Quantitative and Comparative Visualization Applied to Cosmological Simulations

James Ahrens

Katrin Heitmann, Salman Habib

Lee Ankeny, Patrick S. McCormick, Jeff Inman

Los Alamos National Laboratory

Ryan Armstrong, Kwan-Liu Ma

University of California at Davis

SciDAC 2006

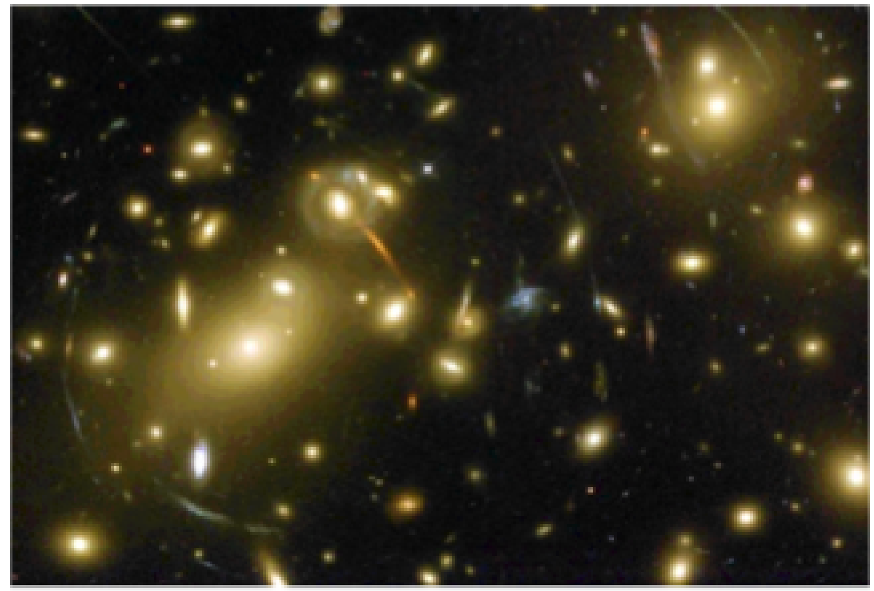
June 25-29 Denver

Los Alamos NATIONAL LABORATORY EST. 1943

The World's Greatest Science Protecting America

The Content of the Universe

- Standard Model of Cosmology
  - ~73% of a mysterious dark energy
  - ~23% of an unknown dark matter component
  - ~4% baryons
  - Constraints on ~20 cosmological parameters, including optical depth, spectral index, hubble constant, …
  - Values are known to an accuracy of +/- 10%
- For comparison: the parameters of the “Standard Model for Particle Physics” are known with 0.1% accuracy



Understanding the Universe

- Science today
  - Theory
  - Simulation
  - Observation / Experiment
- Cosmological simulations follow the formation of nonlinear structure in dark and luminous matter.
- Our goal is to understand sources of inconsistency between different cosmological simulation codes.

Robustness of Cosmological Simulations: Large Scale Structure

- Heitmann, Ricker, Warren and Habib, ApJS, 160, 128, (2005)

- How well do different N-body codes agree on various statistics?
- Test and compare 6 different N-body codes for simulations of structure formation, dark matter only
- Every code starts from identical particle initial conditions

Robustness of Cosmological Simulations: Large Scale Structure - Codes

- Mesh-based Cosmology Code
  - Multi-species particle mesh code (Habib et al. in prep.)
- FLASH
  - Adaptive mesh refinement
    - Hydrodynamics and dark matter code (Fryxell et al. 2000)
- Hashed-Oct Tree
  - Tree code with SPH (Warren & Salmon 1993)
- Galaxies with Dark matter and Gas intEracTions
  - Tree code with SPH (Springel et al. 2001)
- HYDRA, AP^{3}M code with SPH (Couchman et al. 1995)
- TreePM, pure dark matter code (Xu 1995, Bode et al. 2000)

Robustness of Cosmological Simulations: Large Scale Structure - Data

- For each simulation
  - 16 million particles
  - Point, velocity, mass and tag variables
- http://t8web.lanl.gov/people/heitmann/arxiv/

Our Visualization and Analysis Approach

- Scientific method
  1) Form hypothesis
  2) Qualitative – Visualization
     – Intuitive exploration
  3) Quantitative – Analysis
     – Define and measure
- Tight integration

- Bottom-up or top-down focus?
  – Bottom-up application focus
    – Learn and generalize over time
- Work towards significantly improving the scientific analysis process by incorporating quantitative analysis as the driver for visualization.

Initial Approach for Cosmology Problem

- Initially
  - Define halos
    - Particles within 1/5 of the mean distance from each other form a halo
  - Count the halos

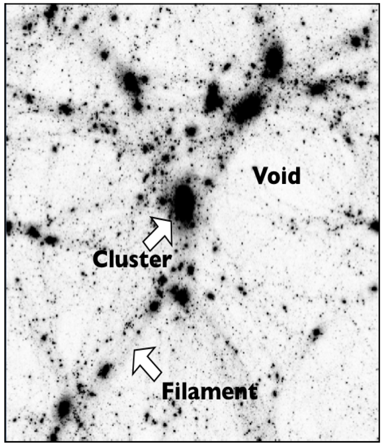
MC²	FLASH	HOT
49087	32494	54417
GADGET	TPM	HYDRA
55854	34367	54840

- Form hypothesis
  - Each simulation should generate the same number of halos

- Quantitative - Analysis
  - MC² (PM code, uniform grid) and FLASH (AMR code) have similar force resolution
  - Highest resolution (after refinement) of FLASH is the same as the MC² resolution throughout
  - FLASH is missing ~40% of the halos! - Why?

Refined Approach for Cosmology Problem

- Form hypothesis
  - Low density regions do not form as many halos as other density regions
- Qualitative - Visualization
  - Comparative visualization
- Quantitative – Analysis
  - Science-based feature definition and manipulation
    - Define density
      - Given a grid, map the particles into the grid elements, density is particle count
    - Count halos as a function of density
    - Also, consider only halos above a certain mass



Additional requirements

- High-performance
  - Reduce time to visual result or analysis
- Scalable
  - Handle massive data sets

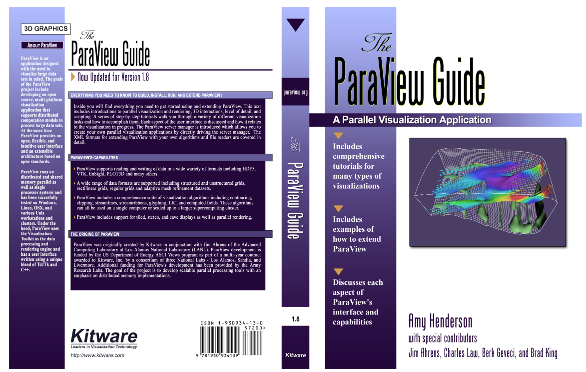
COMPUTER & COMPUTATIONAL SCIENCES

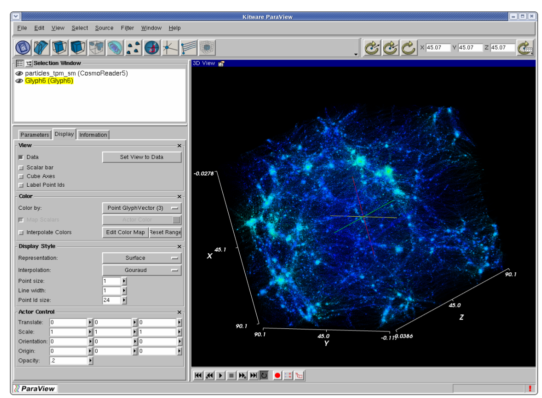
Application of the approach

- Paraview - open-source large data visualization package
  - Scalable
  - Comparative
- Scout - an analysis-language based, hardware-accelerated visualization package
  - High-performance
  - Quantitative

Vtk and ParaView - An Open Source Visualization Tool Suite for Scientists

- VTK
  - An open-source object-oriented visualization toolkit
  - www.vtk.org
- ParaView
  - An open-source, scalable multi-platform visualization application
  - Creates an open, flexible, and intuitive user interface for VTK
  - Project Lead: James Ahrens
  - www.paraview.org





- Agency funding
  - NSF, NIH, DOE, DOD
- Entities using/developing
  - Laboratories
    - ANL, NCSA, EVL
    - LANL, LLNL, SNL
    - CEA, CHCH
    - ARL
  - Commercial Companies
    - GE, DuPont
  - Universities
    - Stanford, UNC, Utah
- ~2000 mailing list participants

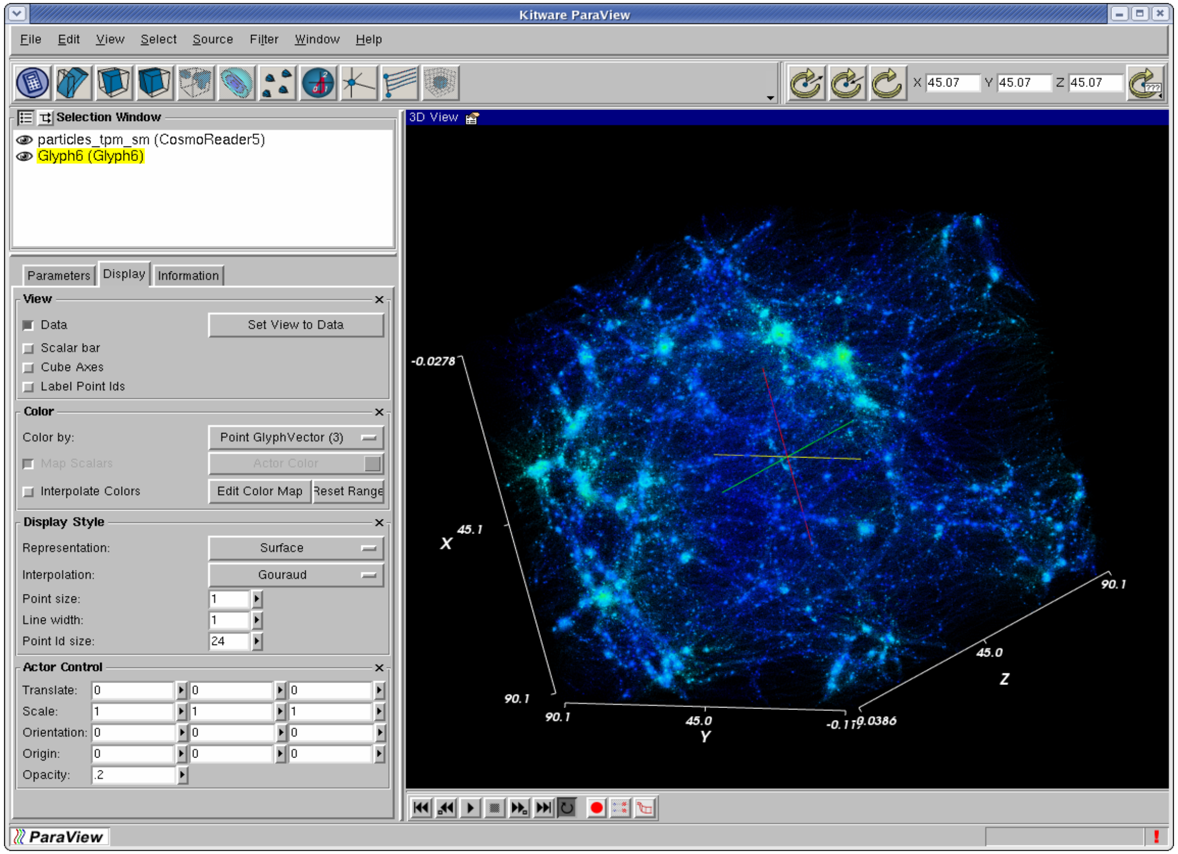
COMPUTER & COMPUTATIONAL SCIENCES

ParaView Overview

- Full functionality
  - Isosurfacing, cutting, clipping, volume rendering...
- Serial and parallel portability
  - Run on most serial and parallel platforms
    - Binaries for Windows, Linux, Mac
  - Distributed-memory execution
    - Commodity clusters
- Scalability
  - Data parallelism and incremental processing
  - Visualized a petabyte-sized test problem in 2001
- Advanced displays and rendering
  - Stereo, Tiled walls, CAVE
  - Automatic level of detail rendering
  - Compression for remote data transfer
- Supercomputing services
  - Parallel data server
  - Parallel rendering server
  - Client
- Visualization research with a real-world impact...

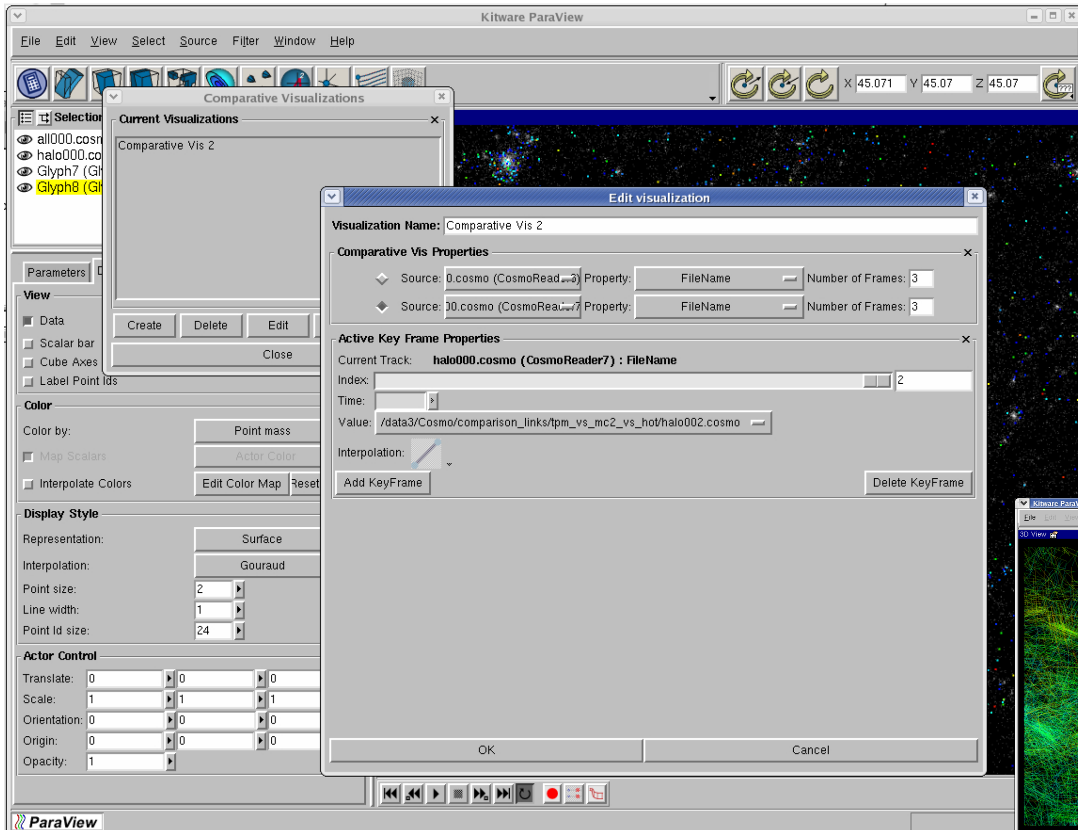
Refined Approach Using ParaView

- Qualitative - Visualization
  - ***Automated*** comparative visualization
- Quantitative - Analysis
  - Create modules and interfaces in ParaView that:
    - Define density, halos
    - Count and query on halos and density

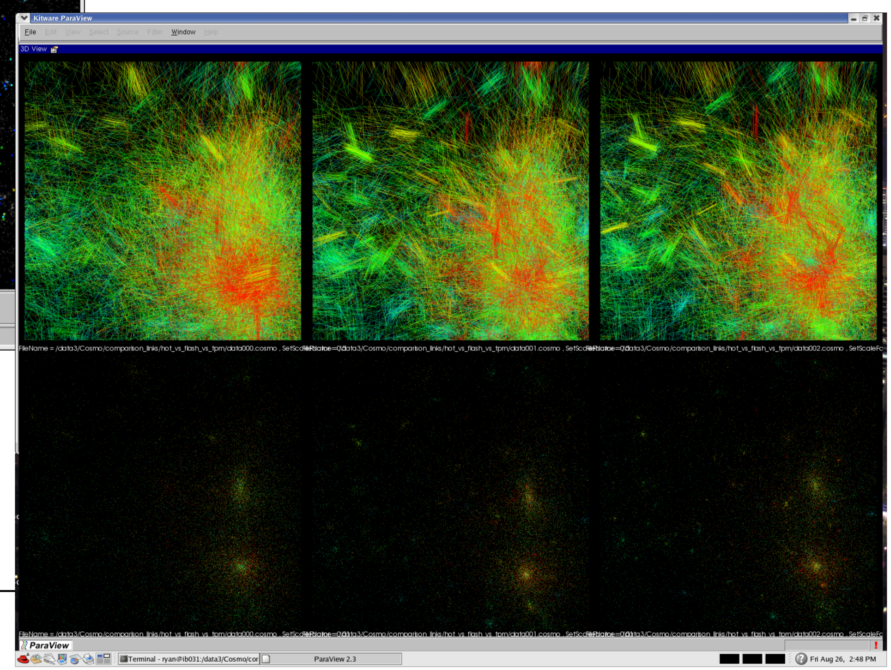


COMPUTER & COMPUTATIONAL SCIENCES

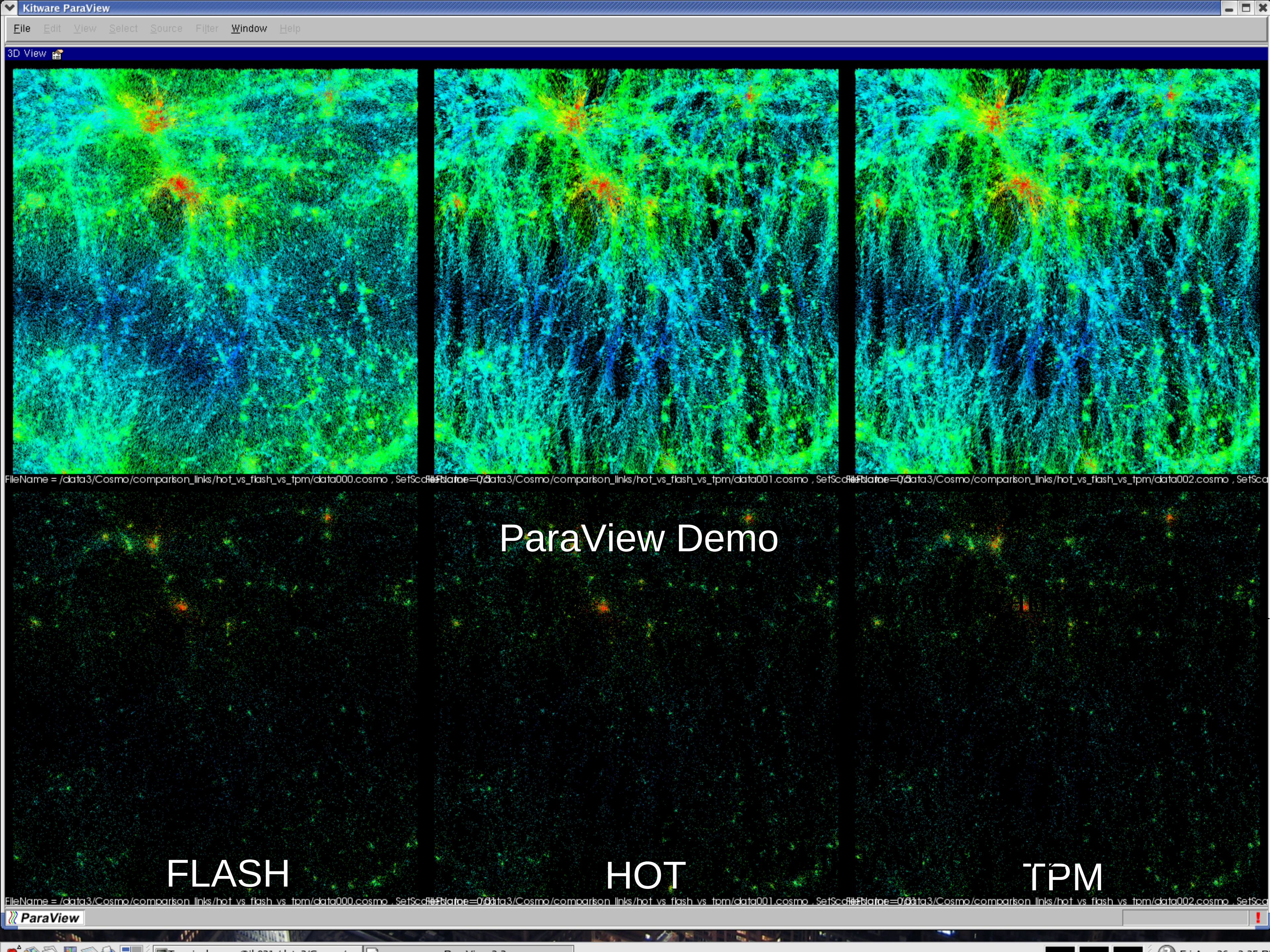
ParaView: Automated Comparative Visualization



- Vary parameters in X and Y
- Create multiple linked visualizations
- Spreadsheet style visual presentation
- Synchronized cameras

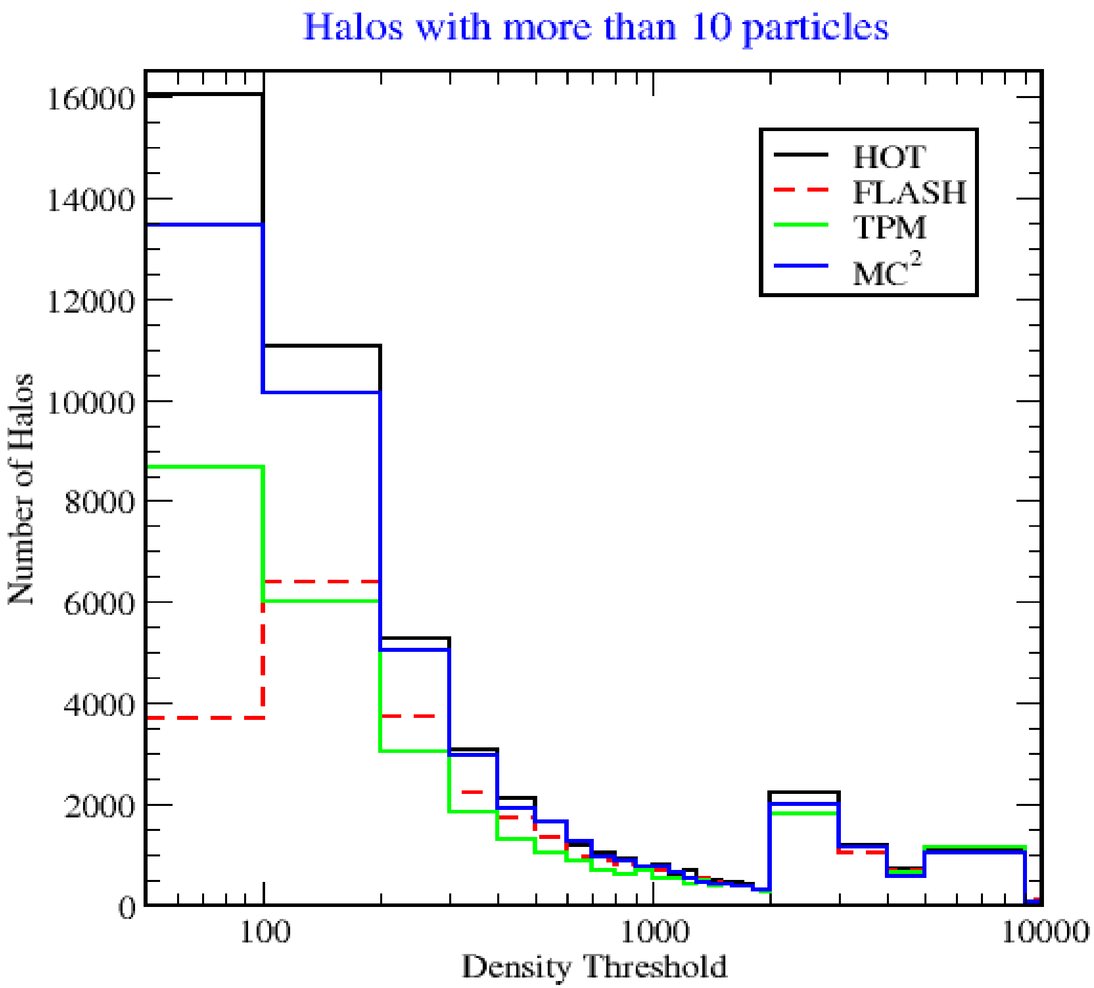


Los Alamos
NATIONAL LABORATORY
EST.1943
The World's Greatest Science Protecting America



Kitware ParaView
File Edit View Select Source Filter Window Help
3D View
FileName = /data3/Cosmo/comparison_links/hot_vs_flash_vs_tpm/data000.cosmo
FileName = /data3/Cosmo/comparison_links/hot_vs_flash_vs_tpm/data001.cosmo
FileName = /data3/Cosmo/comparison_links/hot_vs_flash_vs_tpm/data002.cosmo
ParaView Demo
FLASH
HOT
TPM
ParaView

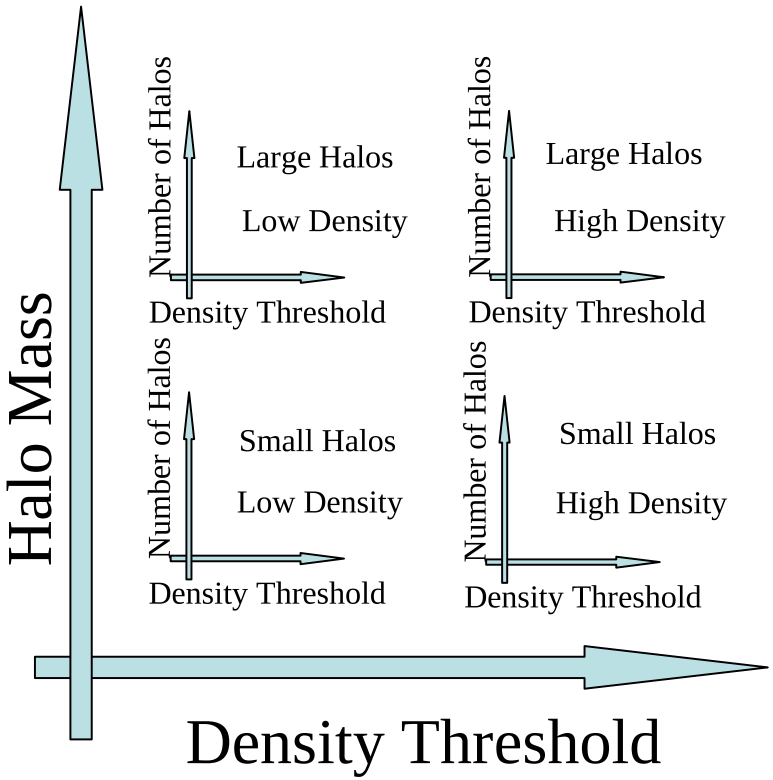
ParaView: Quantitative Results



Note: Bin sizes are not the same in all density regions! This leads to “jumps”, e.g., at 2000.

ParaView Quantitative Results Summary

- FLASH has a severe lack of halos ~40%
- Paraview allows us to identify halos and halo counts in different density regions
  - Qualitative: FLASH loses halos in low density regions
  - Quantitative: confirmed with Paraview (no need for extra analysis codes!)
- Understand the relationship between halo size and density:
  - FLASH has large deficit in low density regions, OK in very high density regions
  - Very small halos live dominantly in low density region



Future: Merging Comparative and Quantitative Visualization Together

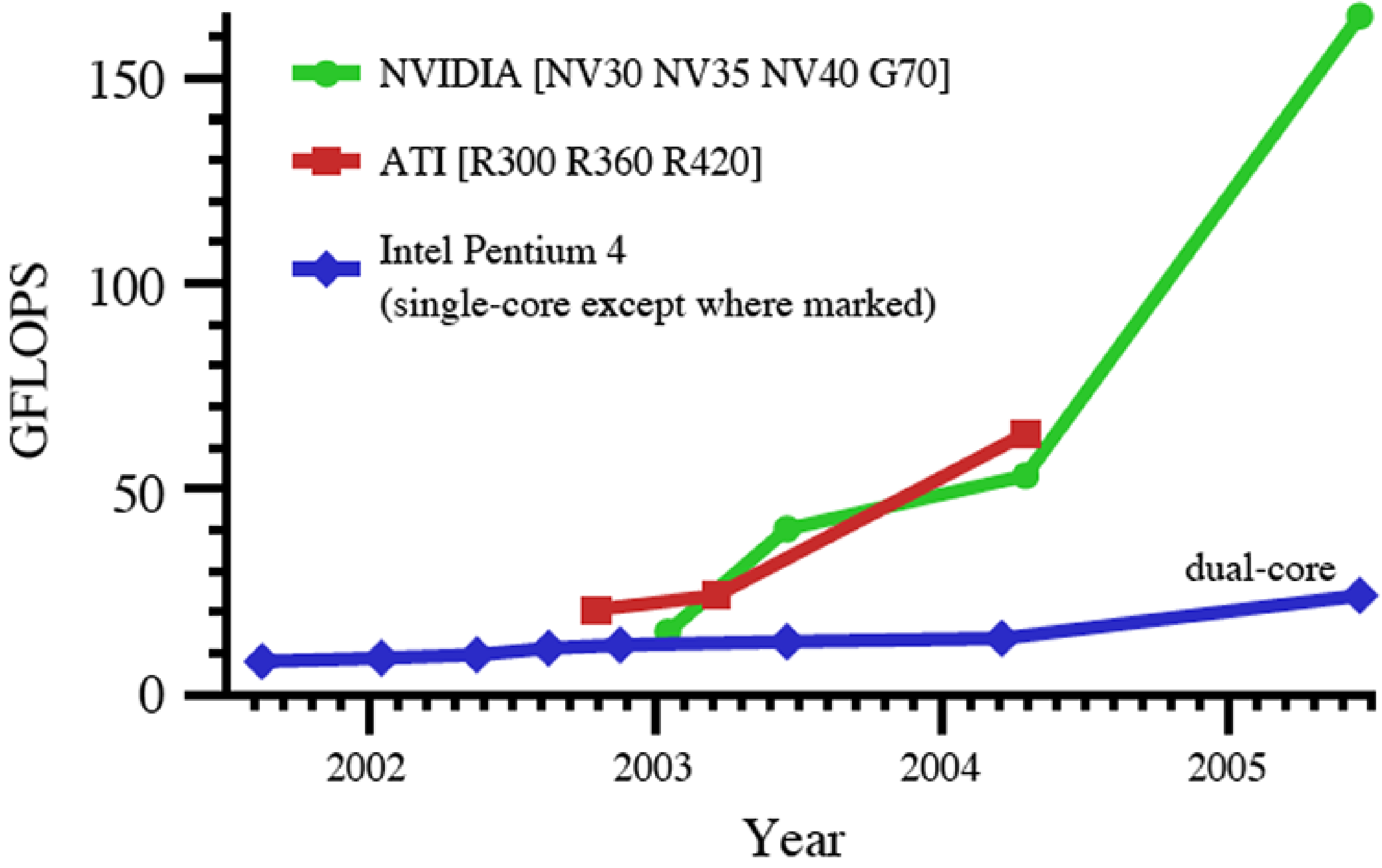
ParaView Quantitative Results Summary

- The current base grid in FLASH allows us to resolve only very large halos (which live in the high density regions)
  - To resolve all halos need a much finer base grid is required
  - Need new force resolution criteria… - refine when appropriate
- Hot topic in cosmology research
  - Study of halo properties and formation as a function of their environment (as defined by density)

Scout Overview

- Patrick McCormick - PI
- High-performance
  - Hardware-acceleration via the multi-core GPU
- Quantitative
  - Define and analyze data via programming language
- Scientist-focused programming language
  - Express both general computations and visualization results
  - Explicit data parallelism
    - Take advantage of data parallel nature of graphics hardware
  - Hide other nuances introduced by graphics API and hardware

Scout: Hardware-acceleration on the GPU



Courtesy of Pat Hanrahan, Ian Buck, and John Owens.

Refined Approach Using Scout

- Qualitative - Visualization
  - Merged as one program
- Quantitative - Analysis
  - Create a program that:
    - Define density, halos
    - Interactively query on halos and density

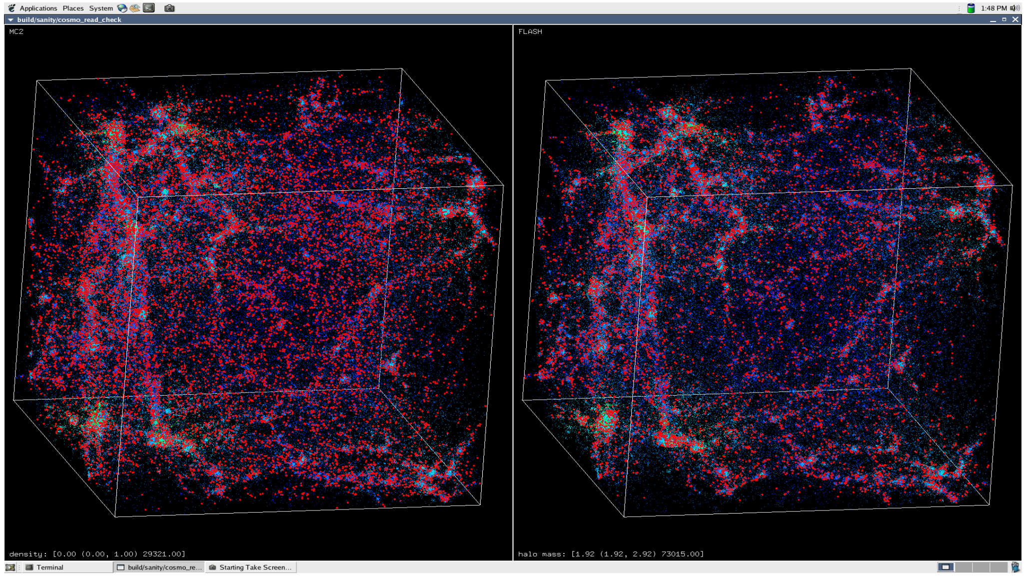
- The Scout Program

```
viewport "MC2" (0.0, 0.0, 0.5, 0.5) {
  float mag(shapeof(mc2_velocity));
  compute with shapeof(mc2_velocity) {
    mag = magnitude(mc2_velocity);
  }
  render points with shapeof(mc2_points) {
    where(density >= … )
      image=hsva(240*(max(mag)-mag) /
          (max(mag)-min(mag)),1,1,1);

      image = null;
  }
  render points with shapeof(mc2_halos) {
    where(mass >= … && density >= …)
      image = rgba(1,0,0,1);
    else
      image = null;
  }
}
```

Scout - Demonstration

- Performance
  - ParaView – halos (~50K) using geometry *(# of visualizations)
  - Scout – halos (~50K), particles (~2 million) using points and queries * (# of visualizations)



Conclusions

- Integrated approach to visualization and analysis
  - Qualitative and quantitative
- Solutions
  - ParaView
    - Open-source large data visualization
    - Comparative visualization
    - www.paraview.org
  - Scout
    - Hardware-accelerated language-based visualization and analysis
    - Contact us - expected binary release end of this year