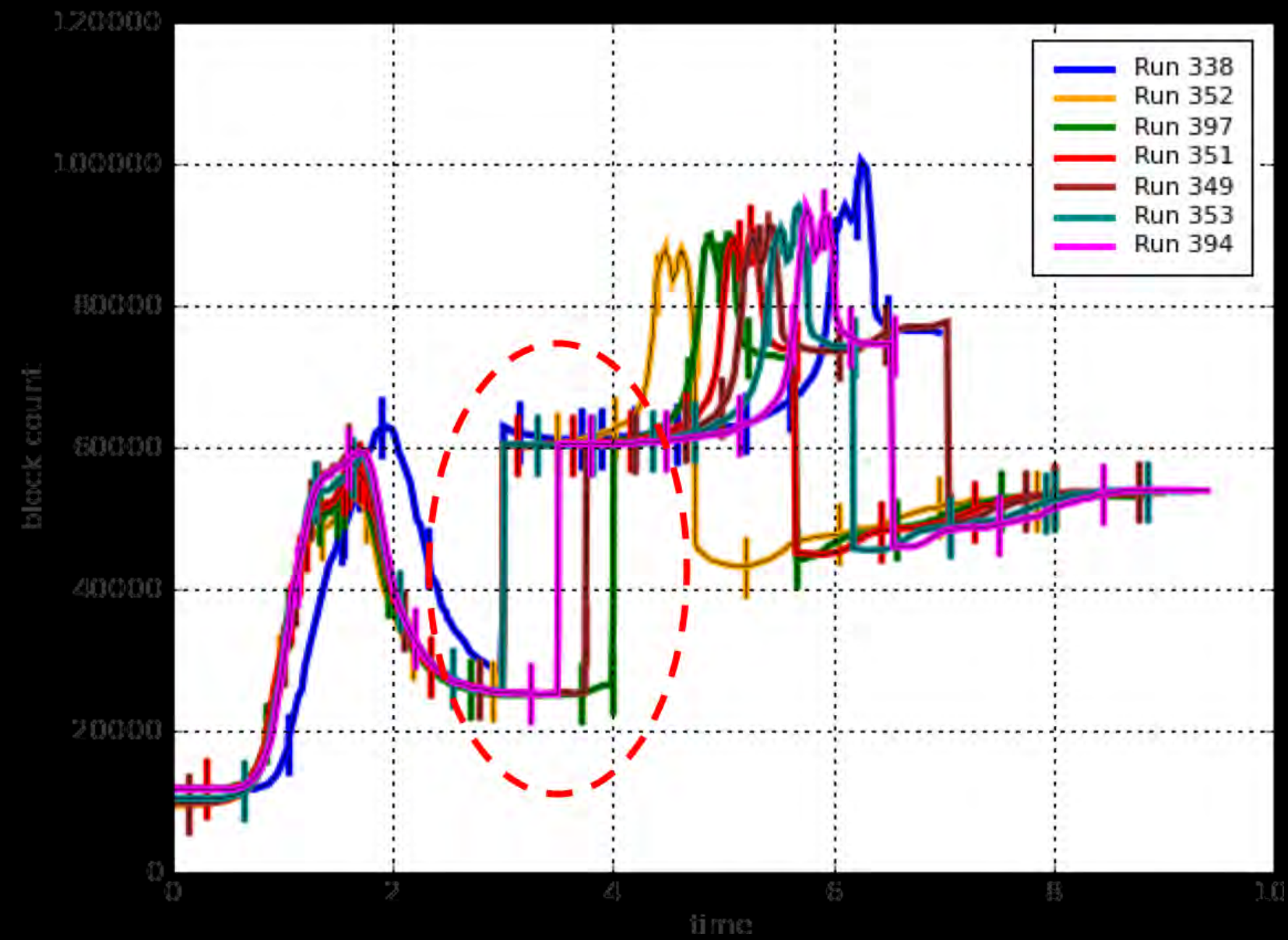
`/intrepid-fs0/users/jnorris/persistent/2010/ResolutionStudy/1km_sl`

<b>System Info</b>	Linux login5 2.6.16.60-0.42.8-ppc64 #1 SMP Tue Dec 15 17:28:00 UTC 2009 pp
<b>Setup Syntax</b>	<code>/intrepid-fs0/users/gjordan/persistent/2010/flameBubble/src/20100610/trunk/bin/s -maxblocks=40</code>
<b>FORTTRAN Compiler Flags</b>	<code>mpif90.ibm -g -O4 -qintsize=4 -qrealsize=8 -qfixed -qnosave -c -qsuffix=cpp=F -q -qsuffix=f=F90:cpp=F90 -qfree=f90 -WF,-DMAXBLOCKS=40 -WF,-DNXB=16 -Wf</code>
<b>C Compiler Flags</b>	<code>mpicc.ibm -I/include -I/soft/apps/hdf5-1.6.6/include -DNOUNDERSCORE -I/bgsy -qarch=450 -qtune=auto -qcache=auto -qmaxmem=16384 -D_FILE_OFFSET_B -DN_DIM=3 -DHAVE_MALLINFO</code>
<b>Max Number of Blocks/Proc</b>	40
<b>Max Number of Particles/Proc</b>	1000

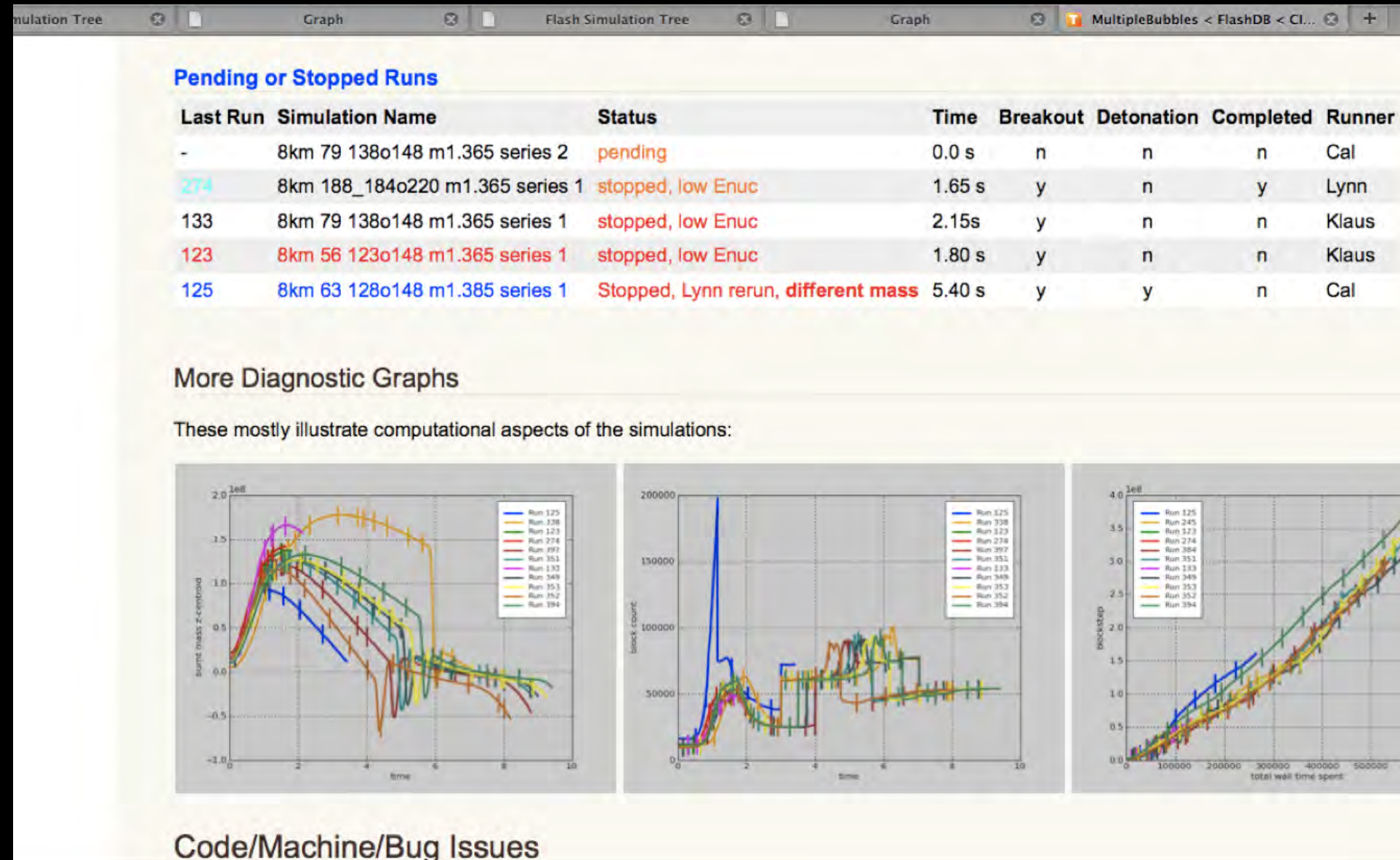
# Smaash Outcomes (Simulation State)



# Smaash Outcomes (Analysis)



# Smaash Outcome (Notebook)





# Smaash Today

Smaash/HACC

Model: MiraU / M019 / L2100

Runs

Run	Model	Simulation	Date	Size
run000	M019	MiraU	1 day, 16 hours ago	1
run001	M019	MiraU	1 day, 16 hours ago	1
run002	M019	MiraU	1 day, 16 hours ago	1

Files

Path	Date	Size	Type
/homes/turam/dev/smaash/data/631/pk.png	09/27/2017 8:21 a.m.	149367	png

Smaash/HACC

Run: MiraU / M019 / run001 / L2100

Timesteps

Timestep	Run	Model	Simulation	Date	Size
496	run001	M019	MiraU	1 day, 16 hours ago	1
495	run001	M019	MiraU	1 day, 16 hours ago	1
494	run001	M019	MiraU	1 day, 16 hours ago	1
493	run001	M019	MiraU	1 day, 16 hours ago	1
492	run001	M019	MiraU	1 day, 16 hours ago	1
491	run001	M019	MiraU	1 day, 16 hours ago	1
490	run001	M019	MiraU	1 day, 16 hours ago	1
489	run001	M019	MiraU	1 day, 16 hours ago	1
488	run001	M019	MiraU	1 day, 16 hours ago	1
487	run001	M019	MiraU	1 day, 16 hours ago	1

Files

Path	Date	Size	Type
/homes/turam/dev/smaash/data/606/hacc_gpu_m019.log	09/27/2017 8:21 a.m.	1.0 MB	txt
/homes/turam/dev/smaash/data/607/hacc_gpu_m019.err	09/27/2017 8:21 a.m.	119.6 KB	txt
/homes/turam/dev/smaash/data/606/indat.params	09/27/2017 8:21 a.m.	7.4 KB	txt

Smaash/HACC

Run: MiraU / M019 / run000 / L2100

Viewing /homes/turam/dev/smaash/data/589/indat.params

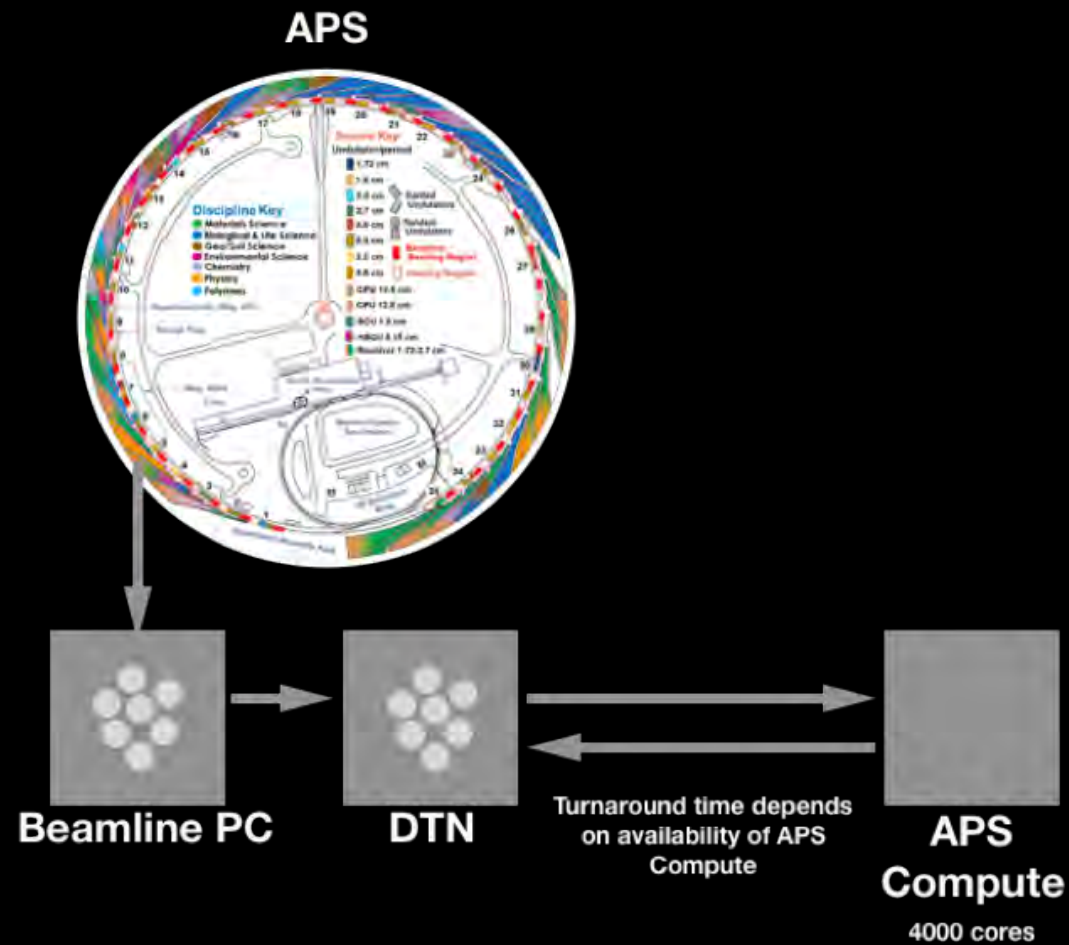
```
#####
# Header version information
#####
HACC_HEADER_VERSION 1.0.0

#####
# Cosmological Parameters
# Length scales are measured in Mpc/h
# OMEGA_CDM and OMEGA_NU given for Omega_cdm and Omega_nu (no $h^2$)
# DEUT=Omegab*h^2
# HUBBLE: Hubble constant/100 km/s/Mpc
# SS8: target value for sigma_8
# NS: index of the primordial power spectrum
# W_DE: constant dark energy equation of state
# Currently flat Universe only
#####
OMEGA_CDM 0.145084
DEUT 0.02217382692
OMEGA_NU 0.00686393
HUBBLE 0.825136069
SS8 0.854654384
NS 1.026482126
W_DE -0.981272302
WA_DE -0.339335368548398
T_CMB 2.726
N_EFF_MASSLESS 0.0
N_EFF_MASSIVE 3.04

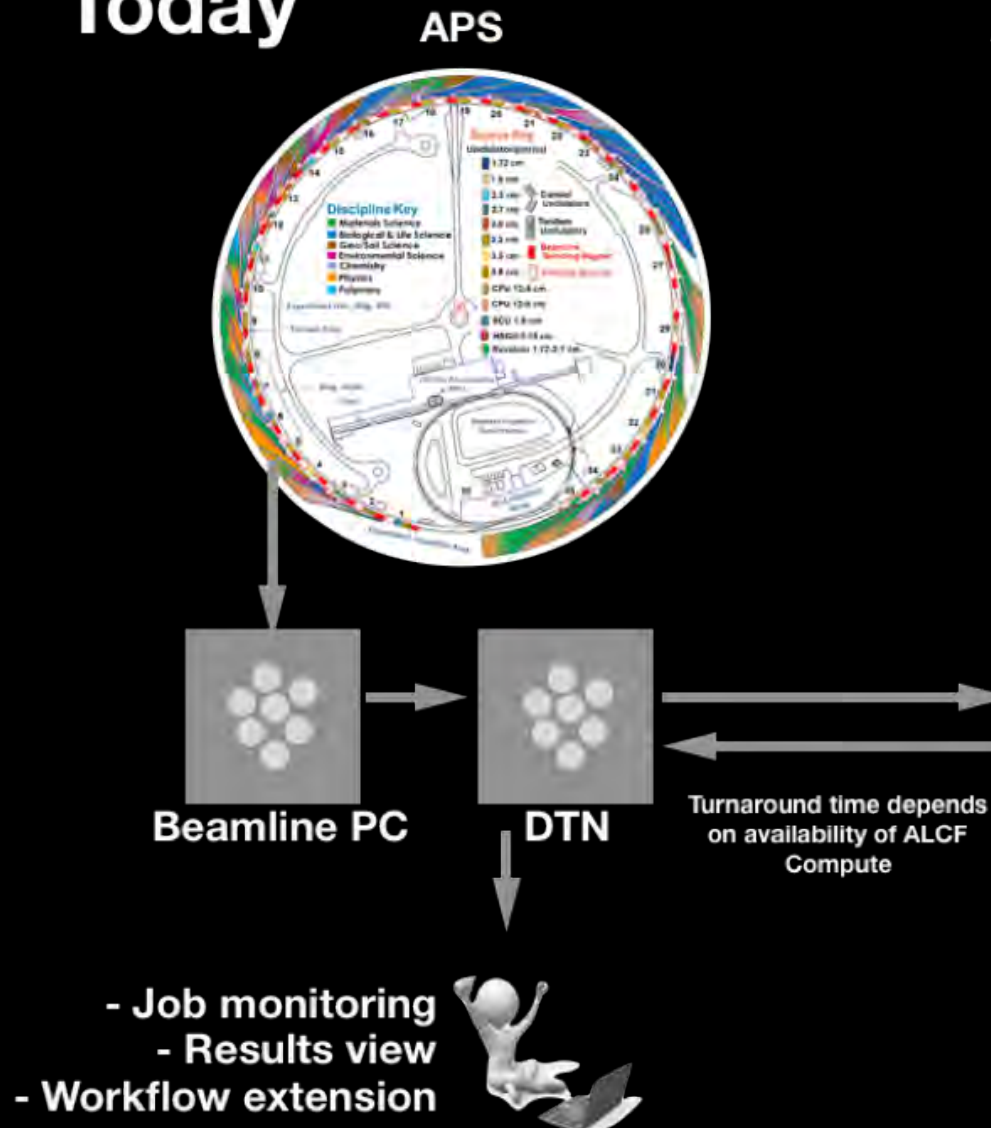
#####
# Initializer Set-up and read-ins
# ZIN: Starting redshift
# USE_WHITE_NOISE_INIT: YES: real space, NO: k space
# input type: INIT|RECORD|BLOCK|COSMO|RESTART
# INIT: generates internal initial conditions, rest if for read-ins
# distrib. type: ROUND_ROBIN|ALL_TO_ALL|ONE_TO_ONE|restart_step
# (ignored if INPUT_TYPE is INIT)
# ROUND_ROBIN indicates particles must be looked at by all processors
# ONE_TO_ONE indicates that particles physically reside on matching processor
# ALL_TO_ALL improved ROUND_ROBIN
# For restart: specify time step and modify INPUT_BASE_NAME
# TRANS: transfer function: Read in CAMB file (specify name in INPUT_BASE_NAME)
# or internal TF (KH, HS, PD, BBKS)
#####
Z_IN 200.0
USE_WHITE_NOISE_INIT YES
TRANS CMB
INPUT_BASE_NAME camb019.tf
INPUT_TYPE INIT
DISTRIBUTE_TYPE LAST
MAX_MINUTES 700
```

# New Efforts in Science

## Yesterday



## Today

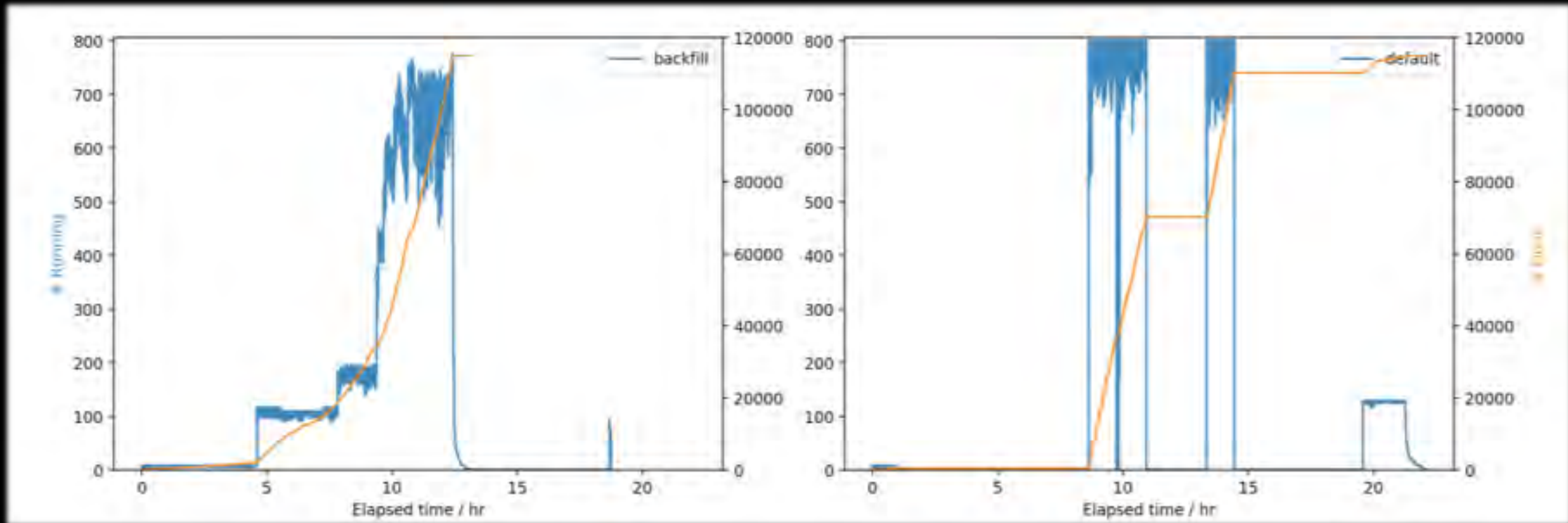


## ALCF Theta (11.7 PetaFLOPs)



- Balsam establishes a deep job reservoir, and flows jobs to ALCF machines based on availability
- Individual APS application runs to individual ALCF jobs
  - Many small APS jobs flow into a single large/long ALCF job
  - Balsam provisions new jobs as needed to satisfy workload
  - Balsam leverages backfill where possible; otherwise, optimizes jobs for scheduler

# New Efforts in Science<sup>u</sup>



<sup>u</sup>M. A. Salim, T. D. Uram, J. T. Childers, P. Balaprakash, V. Vishwanath and M. E. Papka, *Balsam: Automated Scheduling and Execution of Dynamic, Data-Intensive HPC Workflows*, to appear **1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing**, Denver, CO, November 2019.

# Last Topic

## Information Visualization

- Connection to **X** science



# Thank You

Most of my current funding is provided by the Argonne Leadership Computing Facility a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357 with additional support from the National Science Foundation.

Thanks to all the students of the ddiLab and my colleagues at NIU and ANL.



**If I have seen further it is by  
standing on the shoulders of  
giants.**

**— Sir Isaac Newton**

# Extra Slides

# HPC Landscape (Yesterday)

## Simulation Applications

---

64bit floating point

---

memory bandwidth

---

random access to memory

---

sparse matrices

---

distributed memory jobs

---

synchronous input/output multinode

---

scalability limited communication

---

low latency high bandwidth

---

large coherency domains (sometimes)

---

output typically greater than input

---

output rarely read

---

output is data



# HPC Landscape (Today)

## Simulation Applications

64bit floating point

memory bandwidth

random access to memory

sparse matrices

distributed memory jobs

synchronous input/output multinode

scalability limited communication

low latency high bandwidth

large coherency domains (sometimes)

output typically greater than input

output rarely read

output is data

## Big Data Applications

64bit and integer important

data analysis pipelines

databases including NoSQL

MapReduce/SPARK

millions of jobs

input/output bandwidth limited

data management limited

many task parallelism

large-data in and large-data out

input and output both important

output is read and used

output is data

## Deep Learning Applications

lower precision  $\leq 32$ bit

inferencing can be 8bit (TPU)

scaled integer possible

training dominates development

inference dominates pro

reuse of training data

data pipelines needed

dense float point typical SGEMM small DFT, CNN

ensembles and search

single models small

input more important than output

output is models