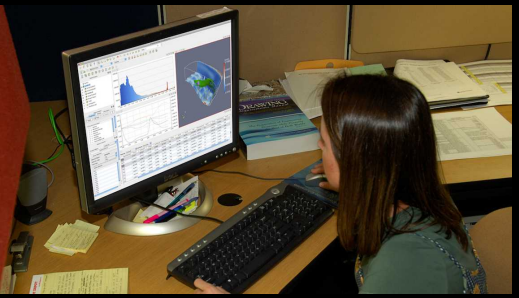
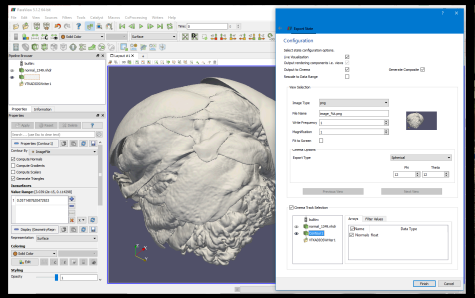
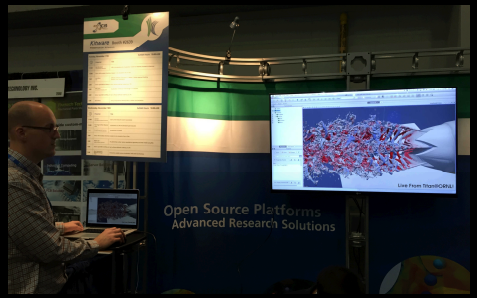
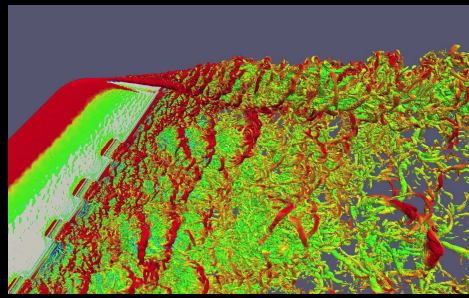
SAND2017-11564PE



Why You *Don't* Want to do In Situ Visualization, and Why You Have To

Computational Science Seminar Series

Kenneth Moreland Sandia National Laboratories

October 24, 2017

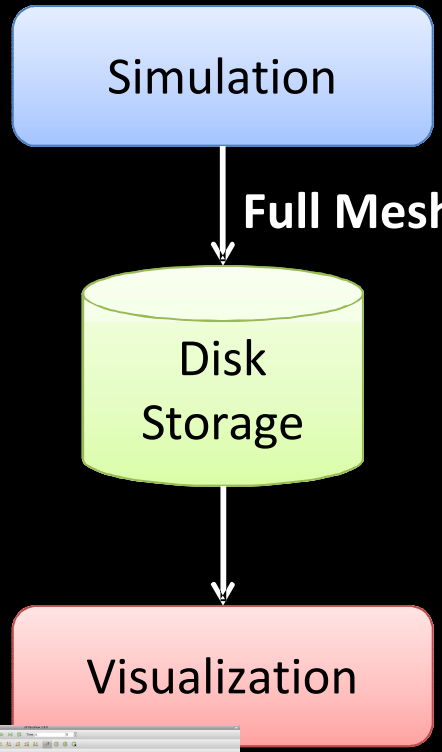
CCR Center for Computing Research

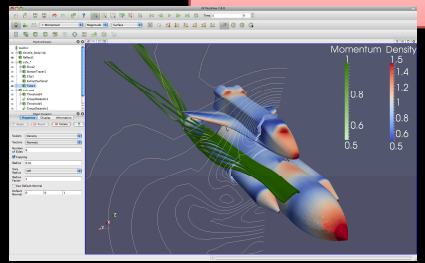
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND 2017-5707 C

Acknowledgements

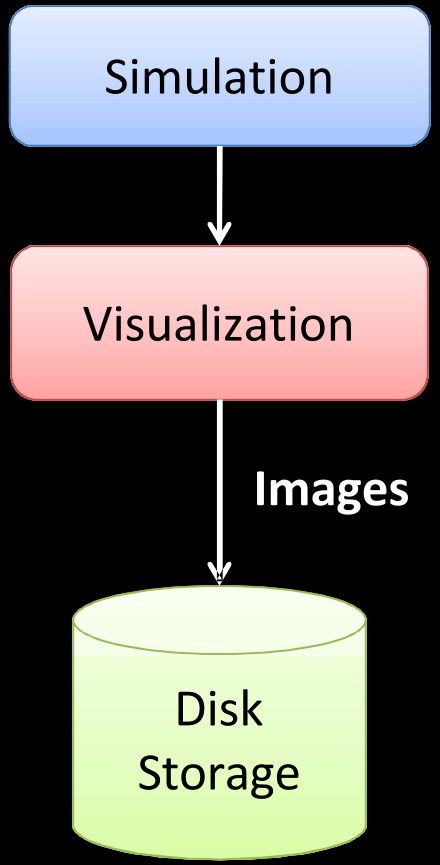
- This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Numbers 10-014707, 12-015215, and 14-017566.
- This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, in support of the nation's exascale computing imperative.
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.
- **Thanks to many, many partners in labs, universities, and industry.**

Post Hoc Vis





In Situ Vis



What Does In Situ Visualization Mean?

Sandia National Laboratories

Co-processing

"Strict" In Situ

In Transit

Simulation Steering

Integrated Problem Solving Environments

On-Line Analysis

Tightly/Loosely Couples

In Situ Visualization is visualization that *necessarily* starts before the data producer finishes.

In Situ Visualization is visualization that *necessarily* starts before the data producer finishes.

Post Hoc Visualization is visualization that can start at some arbitrary time after the data are produced.

In Situ is an Old Idea

E. E. Zajac, “Computer-Made Perspective Movies as a Scientific and Communication Tool”, Communications of the ACM, March **1964**

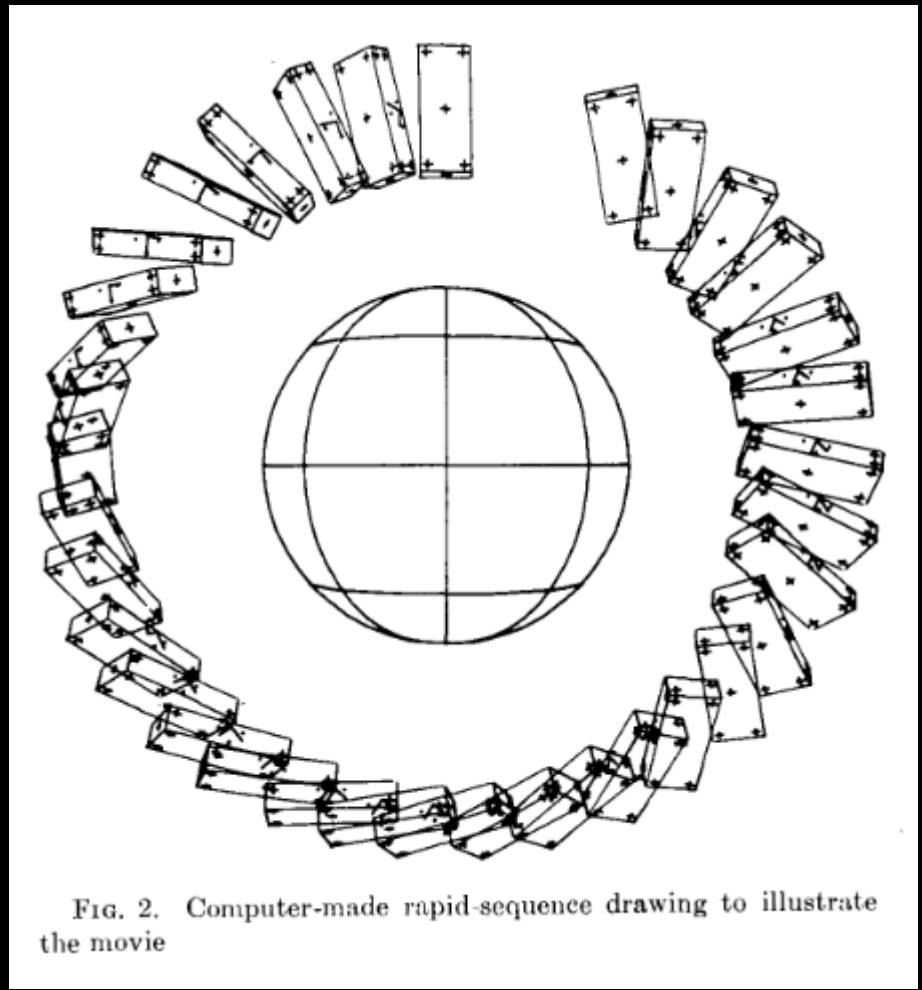
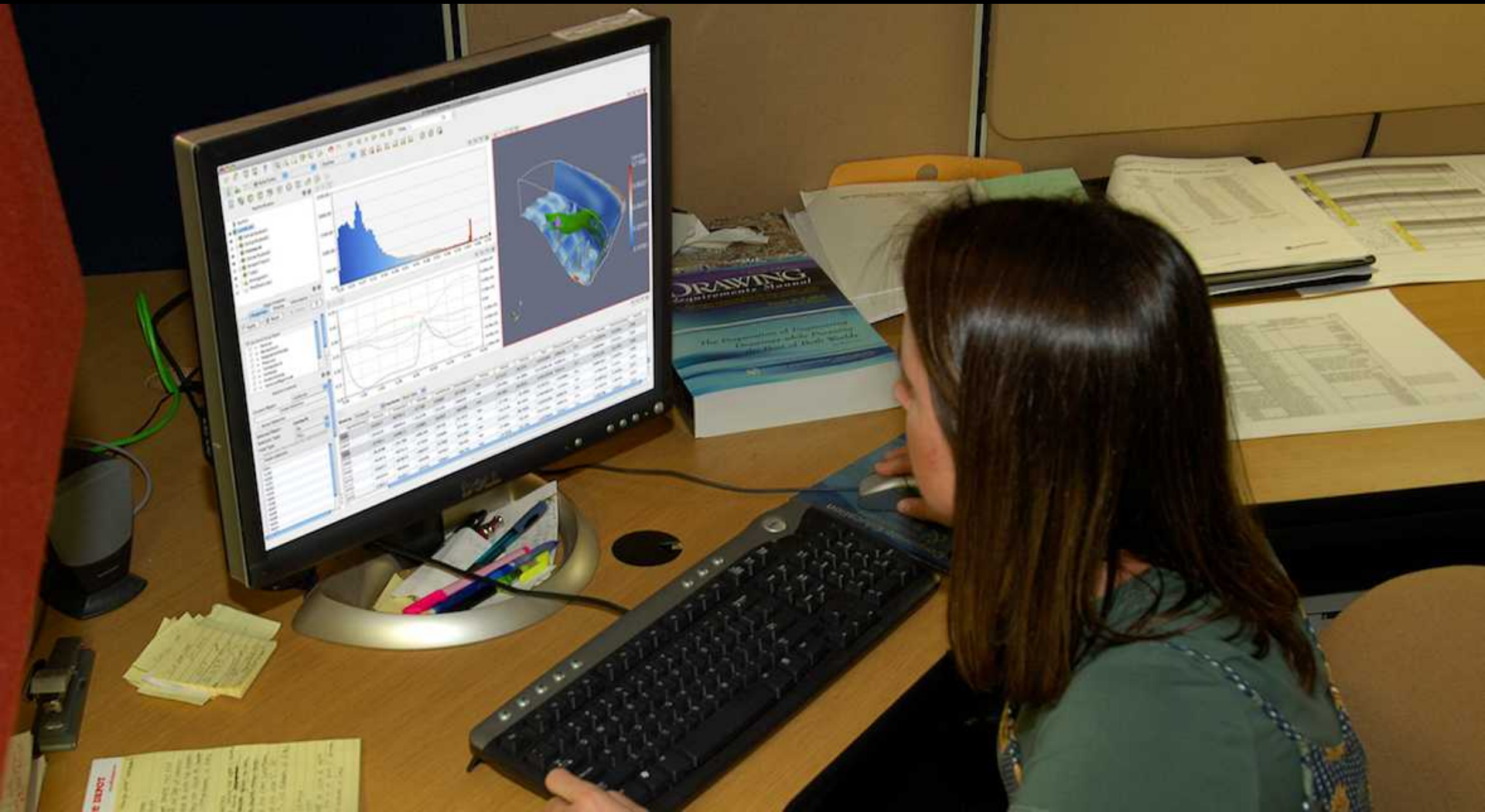


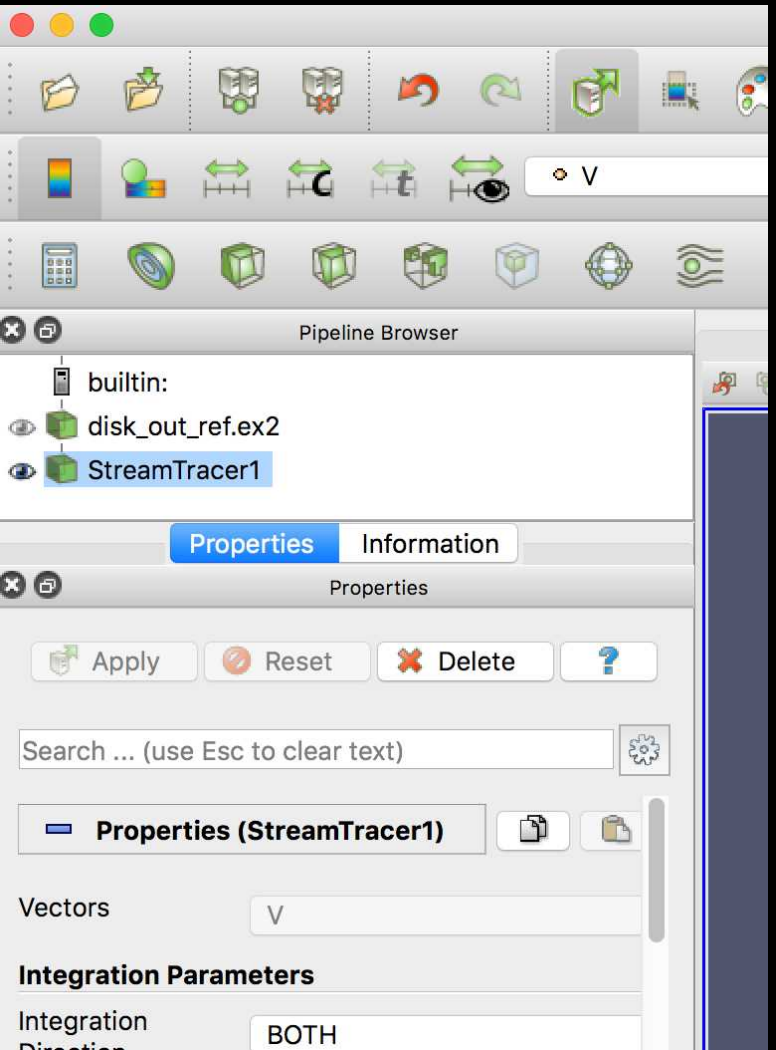
FIG. 2. Computer-made rapid-sequence drawing to illustrate the movie

Reasons Why We Don’t All Do In Situ Visualization All the Time

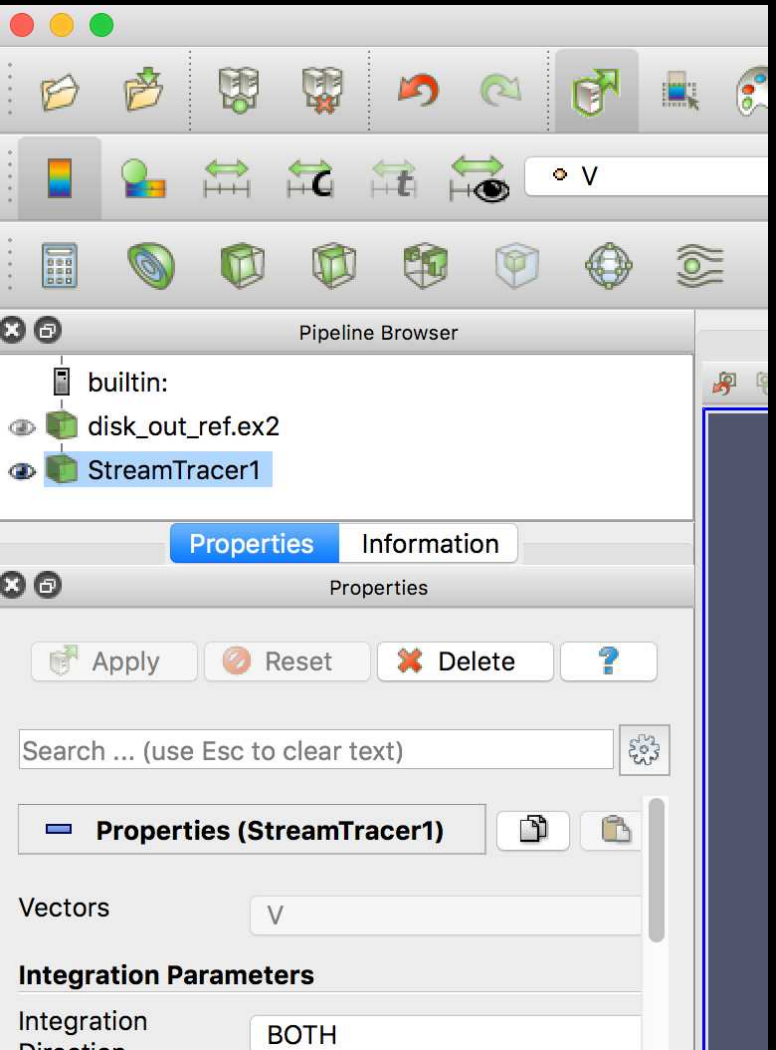
Loss of Interactivity



Expressing the Visualization



Expressing the Visualization




```
cthsource = CTHSource ()

# Create a view of the ball impacting the brick
ballblock = ExtractCTHFragments (cthsource)
ballblock.MaterialFractionThreshold = 0.5
ballblock.SelectMaterialFractionArrays = ["VOLM1", "VOLM2"]
ballblock.SelectMassArrays = ["M1", "M2"]

# Reflect the simulation result about the simulation axis of symmetry
normBall = GenerateSurfaceNormals (ballblock)
normBall.FeatureAngle = 90
normBall.Splitting = 1
normBall.PieceInvariant = 1
reflBall = Reflect (normBall)
reflBall.Center = 0.0
reflBall.Plane = "Y"

# Function to make the appropriate colorbar
def MakeGoldRedLT():
  lt = servermanager.rendering.PVLookupTable ()
  servermanager.Register(lt)
  lt.RGBPoints = [1, 0.56, 0.23, 0.23, 0, 0.9568, 0.6928, 0.3294]
  lt.ColorSpace = "RGB"
  return lt

# Function to initialize the display
def SetupDisplay(proxy, lt):
  SetDisplayProperties (proxy, SuppressLOD = 1)
  SetDisplayProperties (proxy, StaticMode = 1)
  SetDisplayProperties (proxy, LookupTable = lt)
  SetDisplayProperties (proxy, Specular = 0.1)
  SetDisplayProperties (proxy, SpecularPower = 40)
  SetDisplayProperties (proxy, SpecularColor = [1.0, 1.0, 1.0])

# Setup the initial display
```

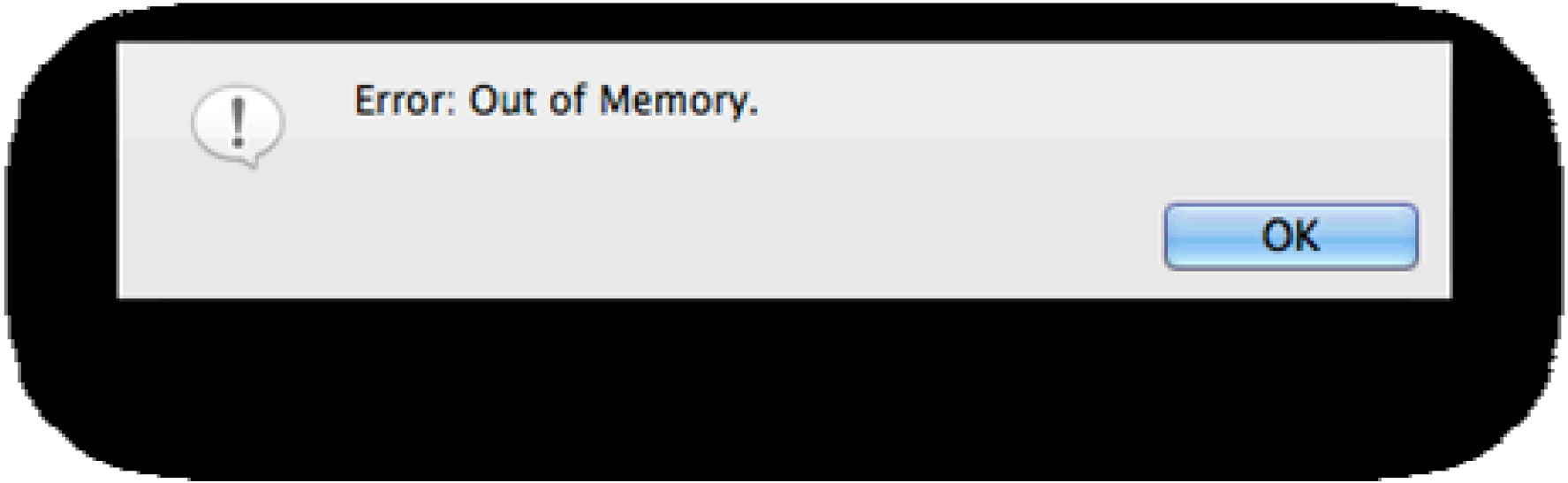
Cramming Codes Together

Sandia National Laboratories

Simulation

Visualization

Using Up Resources

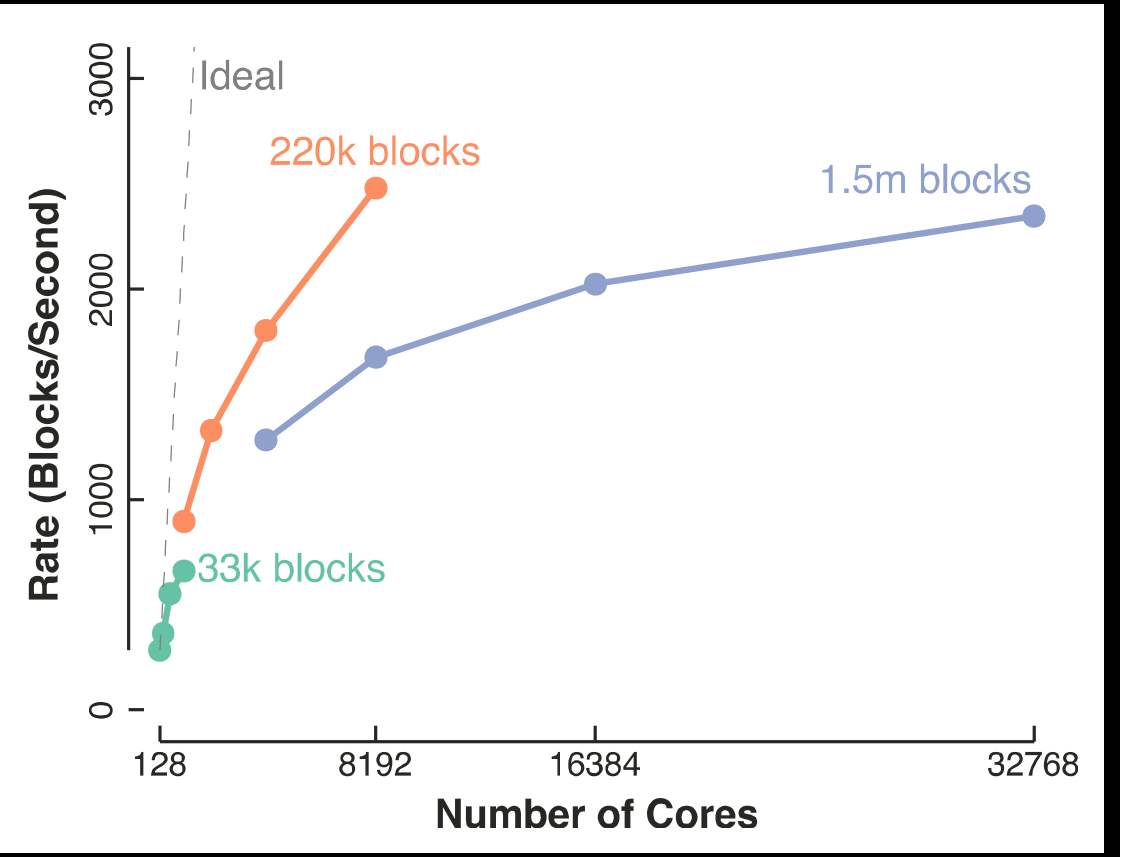


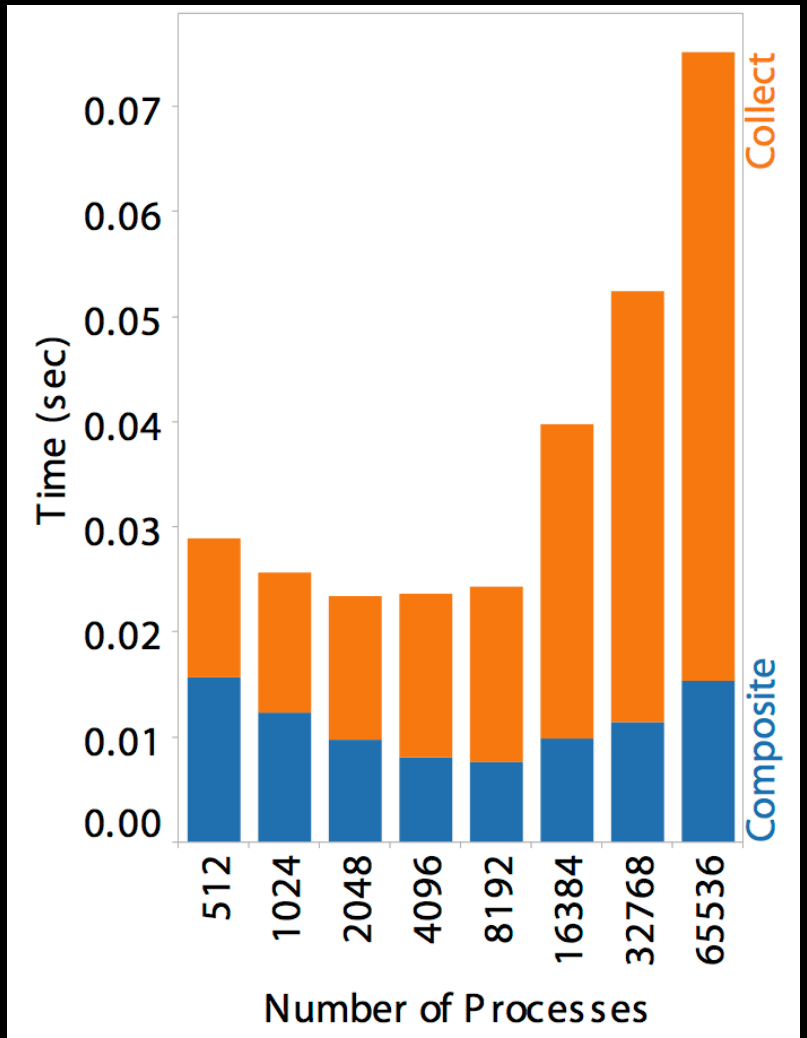
More Complexity = Less Robustness

Sandia National Laboratories

```
grsec: use of CAP_SYS_ADMIN in chroot denied for /sysroot/sbin/load_policy[load_
policy:260] uid/euid:0/0 gid/egid:0/0, parent /init[init:1] uid/euid:0/0 gid/egi
d:0/0
grsec: use of CAP_SYS_ADMIN in chroot denied for /sysroot/sbin/load_policy[load_
policy:260] uid/euid:0/0 gid/egid:0/0, parent /init[init:1] uid/euid:0/0 gid/egi
d:0/0
dracut: FATAL: Initial SELinux policy load failed. Machine in enforcing mode. To
 disable selinux, add selinux=0 to the kernel command line.
dracut: Refusing to continue



Kernel panic - not syncing: Attempted to kill init!
_
```

Scaling


Ideal
220k blocks
1.5m blocks
33k blocks
Rate (Blocks/Second)
0
1000
2000
3000
128
8192
16384
32768
Number of Cores


Time (sec)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Collect
Composite
512
1024
2048
4096
8192
16384
32768
65536
Number of Processes

Should We do In Situ Visualization at All?

In Situ Visualization

Sandia National Laboratories

Pros

- Batch/Automated Processing
- Early Access to Data
- Access to More Data

Cons

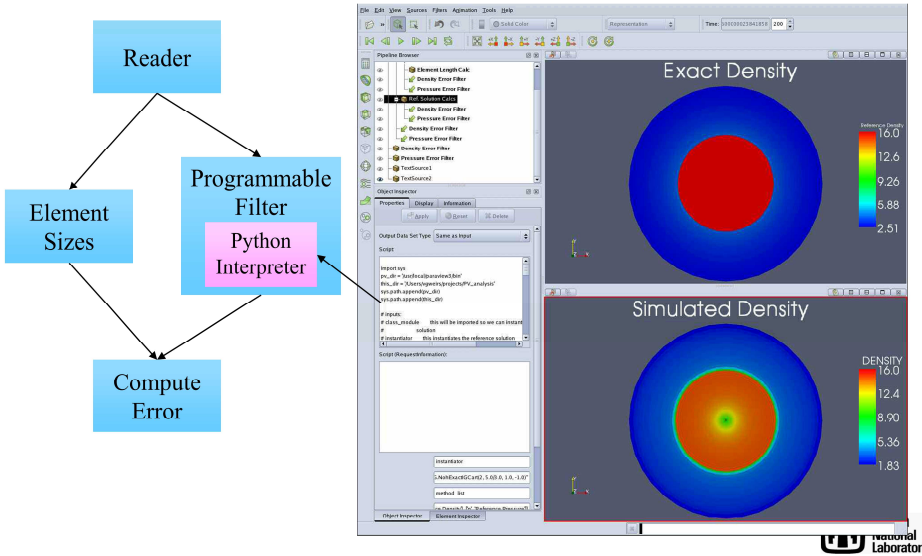
- Loss of Interactivity
- Hard to Express the Visualization
- Difficult Collaborations of Code Teams
- Extra Burden of Resources
- Less Robust
- Requires More Scaling

Post Hoc Batch/Automated Processing

Post-Processing V&V Level II ASC Milestone

- **FY07**
  - **Scripting for automated post processing**
  - **User defined functions at run time**
  - **Scalability**
- **FY08**
  - **Fragment identification and characterization**

Python Scripting, Server Side



In Situ Visualization

Sandia National Laboratories

Pros

- ~~Batch/Automated Processing~~
- Early Access to Data
- Access to More Data

Cons

- Loss of Interactivity
- Hard to Express the Visualization
- Difficult Collaborations of Code Teams
- Extra Burden of Resources
- Less Robust
- Requires More Scaling

In Situ Visualization is visualization that *necessarily* starts before the data producer finishes.

Post Hoc Visualization is visualization that can start at some arbitrary time after the data are produced.

Reload-Button – Custome...

https://paraview.uservoice.com/forums/11350-general/suggestions/315870-reload-button

ParaView
Parallel Visualization Application

I suggest you ...

← General

307
votes

Vote

Reload-Button

It would be nice to have a reload-button for time-dependend datasets. So you don't have to apply all your filters again if the simulation is evolved in time.

R. Prignitz shared this idea · Sep 14, 2009 · Flag idea as inappropriate...

ADMIN

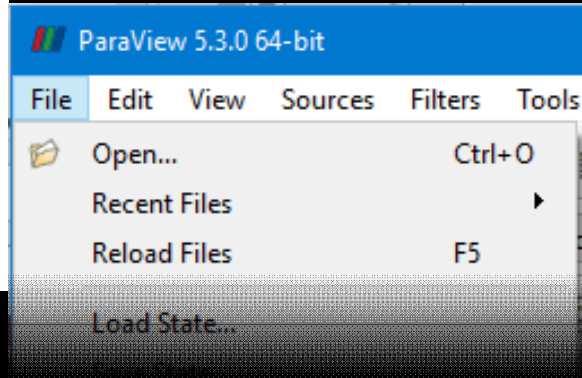
COMPLETED · Utkarsh Ayachit (Admin, paraview) responded · Sep 7, 2016

This feature is now available in the development repository and will be included in the next release, ParaView 5.2.

See the merge-request (now merged) for details: https://gitlab.kitware.com/paraview/paraview/merge_requests/1000

For details see https://blog.kitware.com/refreshingreloading-files-in-paraview/

Show previous admin responses (2)



In Situ Visualization

Sandia National Laboratories

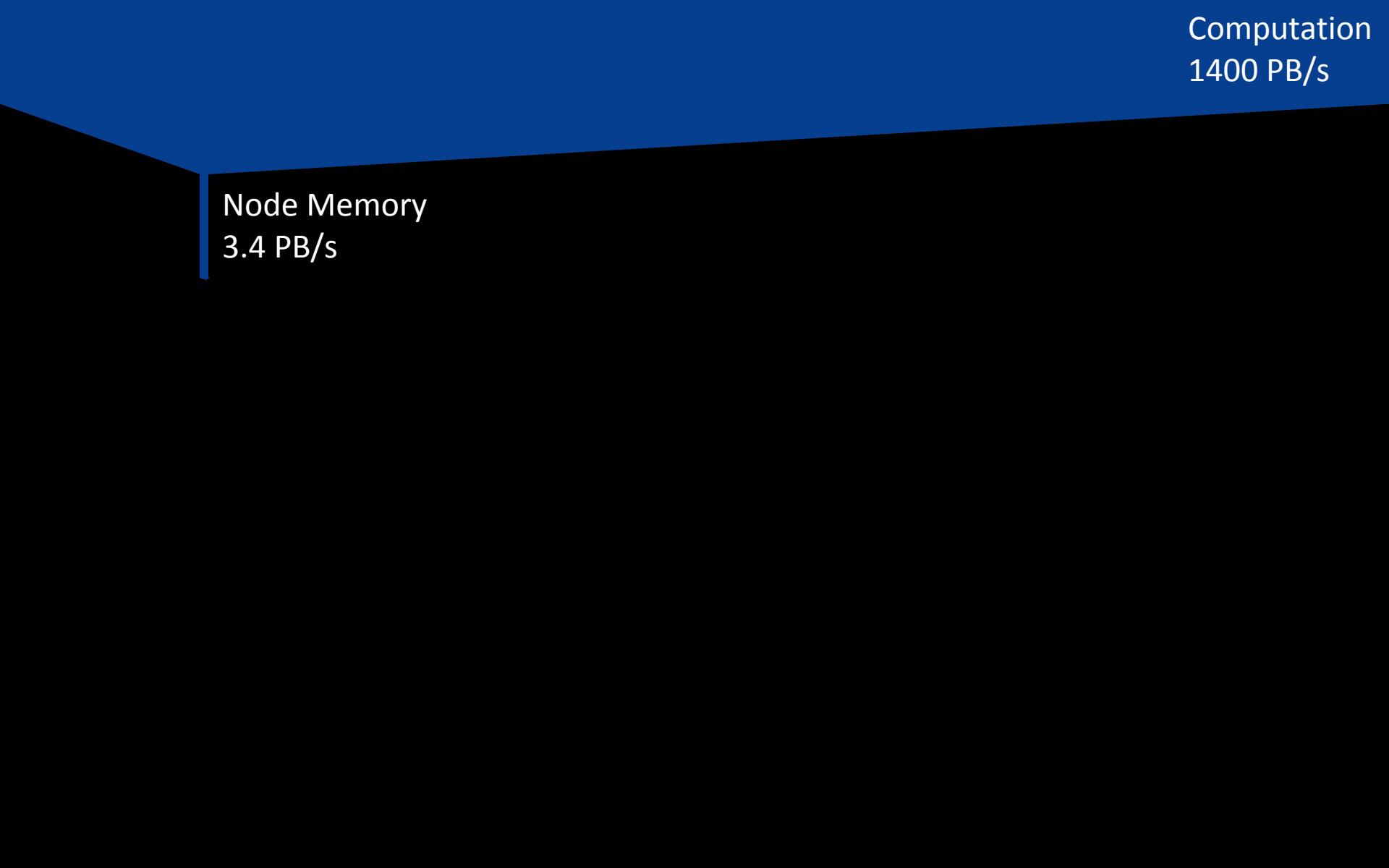
Pros

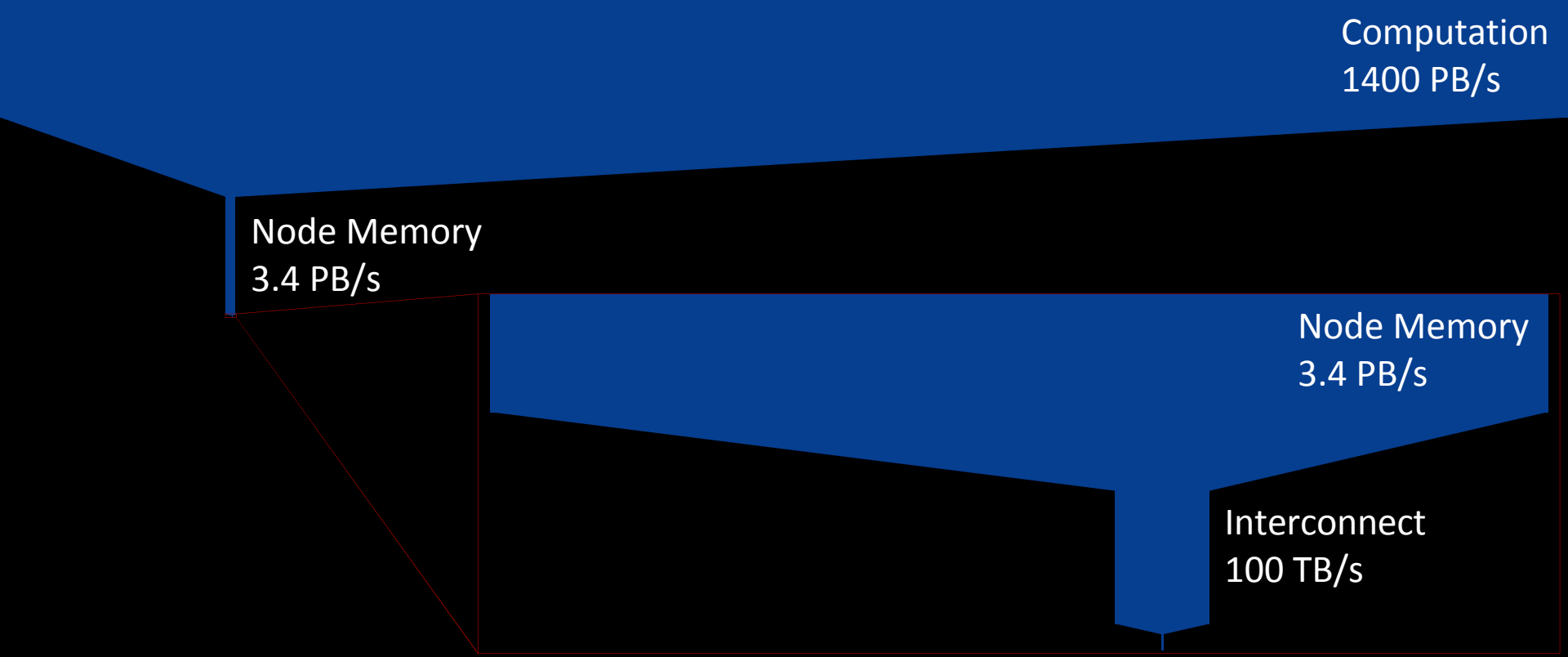
- ~~Batch/Automated Processing~~
- ~~Early Access to Data~~
- Access to More Data

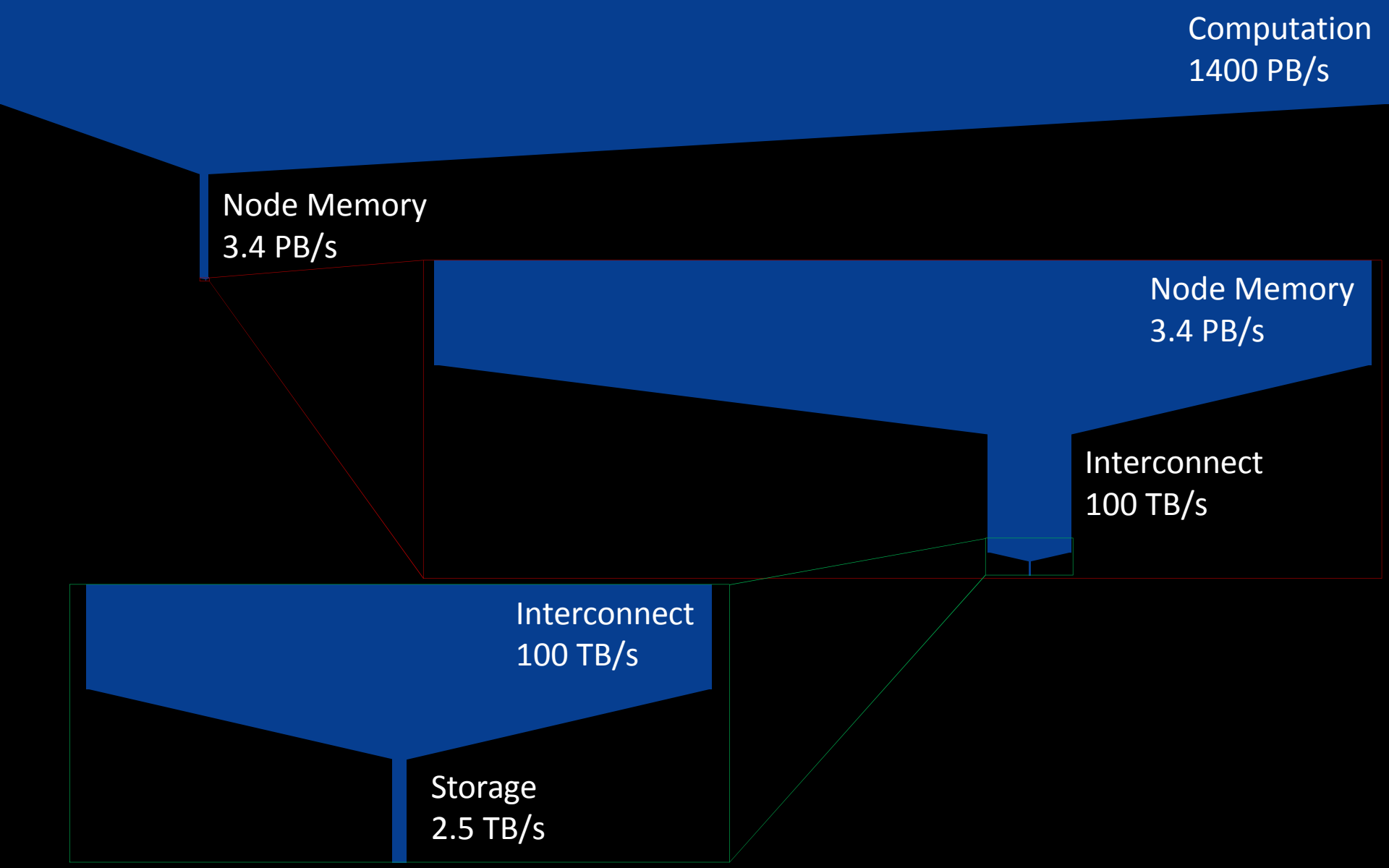
Cons

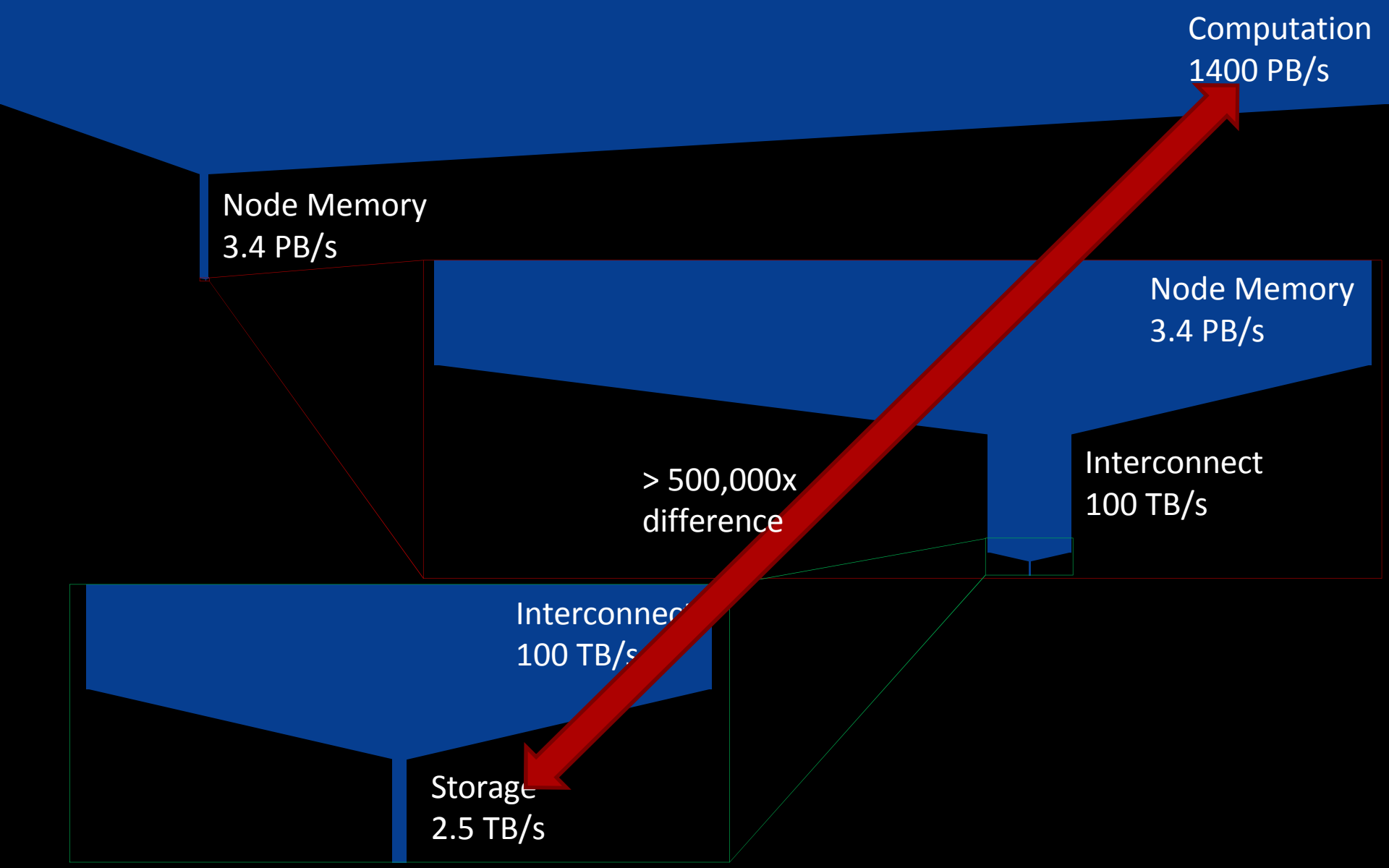
- Loss of Interactivity
- Hard to Express the Visualization
- Difficult Collaborations of Code Teams
- Extra Burden of Resources
- Less Robust
- Requires More Scaling

Why We *Have* to do In Situ Visualization


Computation
1400 PB/s
Node Memory
3.4 PB/s


Computation
1400 PB/s
Node Memory
3.4 PB/s
Node Memory
3.4 PB/s
Interconnect
100 TB/s


Computation
1400 PB/s
Node Memory
3.4 PB/s
Node Memory
3.4 PB/s
Interconnect
100 TB/s
Interconnect
100 TB/s
Storage
2.5 TB/s


Computation
1400 PB/s
Node Memory
3.4 PB/s
Node Memory
3.4 PB/s
> 500,000x
difference
Interconnect
100 TB/s
Interconnect
100 TB/s
Storage
2.5 TB/s

In Situ Visualization

Sandia National Laboratories

Pros

- ~~Batch/Automated Processing~~
- ~~Early Access to Data~~
- Access to More Data

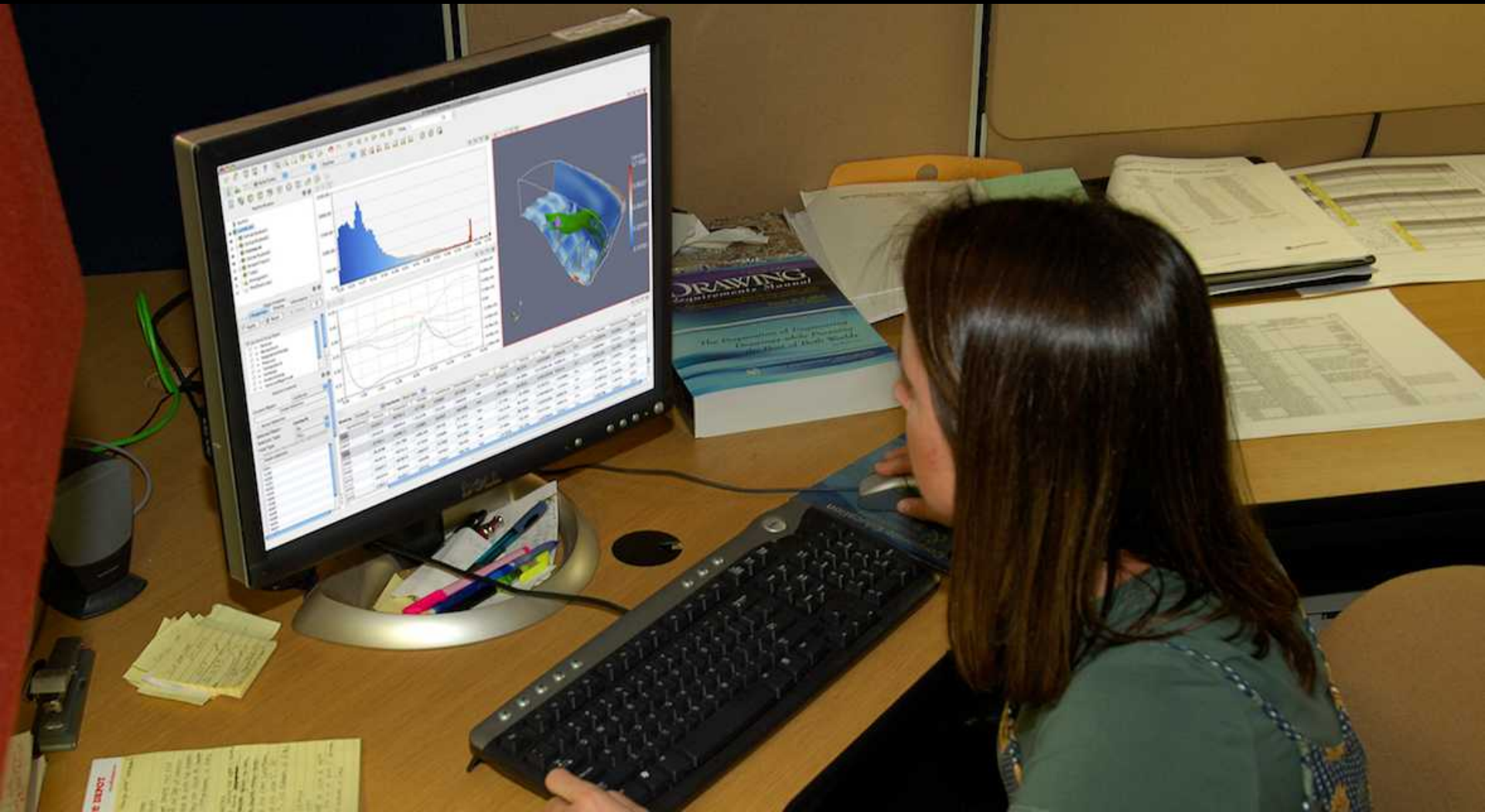
We need this when we cannot save enough to disk.

Cons

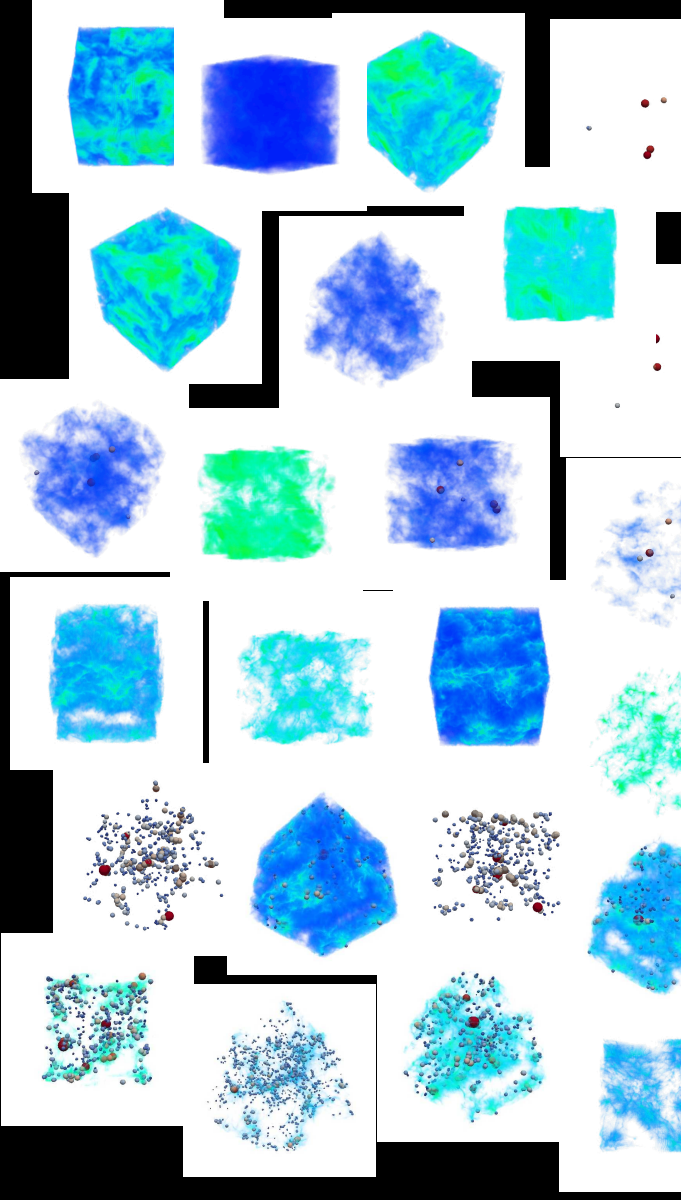
- Loss of Interactivity
- Hard to Express the Visualization
- Difficult Collaborations of Code Teams
- Extra Burden of Resources
- Less Robust
- Requires More Scaling

Addressing Shortcomings of In Situ

Loss of Interactivity




SC15
Kitware Booth #2639
Presentation Schedule
Tuesday, November 17th
Exhibit Hours 10:00-6:00
Wednesday, November 18th
Exhibit Hours 10:00-6:00
Live From Titan@ORNL!
Open Source Platforms
Advanced Research Solutions



Cinema Desktop

Time
Value: 4

Theta
Value: 45

Phi
Value: 315

Visible	Layer Description
	halos
Yes	on
x	off
	volume

volumeIdx
Value: 3

Value Color Lut

None

Import

Geometry Color

Use color

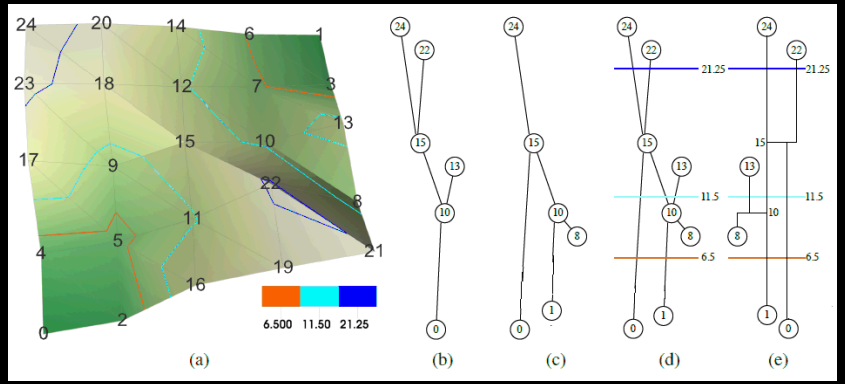
Set Color

Use lighting

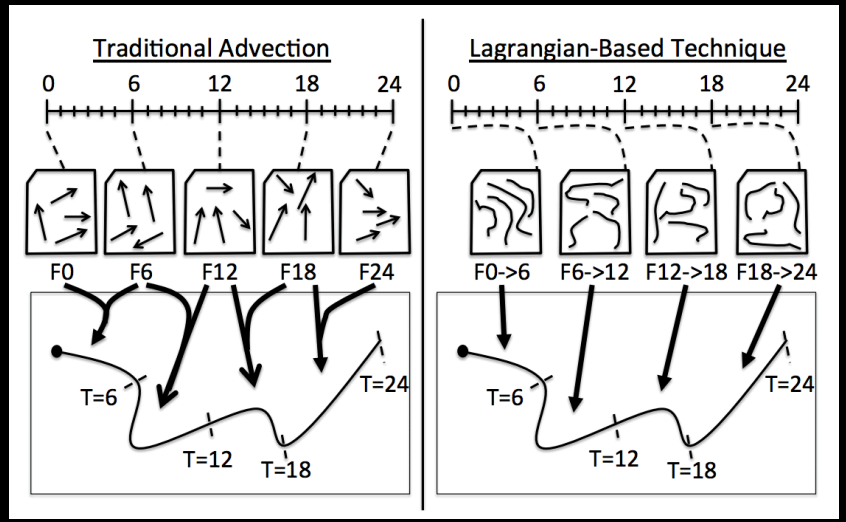
Change Background

Write Smarter, Not Harder

- Use analysis to select important data
  - Entropy, feature detection, topological methods, machine learning
- Transform data for better information extraction
  - Compression, Lagrangian flow representation

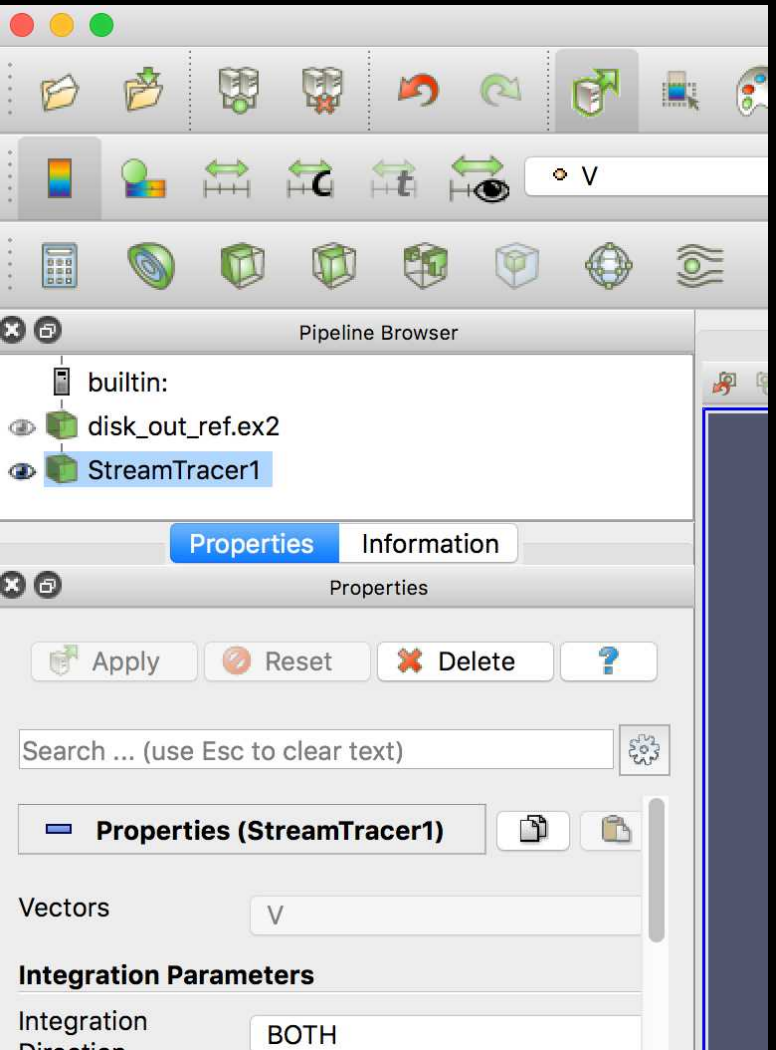


Carr, et al. LDAV 2016. Contour trees can express critical isovalues in spectrum of contours.

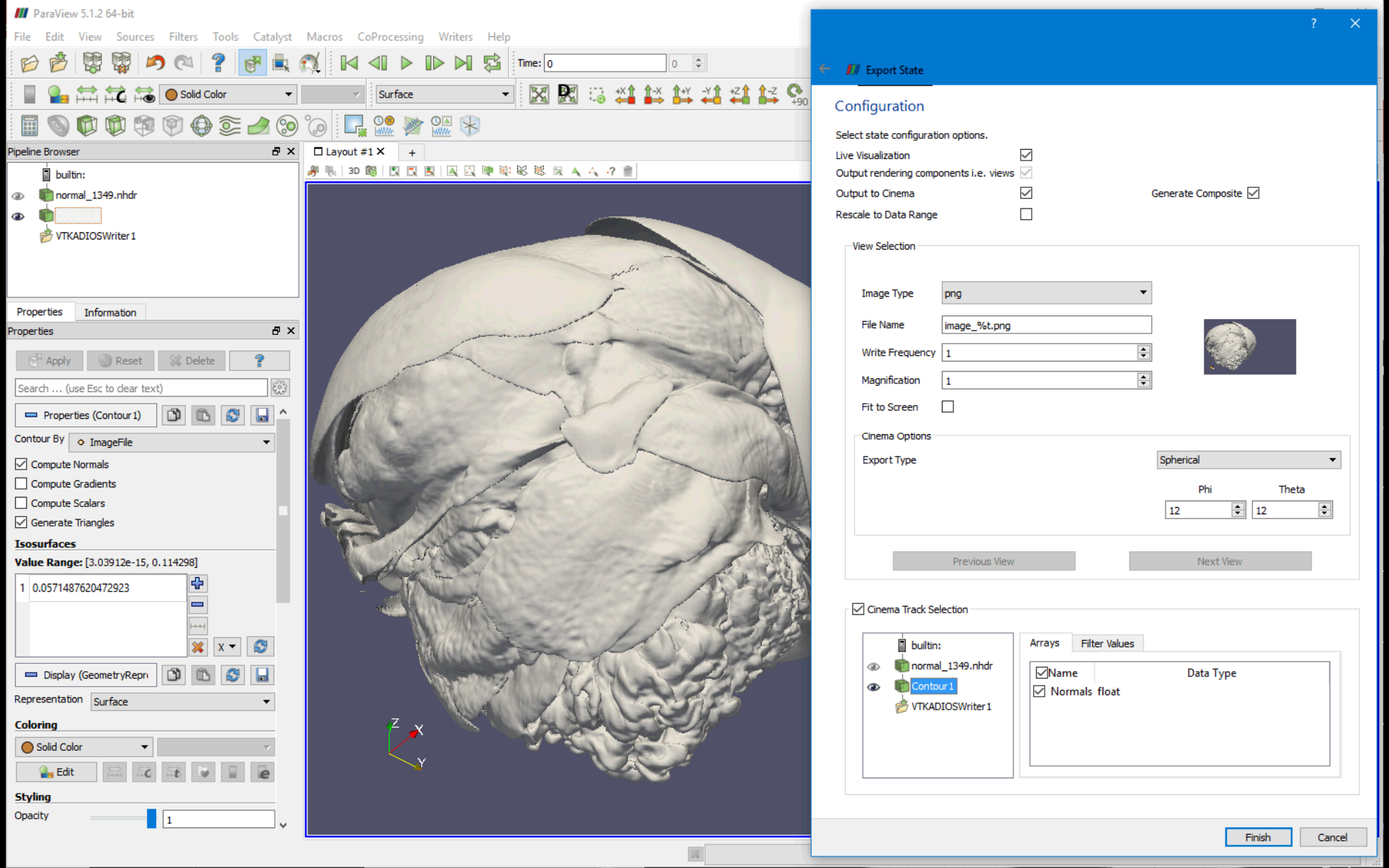


Agranovsky, et al. LDAV 2014. Capturing Lagrangian trajectories instead of field vectors captures flow better with fewer time steps.

Expressing the Visualization




```
cthsource = CTHSource ()

# Create a view of the ball impacting the brick
ballblock = ExtractCTHFragments (cthsource)
ballblock.MaterialFractionThreshold = 0.5
ballblock.SelectMaterialFractionArrays = ["VOLM1", "VOLM2"]
ballblock.SelectMassArrays = ["M1", "M2"]

# Reflect the simulation result about the simulation axis of symmetry
normBall = GenerateSurfaceNormals (ballblock)
normBall.FeatureAngle = 90
normBall.Splitting = 1
normBall.PieceInvariant = 1
reflBall = Reflect (normBall)
reflBall.Center = 0.0
reflBall.Plane = "Y"

# Function to make the appropriate colorbar
def MakeGoldRedLT():
  lt = servermanager.rendering.PVLookupTable ()
  servermanager.Register(lt)
  lt.RGBPoints = [1, 0.56, 0.23, 0.23, 0, 0.9568, 0.6928, 0.3294]
  lt.ColorSpace = "RGB"
  return lt

# Function to initialize the display
def SetupDisplay(proxy, lt):
  SetDisplayProperties (proxy, SuppressLOD = 1)
  SetDisplayProperties (proxy, StaticMode = 1)
  SetDisplayProperties (proxy, LookupTable = lt)
  SetDisplayProperties (proxy, Specular = 0.1)
  SetDisplayProperties (proxy, SpecularPower = 40)
  SetDisplayProperties (proxy, SpecularColor = [1.0, 1.0, 1.0])

# Setup the initial display
```


ParaView 5.1.2 64-bit
File Edit View Sources Filters Tools Catalyst Macros CoProcessing Writers Help
Time: 0
Solid Color
Surface
Pipeline Browser
builtin:
normal_1349.nhdr
Contour1
VTKADIOSWriter1
Properties Information
Properties
Apply Reset Delete
Search ... (use Esc to clear text)
Properties (Contour1)
Contour By ImageFile
Compute Normals
Compute Gradients
Compute Scalars
Generate Triangles
Isosurfaces
Value Range: [3.03912e-15, 0.114298]
1 0.0571487620472923
Display (GeometryRepr
Representation Surface
Coloring
Solid Color
Edit
Styling
Opacity 1
Layout #1
3D
Export State
Configuration
Select state configuration options.
Live Visualization
Output rendering components i.e. views
Output to Cinema
Generate Composite
Rescale to Data Range
View Selection
Image Type png
File Name image_%t.png
Write Frequency 1
Magnification 1
Fit to Screen
Cinema Options
Export Type Spherical
Phi 12
Theta 12
Previous View
Next View
Cinema Track Selection
builtin:
normal_1349.nhdr
Contour1
VTKADIOSWriter1
Arrays
Filter Values
Name
Data Type
Normals float
Finish
Cancel

Cramming Codes Together

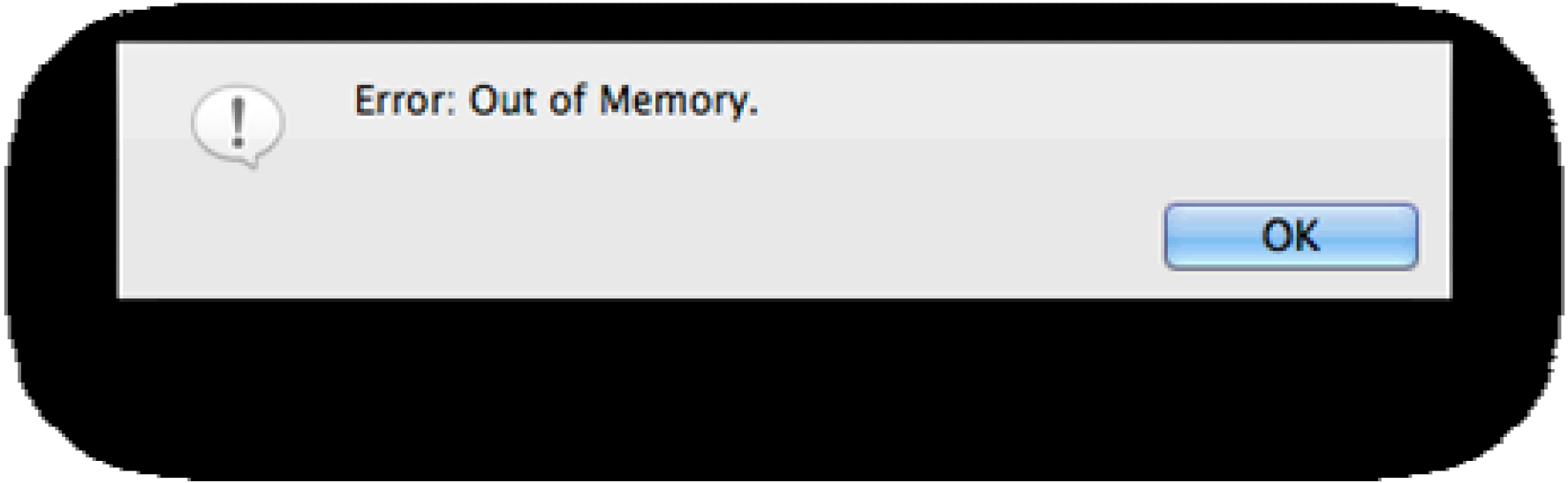
Sandia National Laboratories

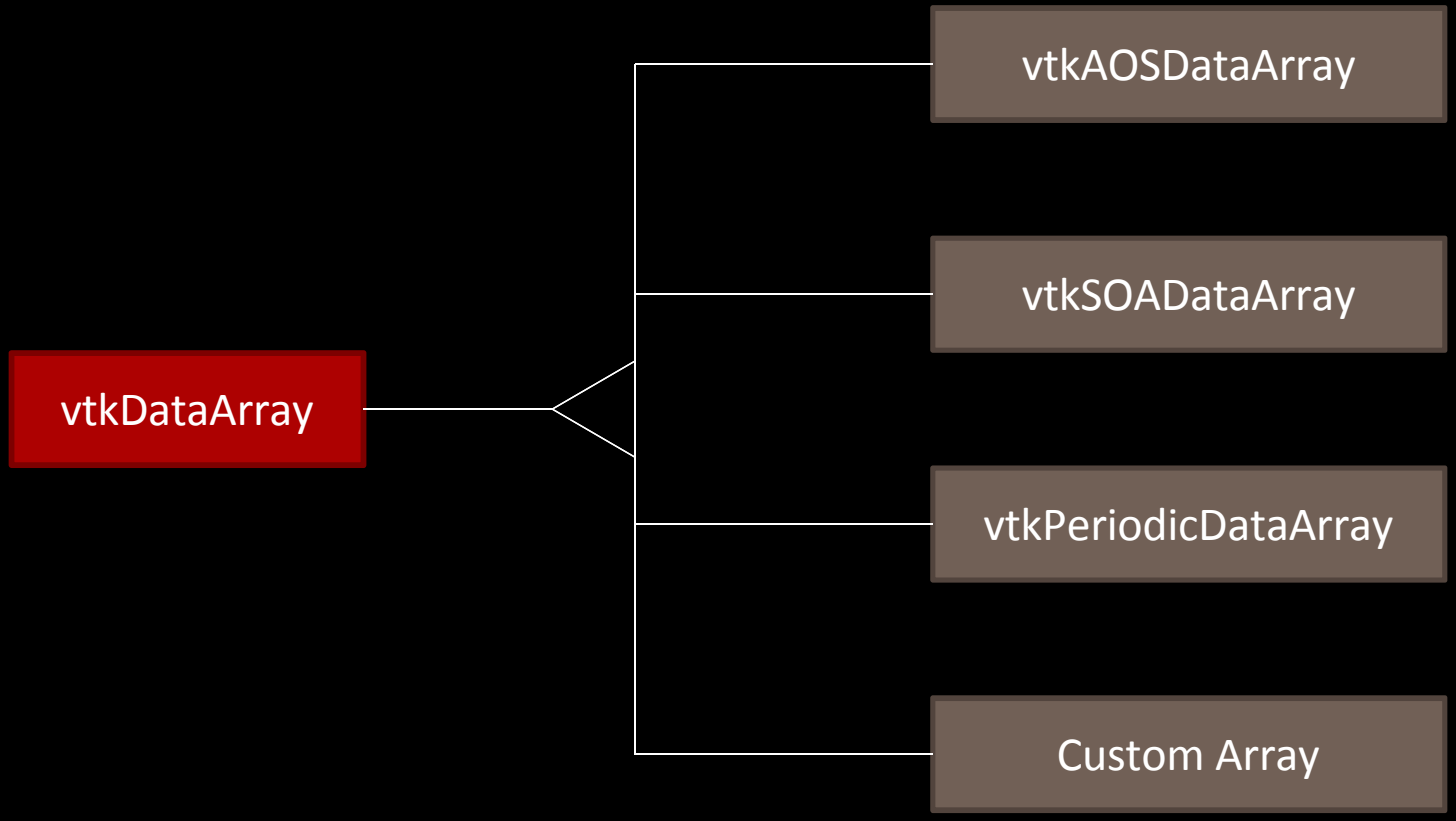
Simulation

Visualization

Photo credit: Blogspot.com

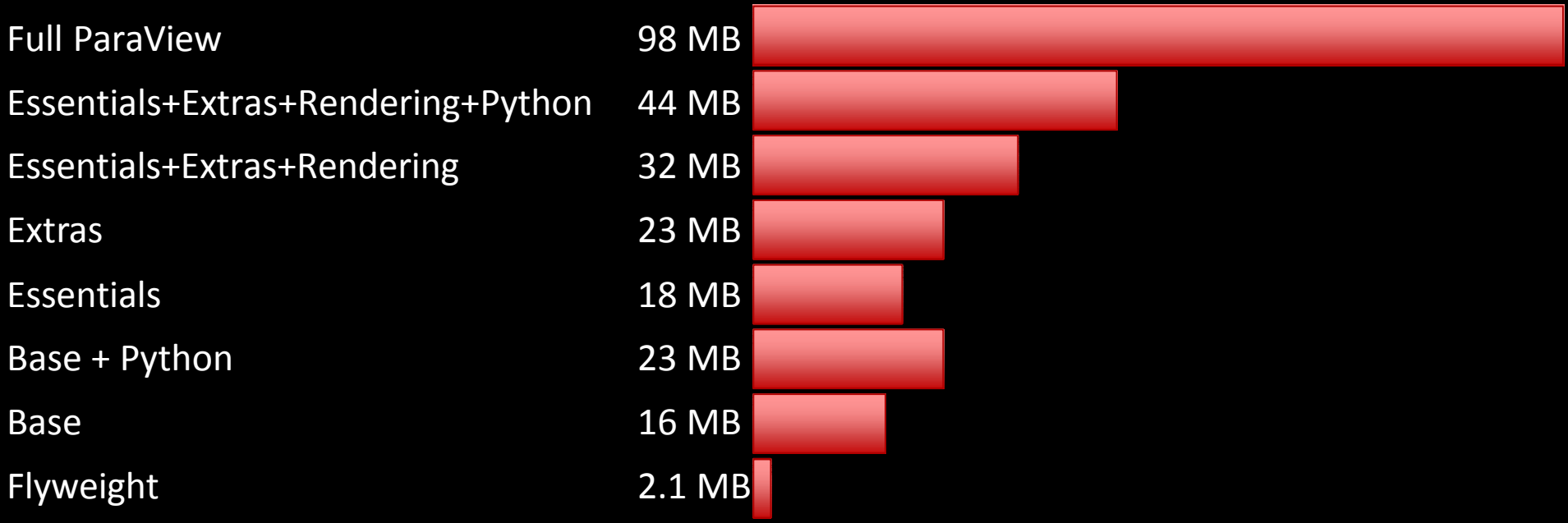
Using Up Resources




vtkDataArray
vtkAOSDataArray
vtkSOADataArray
vtkPeriodicDataArray
Custom Array

ParaView Catalyst Editions

Sandia National Laboratories

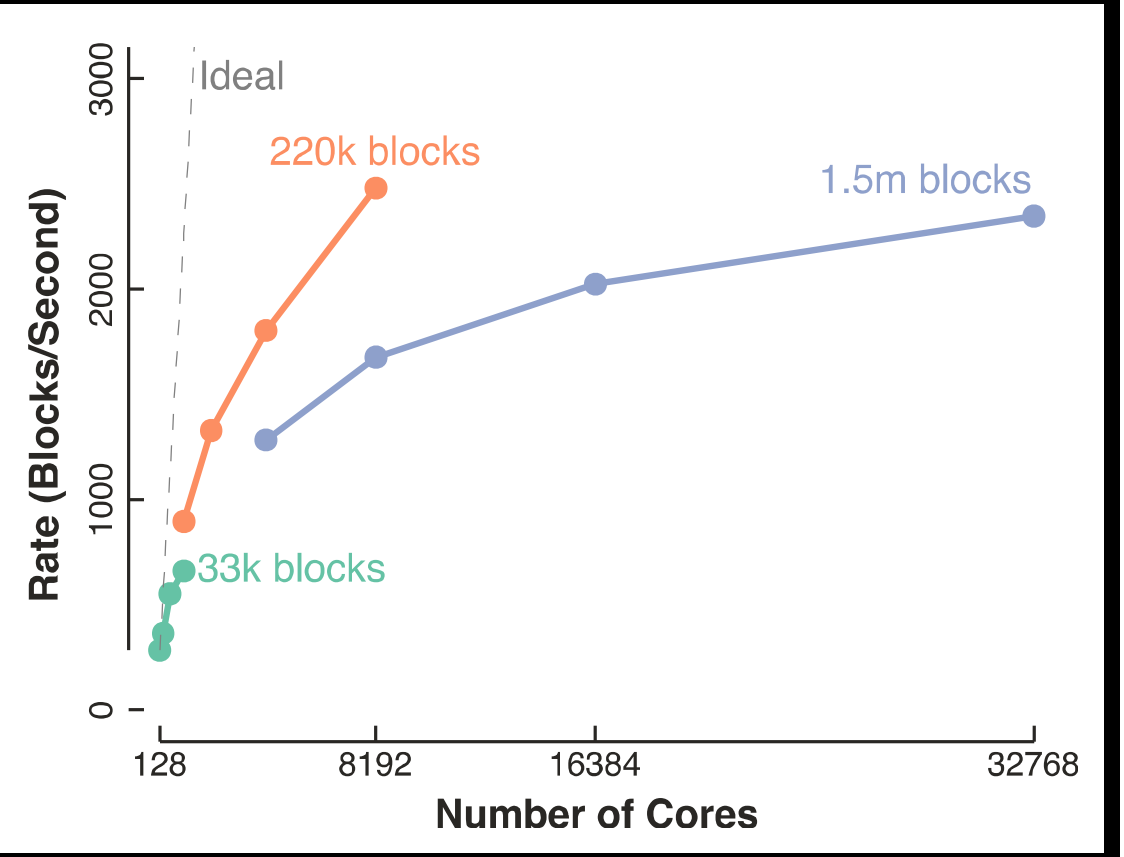


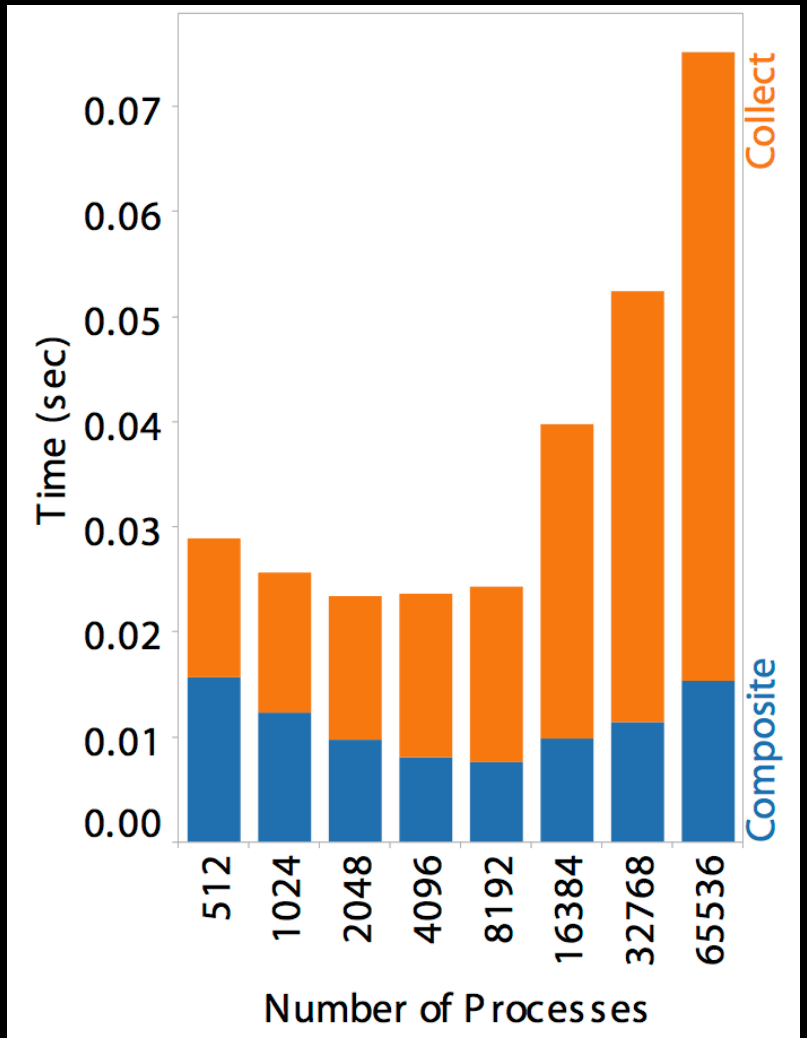
More Complexity = Less Robustness

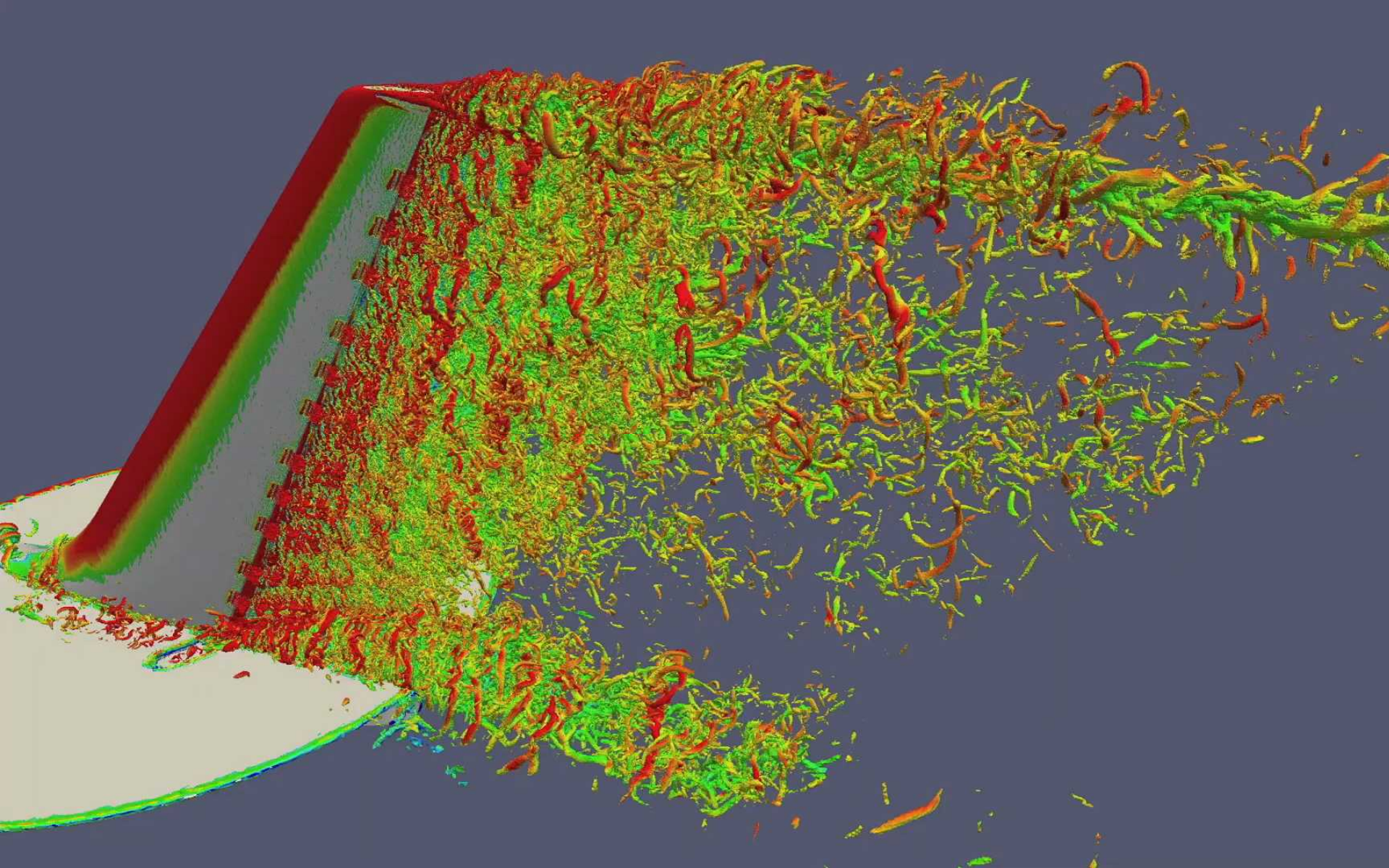
Sandia National Laboratories

```
grsec: use of CAP_SYS_ADMIN in chroot denied for /sysroot/sbin/load_policy[load_
policy:260] uid/euid:0/0 gid/egid:0/0, parent /init[init:1] uid/euid:0/0 gid/egi
d:0/0
grsec: use of CAP_SYS_ADMIN in chroot denied for /sysroot/sbin/load_policy[load_
policy:260] uid/euid:0/0 gid/egid:0/0, parent /init[init:1] uid/euid:0/0 gid/egi
d:0/0
dracut: FATAL: Initial SELinux policy load failed. Machine in enforcing mode. To
 disable selinux, add selinux=0 to the kernel command line.
dracut: Refusing to continue



Kernel panic - not syncing: Attempted to kill init!
_
```

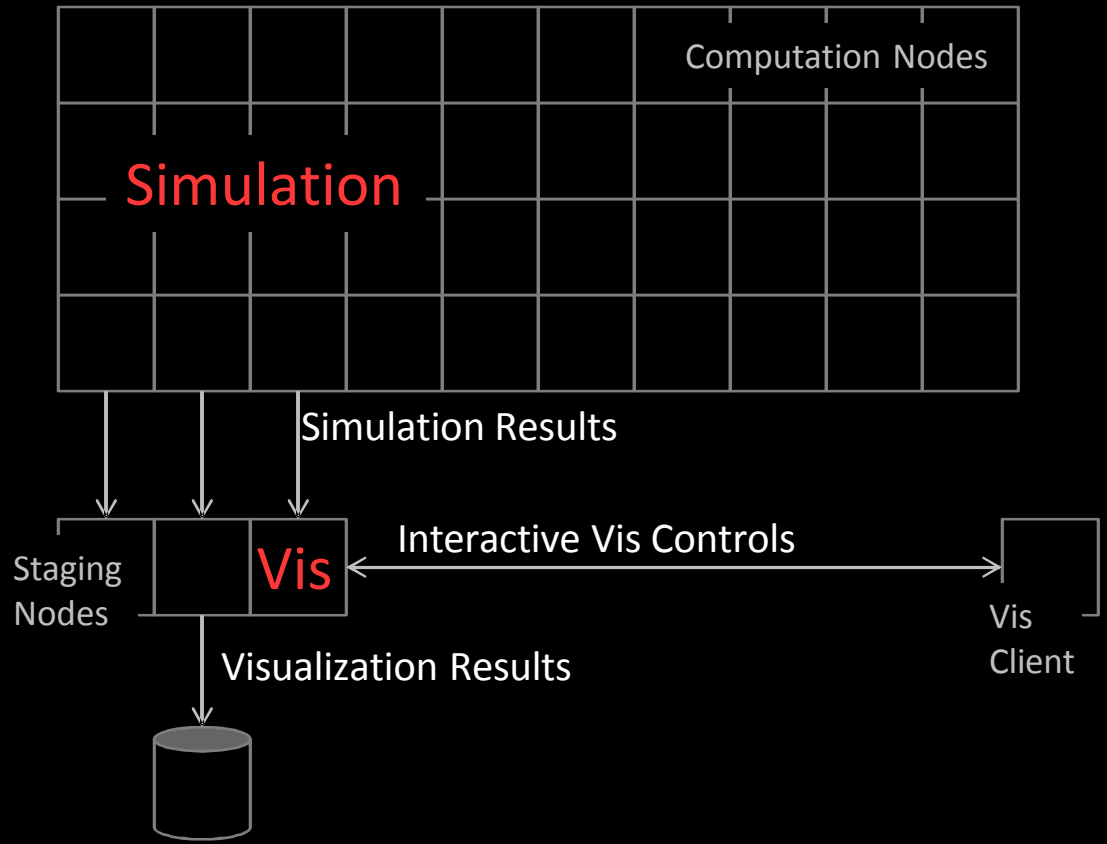
Scaling


Ideal
220k blocks
1.5m blocks
33k blocks
Rate (Blocks/Second)
0
1000
2000
3000
128
8192
16384
32768
Number of Cores


Time (sec)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Collect
Composite
512
1024
2048
4096
8192
16384
32768
65536
Number of Processes



Alternative to Scaling



Conclusions

Sandia National Laboratories

- In Situ takes away lots of beloved features of visualization
  - This has limited its adoption to specific use cases for almost 50 years
- The resurgence of interest (why we are here today) is because there is a pressing need in HPC
- File I/O speed has been losing ground since forever
  - From a practical standpoint, we are hitting resource limits that force us to move to in situ
- Understand why you are doing in situ and what you are giving up to be successful

Acknowledgements

- This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Numbers 10-014707, 12-015215, and 14-017566.
- This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, in support of the nation's exascale computing imperative.
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.
- **Thanks to many, many partners in labs, universities, and industry.**