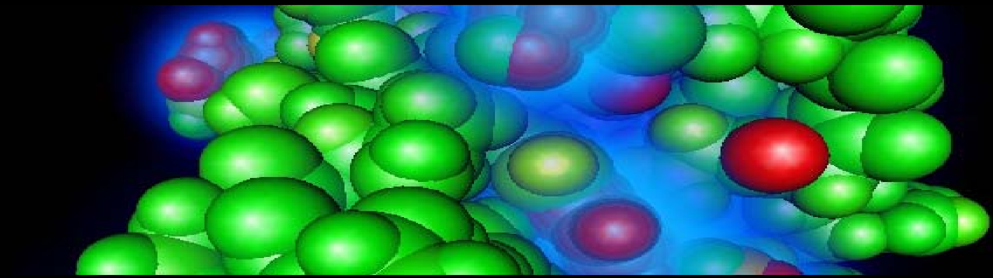


Science-Driven Visualization Research Challenges

9 Nov 2004
SC2004 – Pittsburgh PA

Wes Bethel
with help from Friends at
Lawrence Berkeley National Laboratory
vis.lbl.gov

Outline

- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Outline

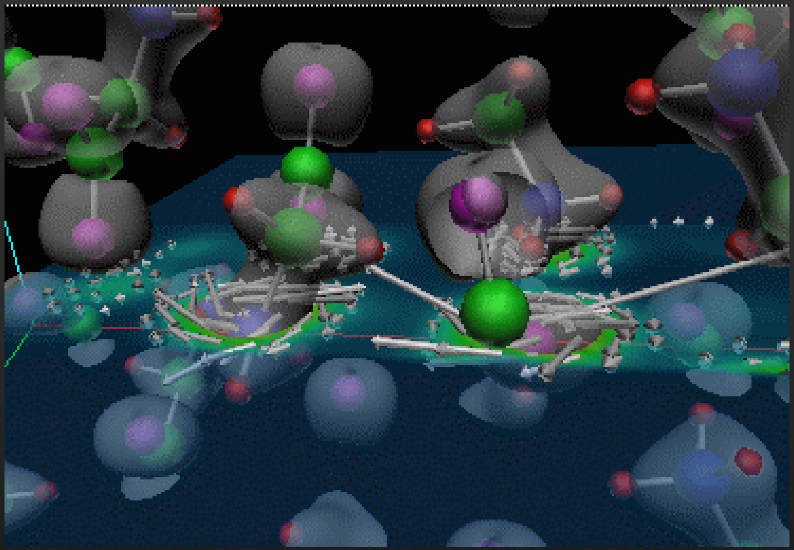
- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization – Introduction and Approach.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Science-Drive Visualization Challenges – Outline

- Role of visualization in science, and what users really want?
- Challenges of user needs.
- What efforts targeted at meeting those needs?
- Is the current approach meeting user needs?

Role of Visualization in Science

- An instrument to "see data" that is otherwise unseeable.
- A vehicle to communicate findings and results.
- Plays an integral part of the scientific process and scientific workflows.



Something doesn't "look right" in this picture – what happened?

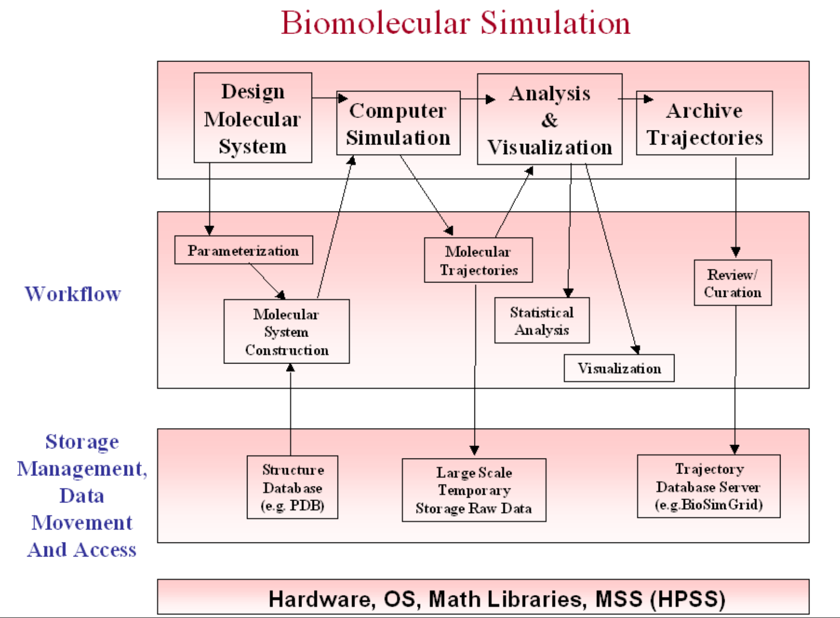
Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Introduction – The Scientific Process and Workflows

- Hypothesize – experiment/test – refine.
- Workflows are the sequence of tasks in the scientific process.
- Visualization serves as the "instrument" to aid in "seeing results" at each stage in the workflow.



Office of Science U.S. DEPARTMENT OF ENERGY

SC2004 PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

What Do (Science) Users Need?

- Easy to use software.
- That is free (and works).
- That is supported.
- Help learning/using/applying the software to their problem.
- New visualization capabilities for their problem.
- Support for remote and distributed operations, capacity to analyze large and complex data.

Challenges of User Needs

- For many modern computational science projects, there exists no “canned” visualization solution. Tools and technology must be created.
- Such efforts require expertise in a wide range of specialties: computer science, software engineering, cognitive science, people skills, etc.
- Creating such tools requires close and ongoing effort between researchers of many disciplines.
- Few, if any, “standards” to help provide a stable environment for visualization.

Science-Drive Visualization Research Problem Statement

- Trend is towards remote and distributed data analysis and visualization.
- Domain-specific solutions required.
- Such solutions are inherently multidisciplinary and extremely complex.

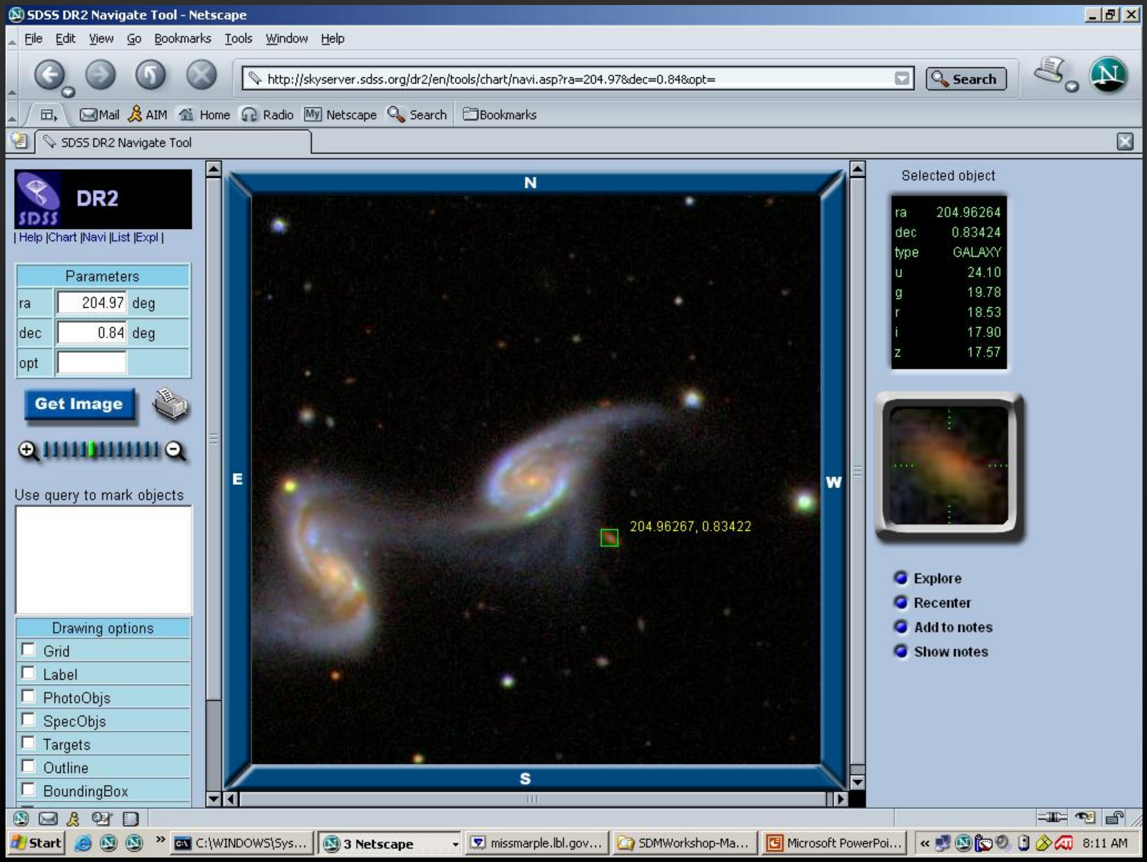
Efforts Targeted at Meeting Science Needs

- **Individual P.I. Funded to perform some visualization research.**
  - A fraction of a P.I. and a graduate student.
  - Publish a research paper, might release a research prototype of their software (or might not).
  - Their reward is the technical publication.
- **Institutional visualization support.**
  - NERSC, ASCI/Views, etc.
- **Missing: large, program-wide coordination of activities.**

Examples of Success

Sloan Digital Sky Survey Portal

- Interface and operations tailored to astronomy community.



Office of Science
U.S. Department of Energy

Does the Current Approach Work?

- **Generally, no:**
  - Duplication of effort across disparate programs.
  - Little impetus to share work, to leverage others work.
- **What’s Missing?**
  - Critical visualization infrastructure: community-centric data models, fungible visualization technology that can be shared and reused across program areas.
  - Program-wide emphasis upon coordinated visualization activities.
  - Requires conscious engineering – coordinated activities will not “emerge” from many small visualization projects.

Outline

- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization – Introduction and Approach.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Office of Science
U.S. DEPARTMENT OF ENERGY

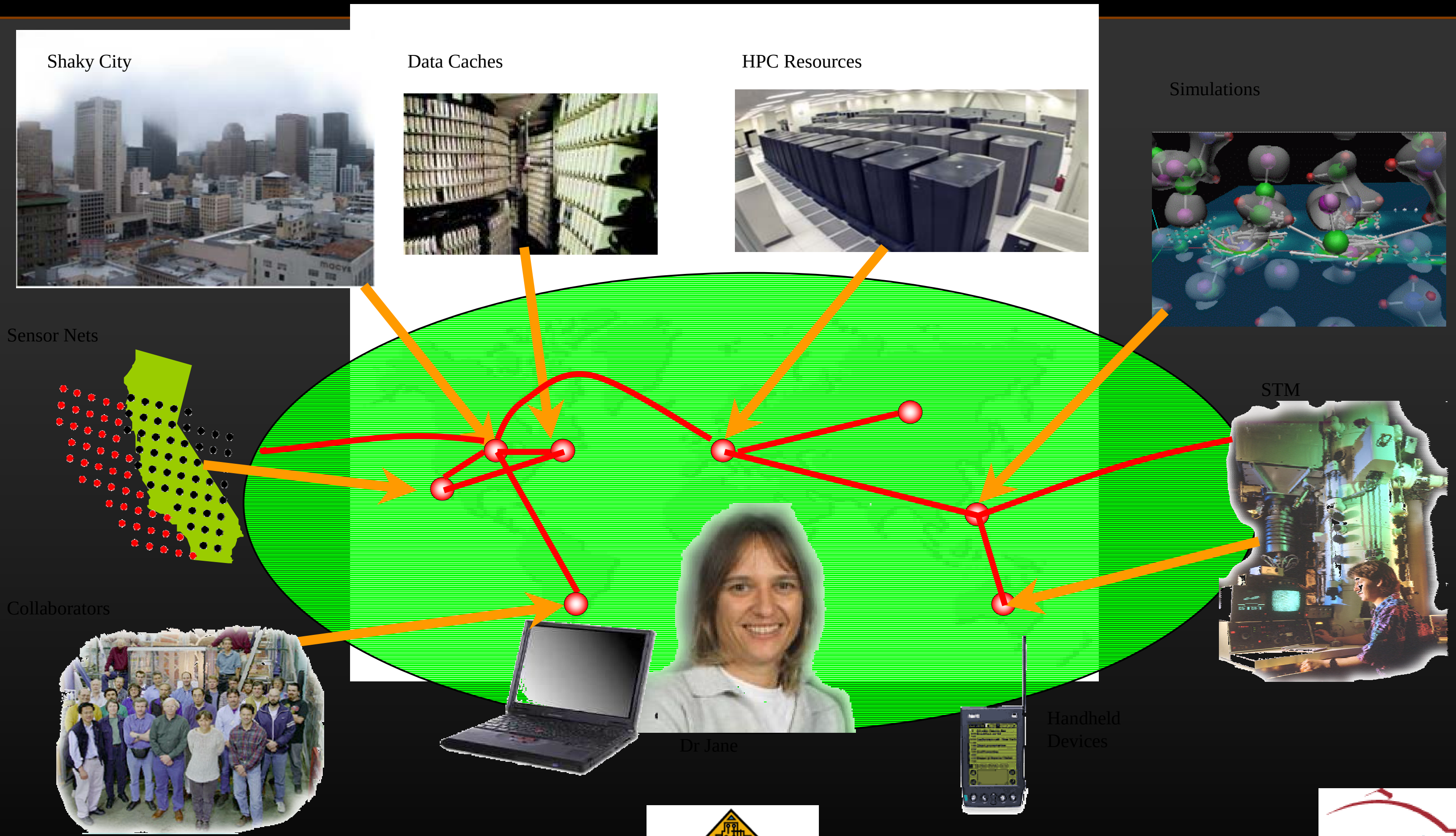
SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

LBNL Visualization Research – Outline

- **The LBNL Visualization Research Vision.**
- **The Research Strategy and Tactics.**
- **Near-term and long-term goals.**
- **Results:**
  - Domain-specific solutions.
  - Remote and Distributed visualization research results.
  - Computer Science Research.

LBNL Visualization Research Vision



Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Problem Statement – Repeated

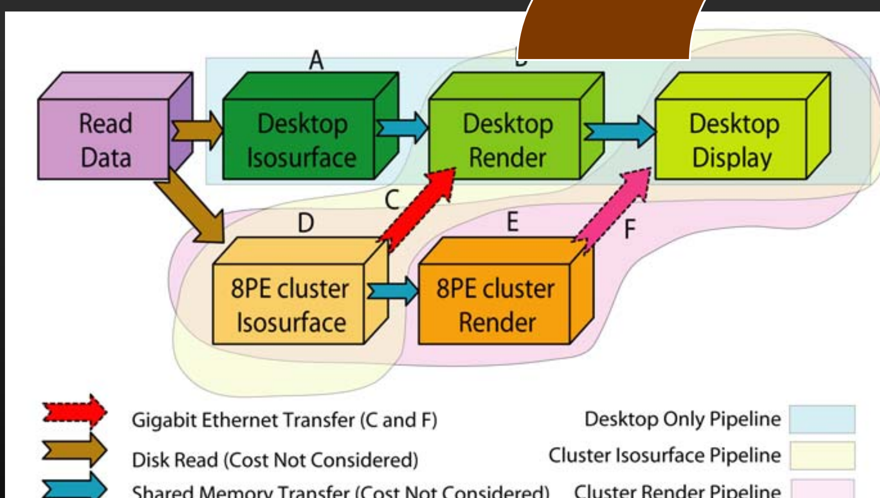
- Trend is towards remote and distributed data analysis and visualization.
- Domain-specific solutions required.
- Such solutions are inherently multidisciplinary and extremely complex.

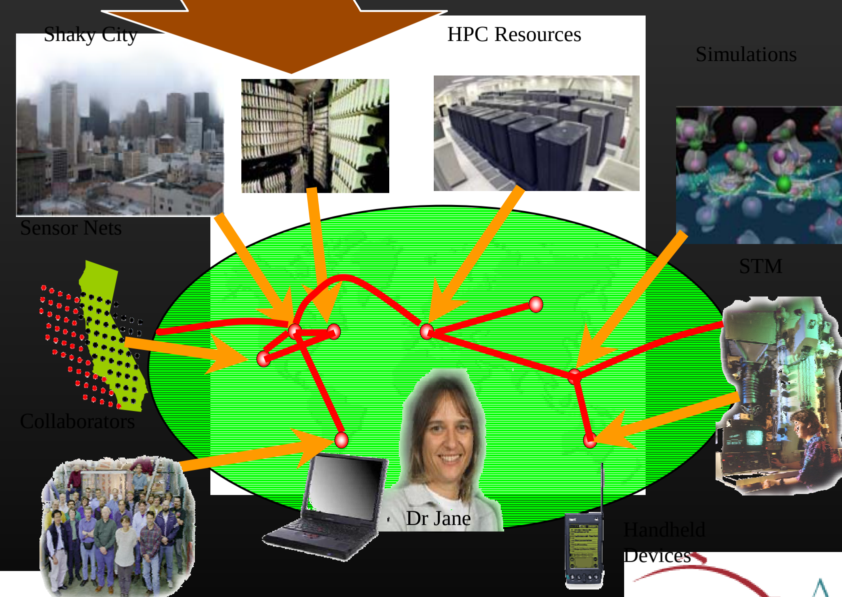
Research Challenges for Remote and Distributed Visualization

- Community-centric data models, component interfaces, execution frameworks.
- Visualization algorithms, delivery mechanisms.
- Effective and simplified use of parallel and distributed resources.

LBNL Visualization Research Strategy

- **Map the canonical visualization pipeline into remote & distributed use model.**





LBNL Visualization Research Tactics

- Close relationships with DOE science projects to deliver domain-specific (useful) technologies.
- Research advances on the visualization pipeline to realize the dream of "vis anywhere, anytime, by anybody."
- Fundamental CS research to complement visualization research.

LBNL Visualization Research Tactics

- Components encapsulate algorithms, frameworks marshal data and mediate execution (see HECRTF).
- Bottom-up: focus on specific application-driven projects. E.g., Accelerator SciDAC.

LBNL Visualization Research Tactics

- Distributed and parallel architectures offer new algorithmic opportunities (Visapult).
- Interaction methodology important for large data exploration, cuts across data management, visualization, applications.
- Delivery mechanisms are “the handles” provided to the user to guide data exploration and analysis.

Office of Science U.S. DEPARTMENT OF ENERGY

SC2004 PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Outline

- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization – Introduction and Approach.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Domain-Specific Solutions

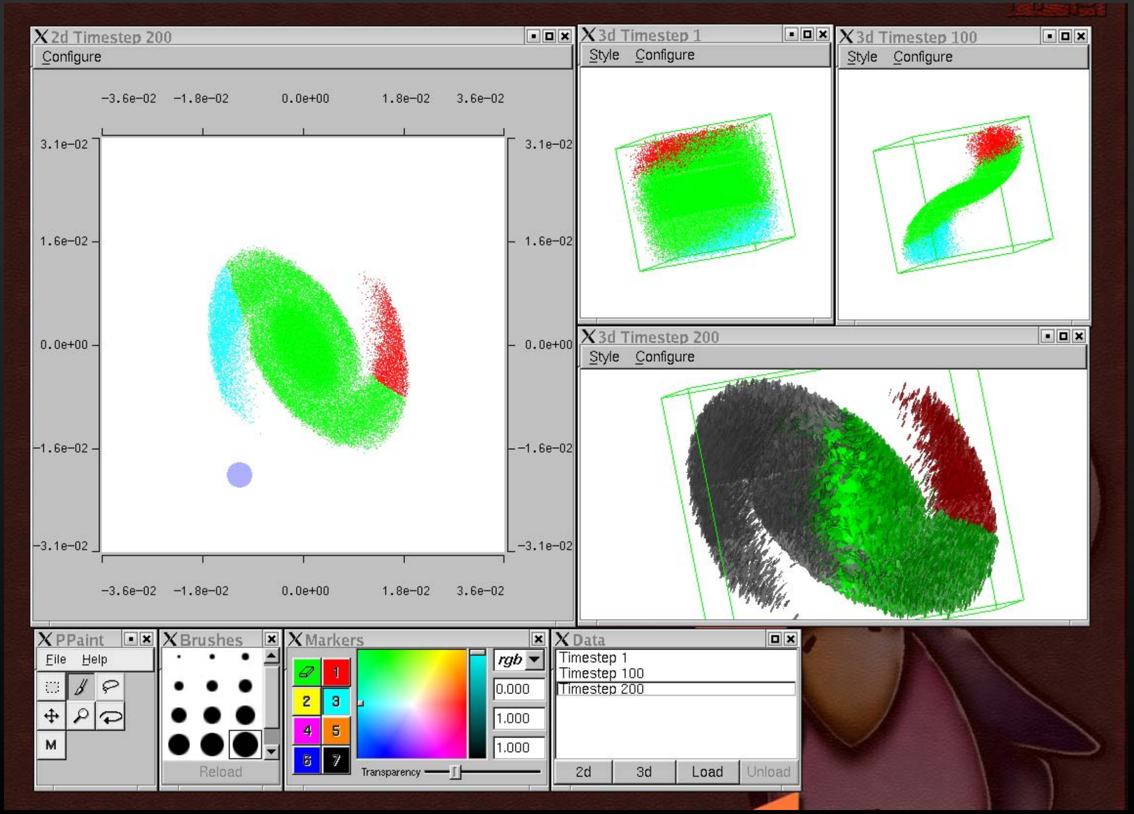
- 21st Century Accelerator Modeling (SciDAC)
- Center for Extended MHD (SciDAC)
- Protein Structure Prediction

Accelerator Simulation Visualization

- Data: time-varying, 6D, multi-species.
- Typical visualization: scatterplots of one dimension against another. E.g., x-position vs. x-phase.
- Need: ability to explore, to subset, to visually comprehend science.

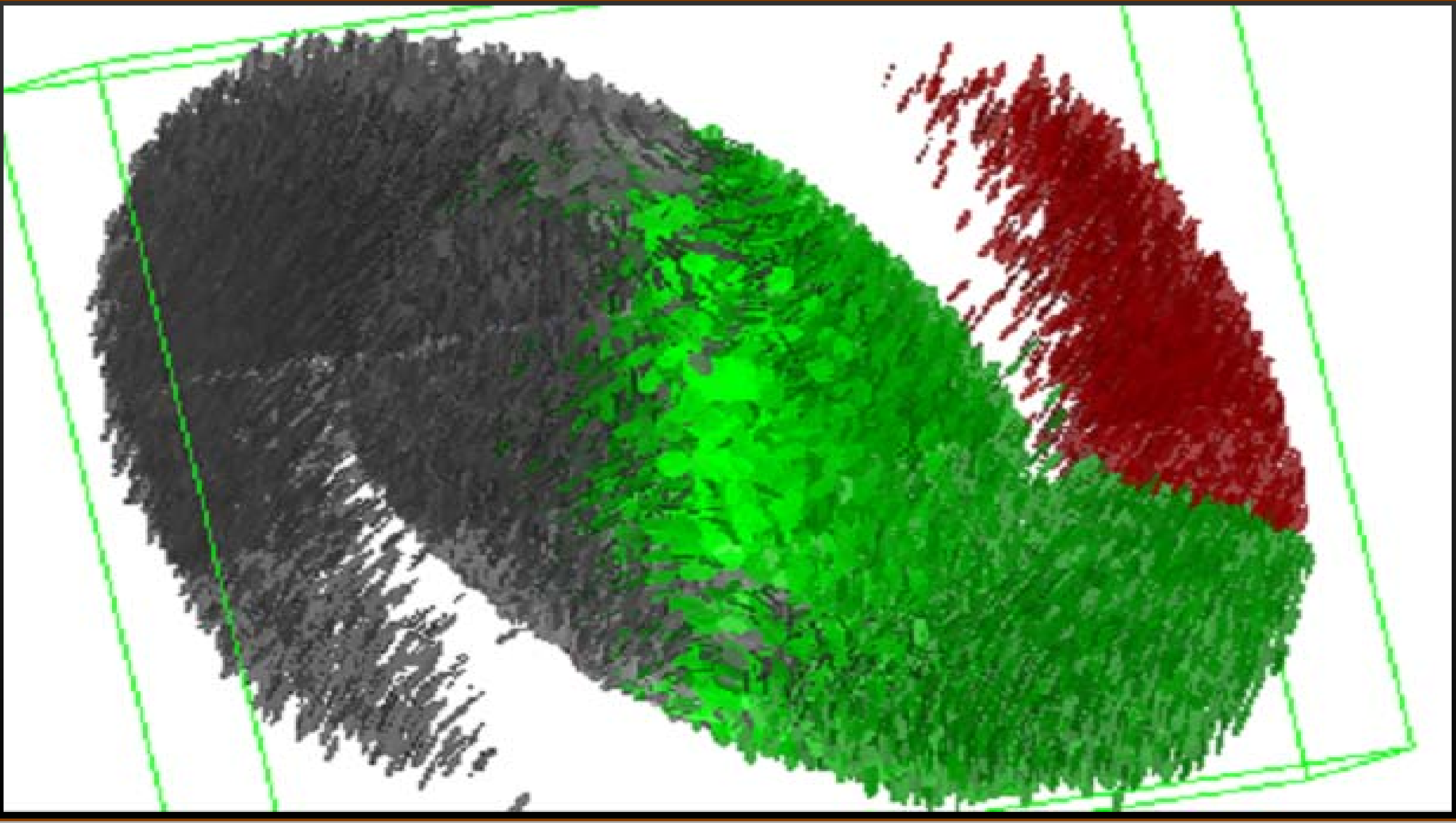
Accelerator Simulation Visualization, ctd.

- **Interactive data subsetting and selection.**
  - Paint metaphor
  - Using domain knowledge.
- **Novel visualization technique well-suited for 6D data (next slide).**

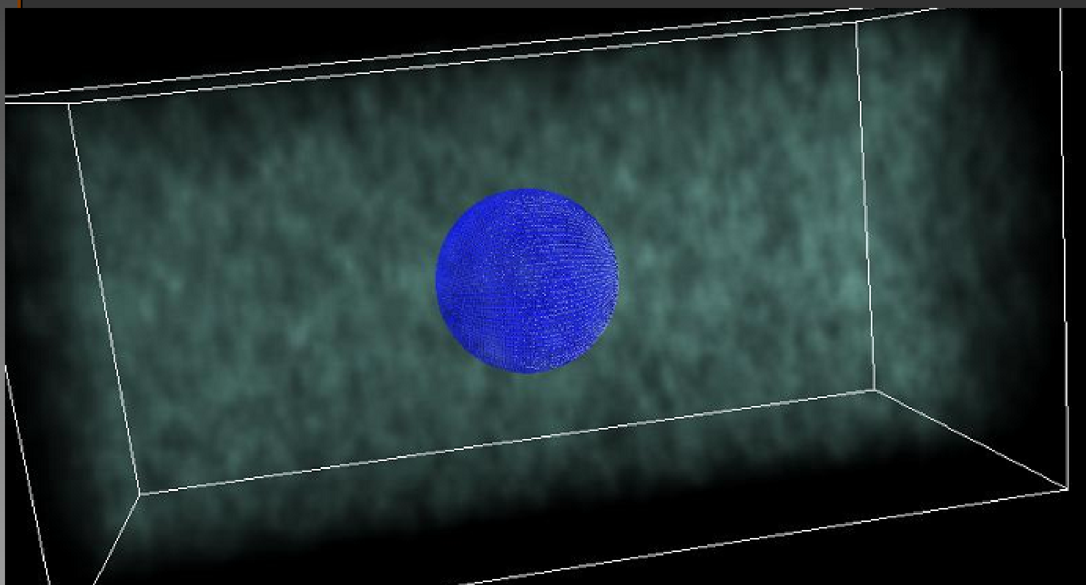


Office of Science
U.S. DEPARTMENT OF ENERGY

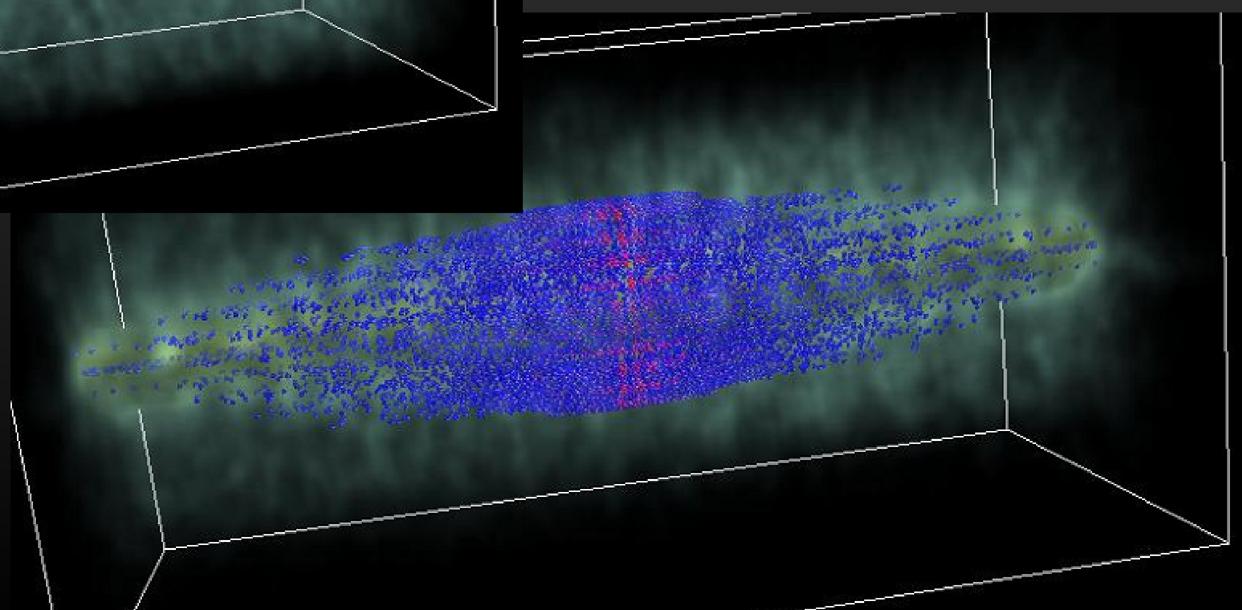
Accelerator Simulation Visualization, ctd.



Accelerator Simulation Visualization, ctd.

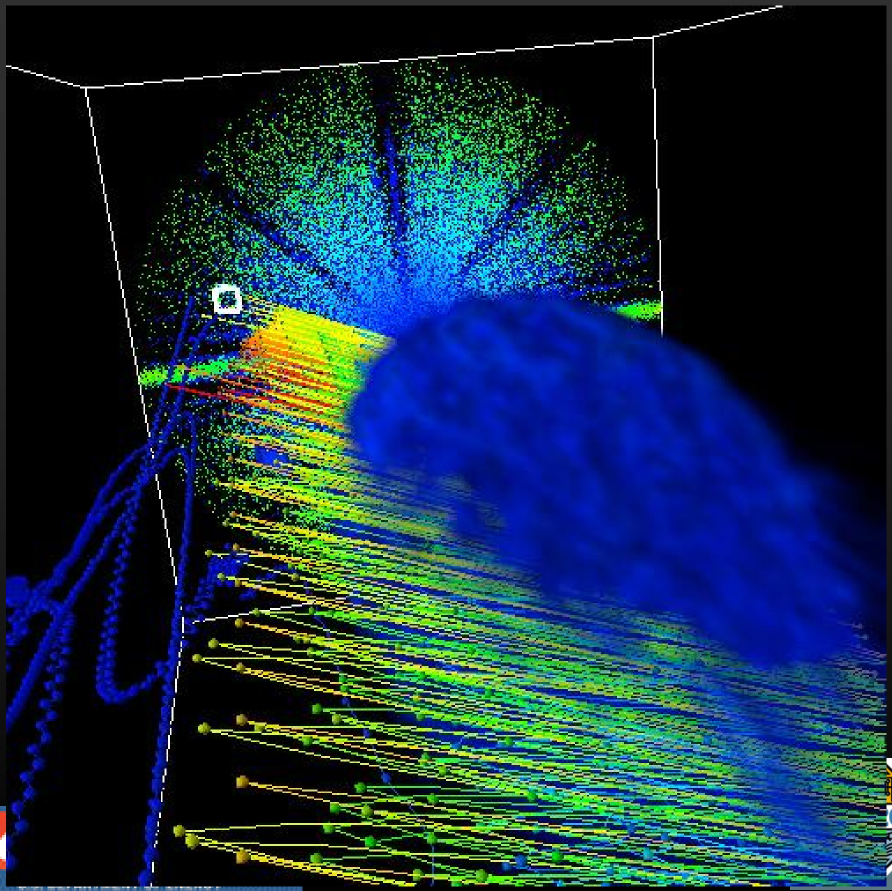


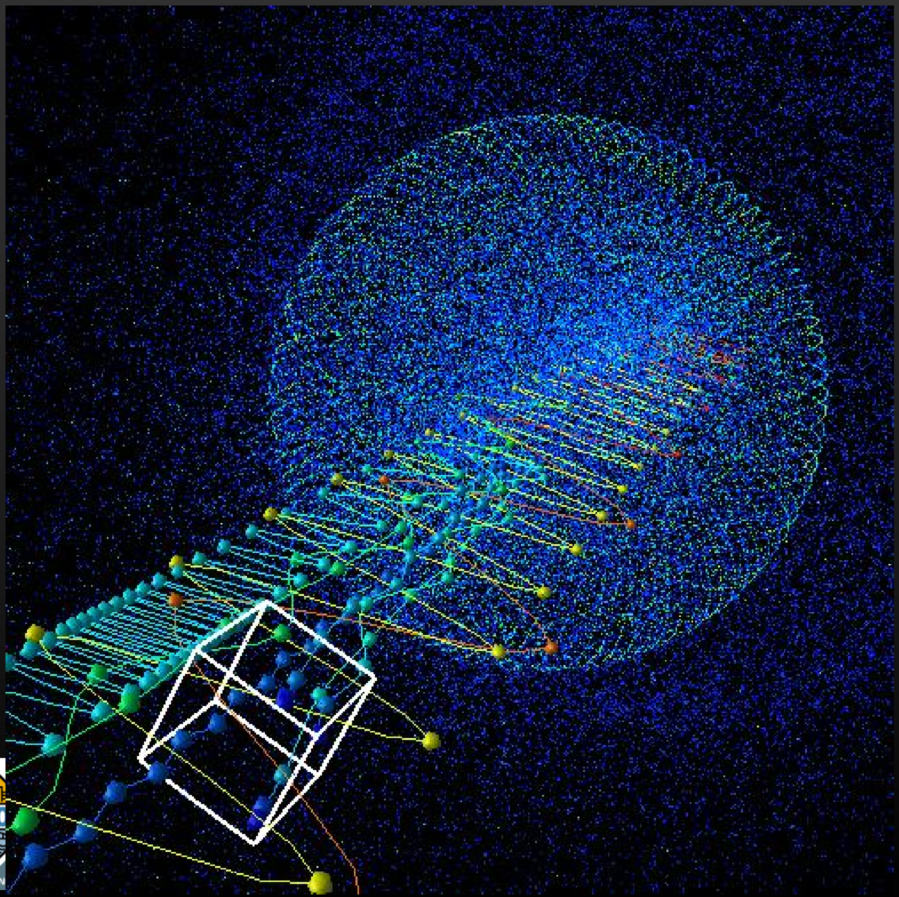
Proton beam (particles) passing through a cloud of electrons (volume rendering).



Accelerator Simulation Visualization, ctd.

Electron trajectories

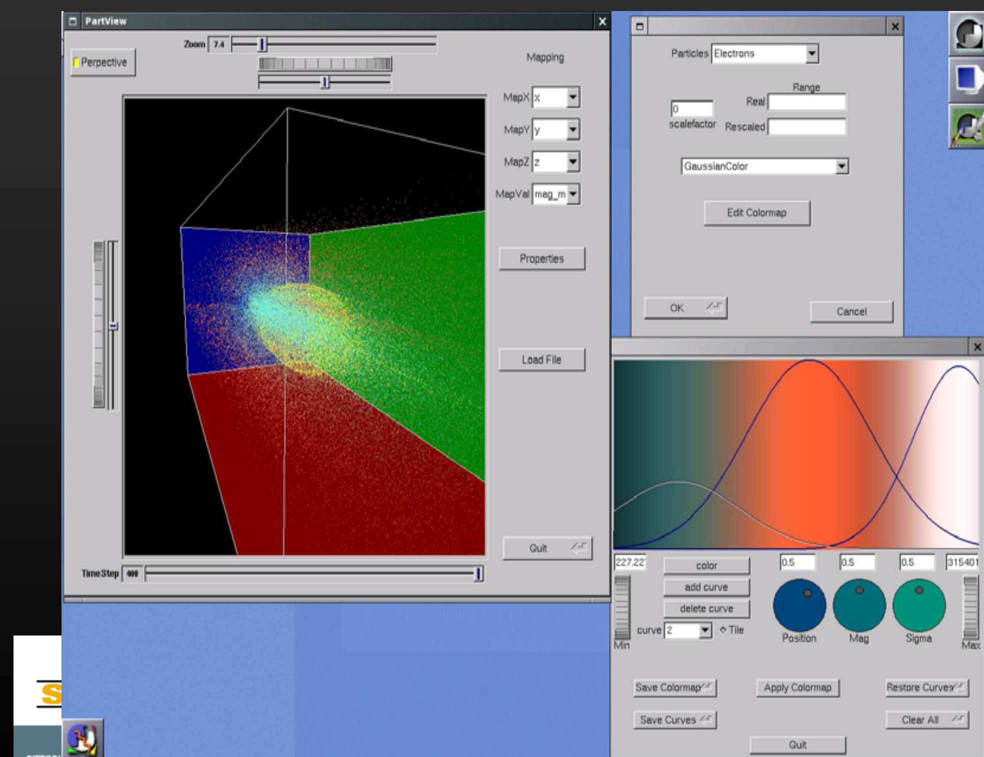




Accelerator Modeling: Remote and High Performance Visual Analysis

- **User-requested domain-specific tool for browsing data.**
- **Distributed, pipelined architecture to scale with increasing data sizes.**

Remote data storage ⟷ workstation



Office of Science
U.S. DEPARTMENT OF ENERGY

Accelerator Modeling: Remote and High Performance Visual Analysis

- **Our group engineered a HDF5 file format for the computational scientists.**
  - They were using ASCII files.
- **Our group also engineered parallel I/O capabilities using HDF5.**
- **A common data model/format is the basis for a family of high performance analysis software technology.**

Accelerator Modeling Visualization: Conclusion

- Close interaction with scientists resulted in domain-specific technologies as well as new visualization research.
- The “unglamorous work” of data models/formats and I/O is the underpinning for the much of the project.
- We are in a good position to move forward with additional tools based upon a community-centric data model.

Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Remote Visualization of Fusion Simulation Results

Problems:

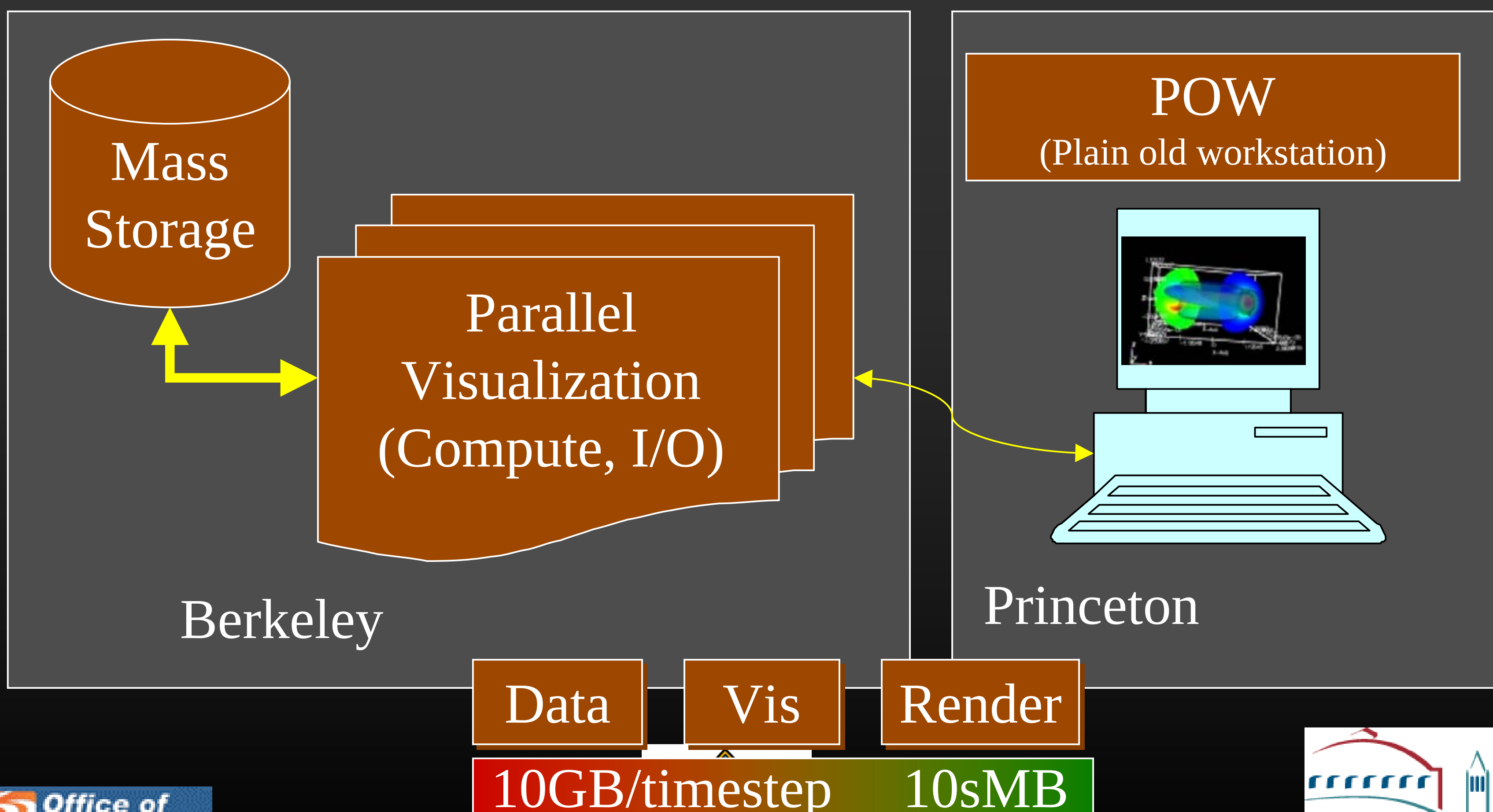
- Simulations run at centralized supercomputing facilities generate large, complex data.
- Analysis to be performed by remotely located scientists.
- Science teams are themselves geographically distributed, and have requested some form of collaborative investigation/visualization.

Remote Visualization of Fusion Simulation Results

Approach:

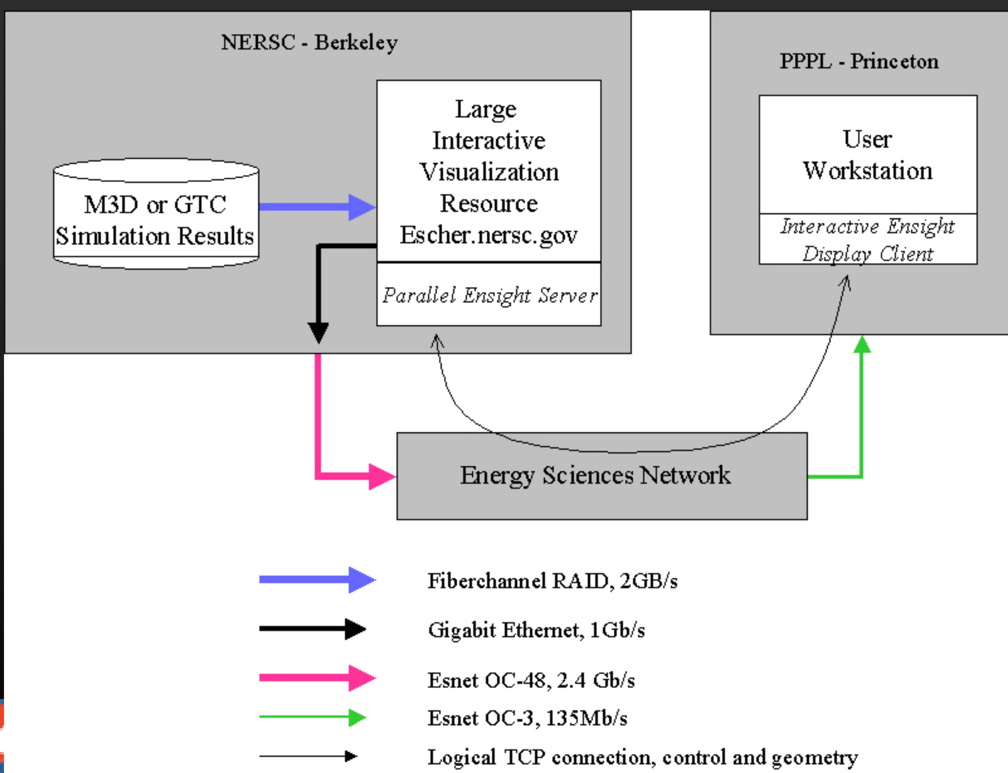
- Use high performance, parallel resources located “close to” the data.
- Where plausible, retain the high performance rendering capabilities of desktop workstations.
- Partition the visualization pipeline (more later) across sites in multiple ways. Which works best?

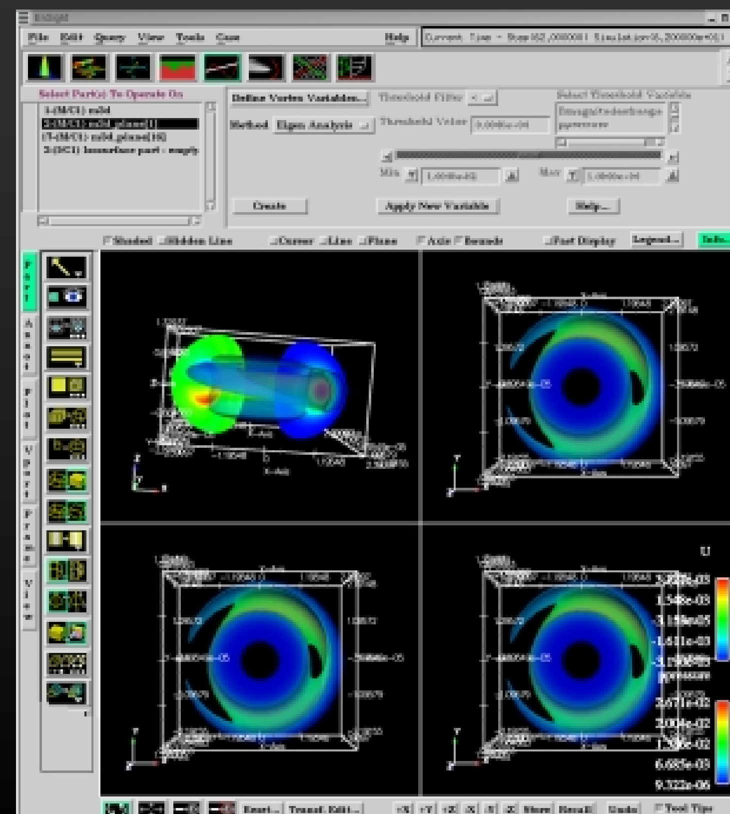
Fusion Visualization: Pipelined, Distributed and Parallel Architecture



Fusion Visualization: Pipelined, Distributed and Parallel Architecture

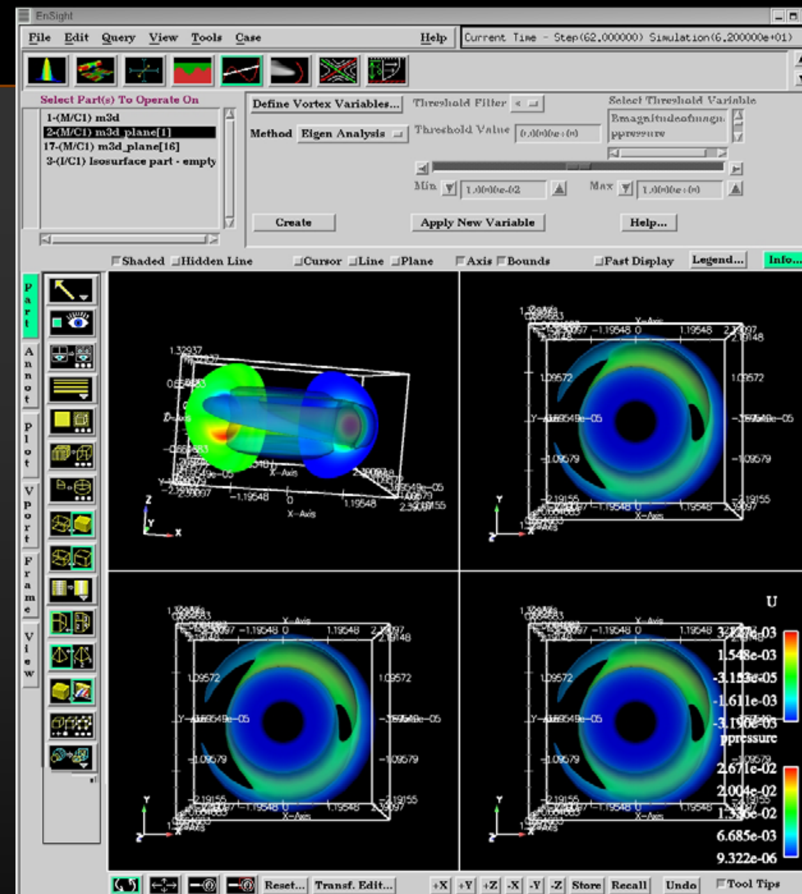
- High capacity I/O and compute located “near” large data source.

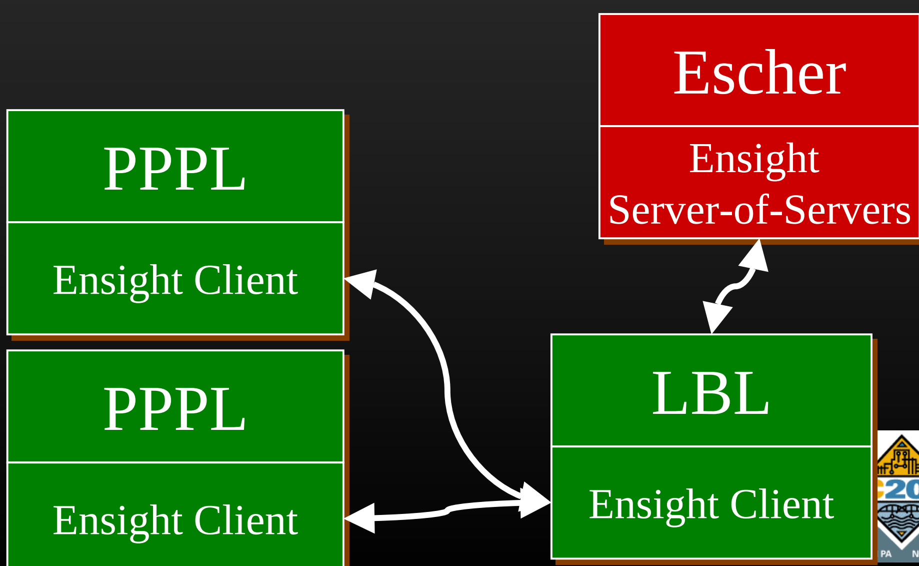




Collaborative Visualization

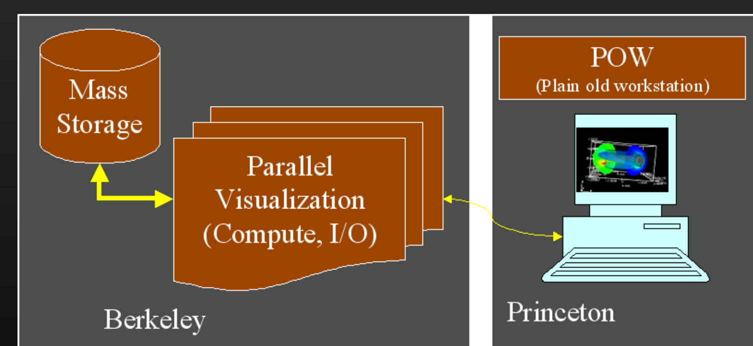
- Rapid inspection of data too large to move:
- Saves having to transfer 100s of GB across country.
- Multiple simultaneous participants (roundtable model).





Remote Fusion Simulation Visualization – Sending Images

- **~50fps 800x600, 24-bpp**
  - Over 100BaseT, low latency connection (LAN)
  - Freely running image generator – only framebuffer contents sent; no mouse events, etc.
  - Frame rate relatively insensitive to compression algorithm, as long as some compression is used.
- **4-5fps "full screen interactive application"**
  - 100BaseT Ethernet, 50ms latency (WAN between LBNL – PPPL)
  - Interactive application.

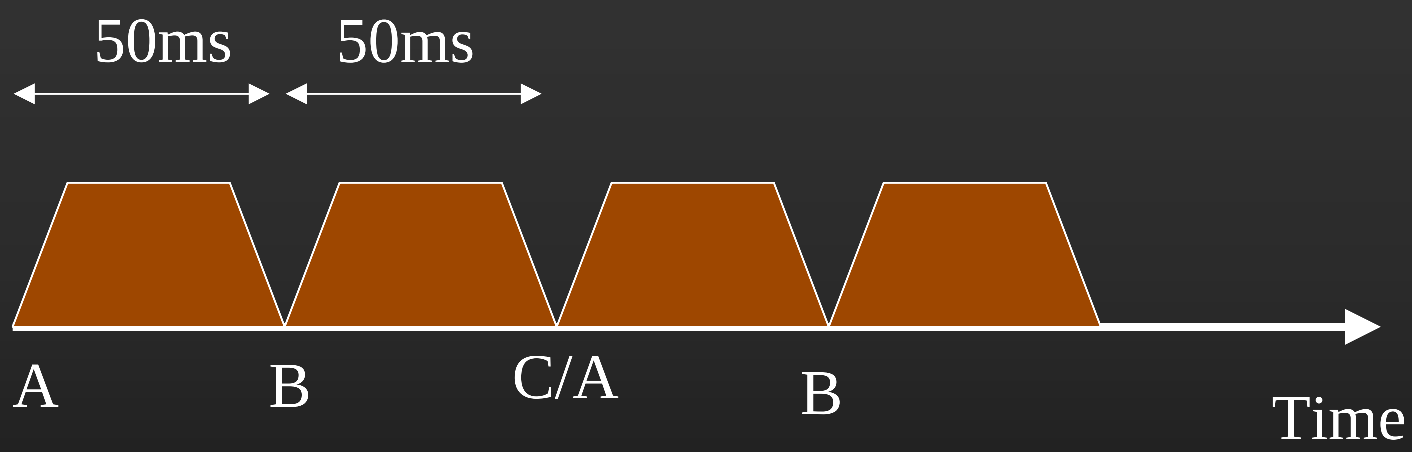


4-5fps “not unexpected”

- 50 ms one-way latency is 100ms RTT
- Maximum possible framerate: 10fps
- Add in more latency due to fb reads, detect and package mouse events, etc.

- Conclusion: latency is a killer.

Frame Rate Limit Due to Latency: 1000/2*latencyMS.



A – user drags the mouse, mouse event sent to server.

B – "instantaneous" frame render, grab, compress, send and receipt by client.

C – client decompresses, displays image, grabs next mouse event, etc.

Fusion Visualization: Conclusions

- Using high capacity visualization resources located "close to" the source data for remote use appears promising.
- Different approaches, each with advantages and disadvantages.
- Functional results: good.
- Performance results: mixed.

Protein Structure Prediction Outline

- **Problem Description.**
- **Approaches to help solve an NP-hard problem:**
  - Better initial configurations.
  - Visualization and intervention to guide optimizations.

Protein Structure Prediction

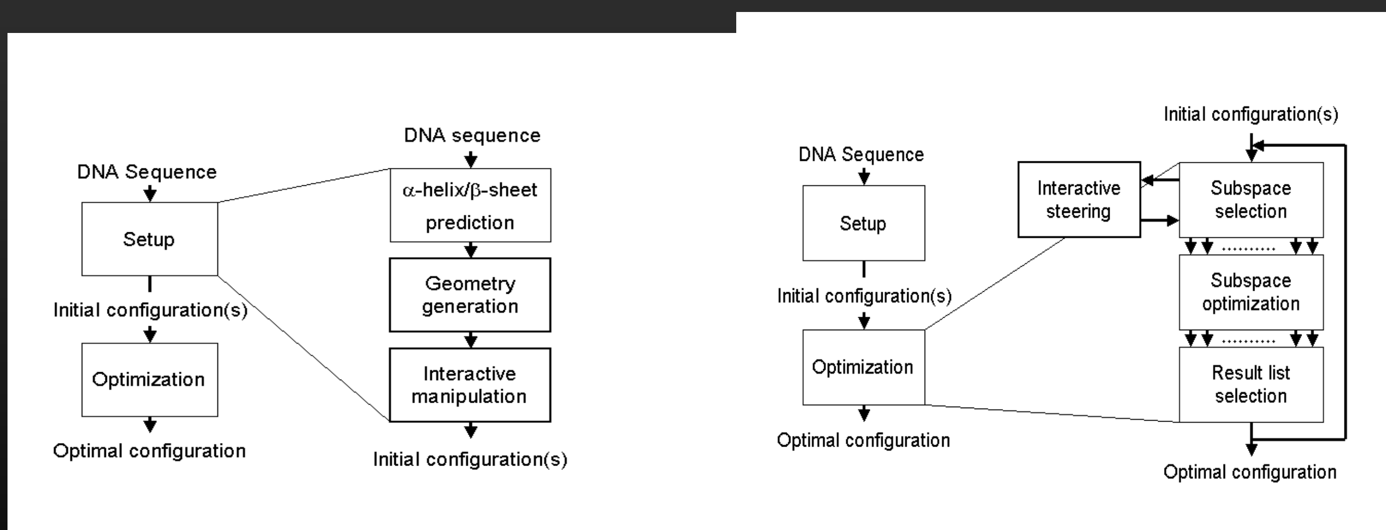
Challenges

- Protein structure prediction is difficult (NP-hard) – it is one of the grand challenges in computational biology.
- Visualization and interactive techniques can accelerate the process.
- No “off-the-shelf” technologies exist – they must be created.

Protein Structure Prediction, ctd.

Given: an amino acid sequence,

Find: an optimal protein conformation.



Protein Structure Prediction, ctd.

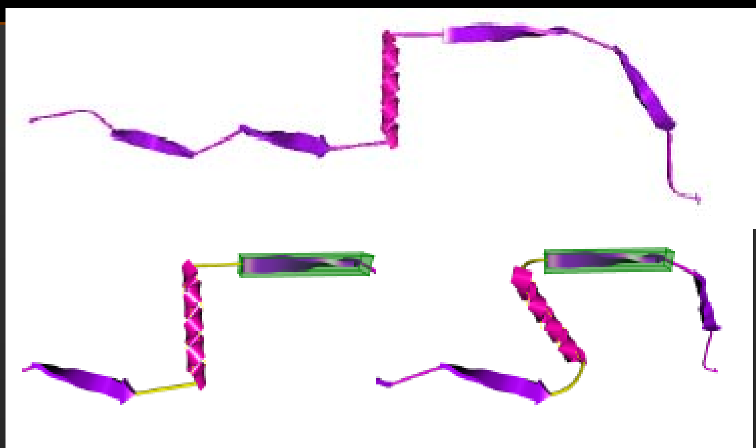
Problem: what is the minimal-energy structure of a sequence of amino acids?

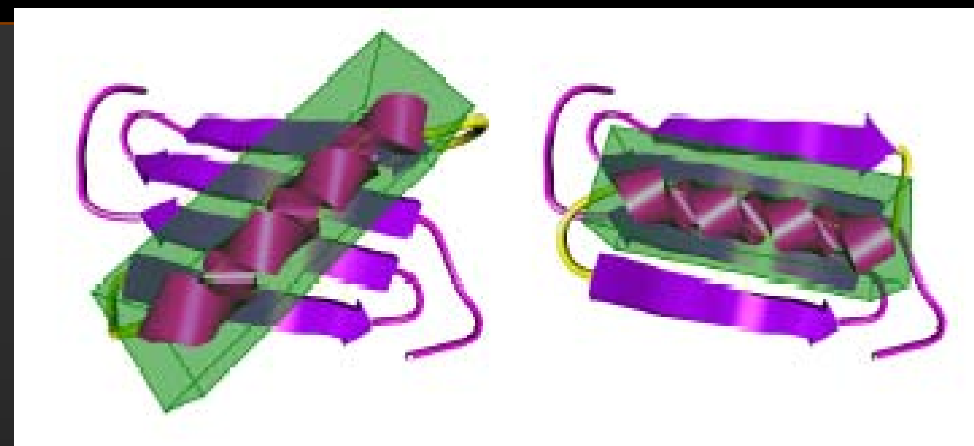
Solution: Nature knows, but computing an answer is NP-hard (not solvable).

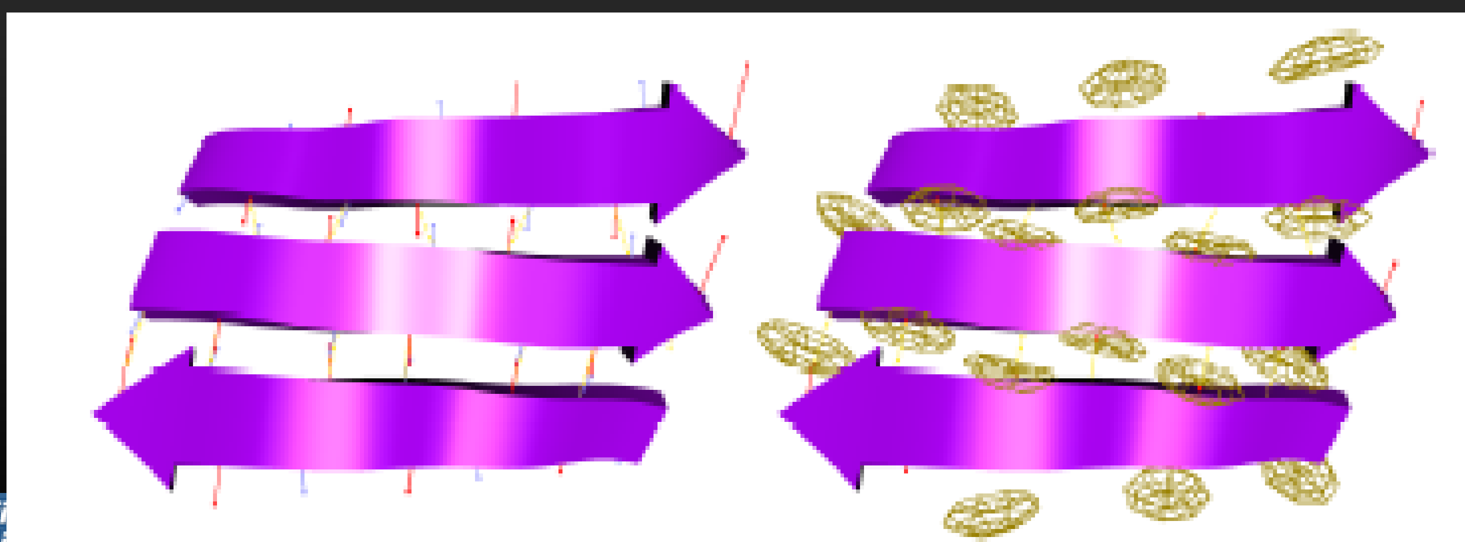
Approach: Human-guided setup, computer-aided energy optimization and minimization.

```
Conf: 999999999999999999999999999999999999
Pred: HHHHHHHCCCEEEEEEEECCCEEEEEEEEECCCCCCC
AA:   FKQYANDNGVDGVWTYDDATKTFTVTEMVTEVPVA
```

Protein Structure Prediction, ctd.

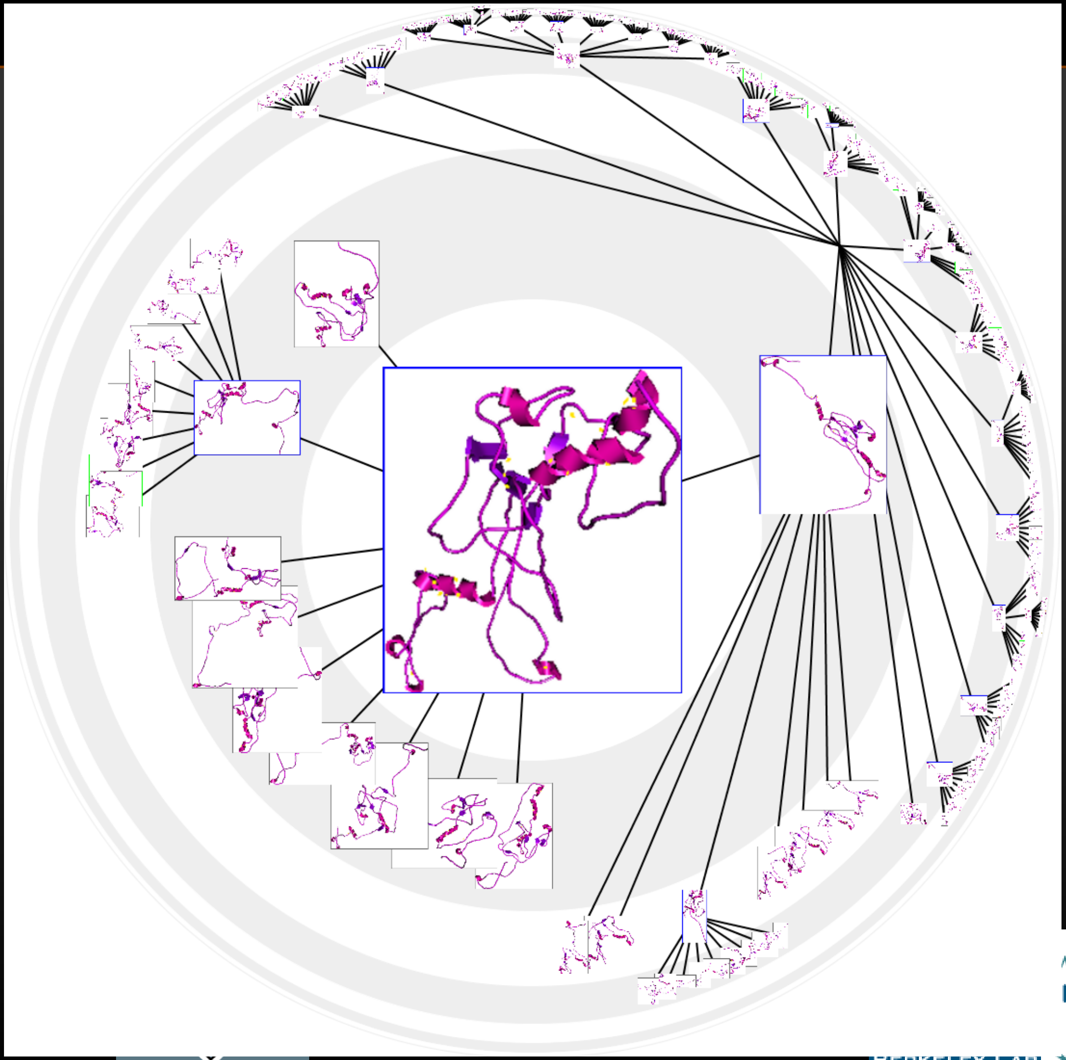






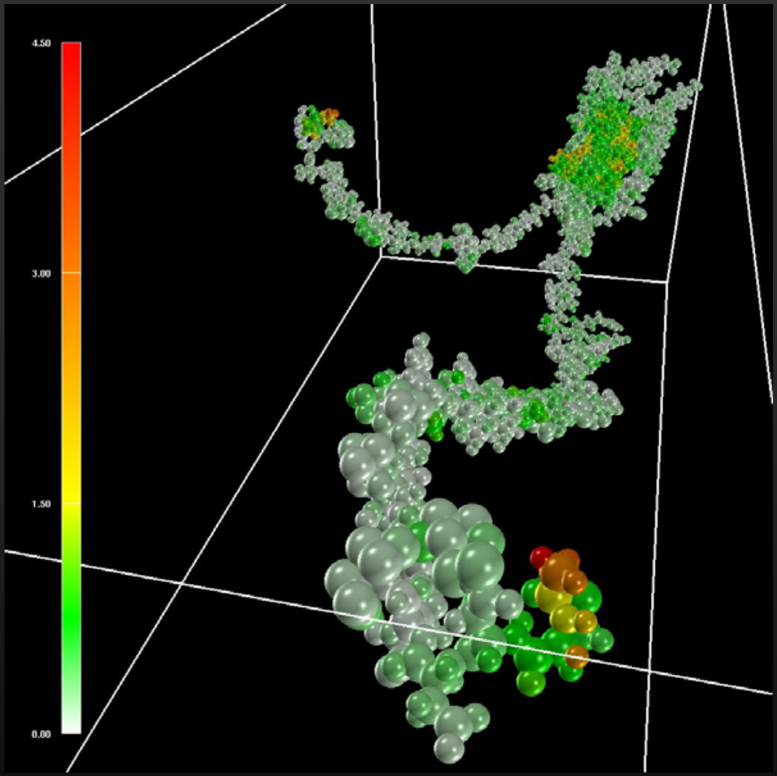
Protein Structure Prediction, ctd.

- Optimization and computational steering
- Initial configurations used as “seed points” for optimization.
- Intermediate results – the “search tree” – is displayed for inspection.
- A human may intervene in the optimization.



Protein Structure Prediction – Energy Visualization

- Energy gradient
- (Movie)

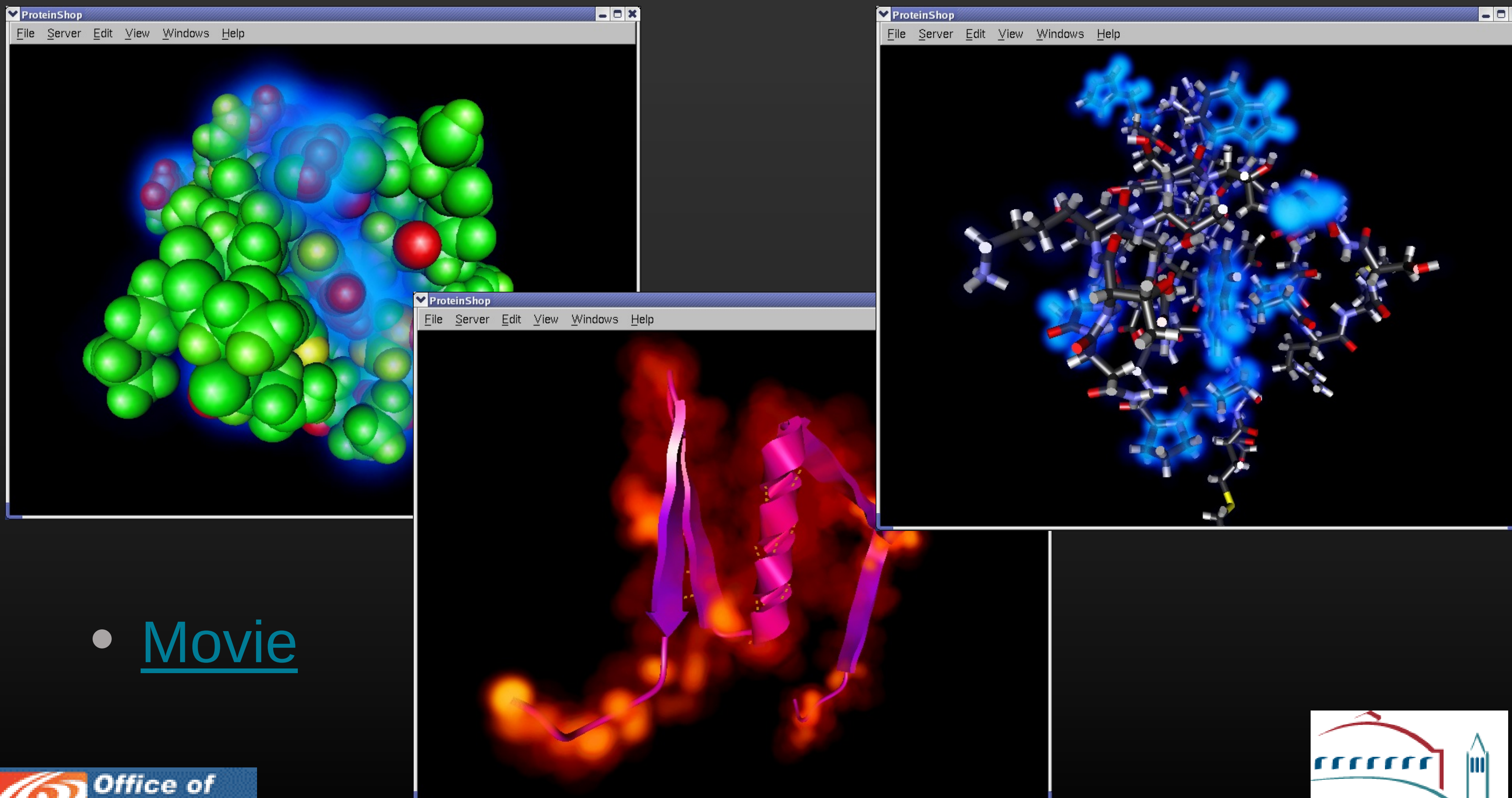


Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Protein Structure Prediction – Energy Visualization



- Movie

Protein Structure Prediction – Conclusion

- **Increased scientific capacity and capability.**
  - CASP4 2000 – days; CASP6 2004 – hours.
- **New scientific opportunities:**
  - Multiple molecule interactions – drug design.
- **Visualization impact:**
  - Best Application Paper award, IEEE Visualization 2003.

Outline

- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization – Introduction and Approach.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Office of Science U.S. Department of Energy

SC2004 Pittsburgh PA November 6 - 12

Berkeley Lab

Computer Science/Visualization Research - Outline

- Research Challenges.
- Query-based visualization.
- Desktop delivery R&D.
- Remote and distributed visualization pipeline optimization.

Fundamental Remote and Distributed Visualization Research Challenges

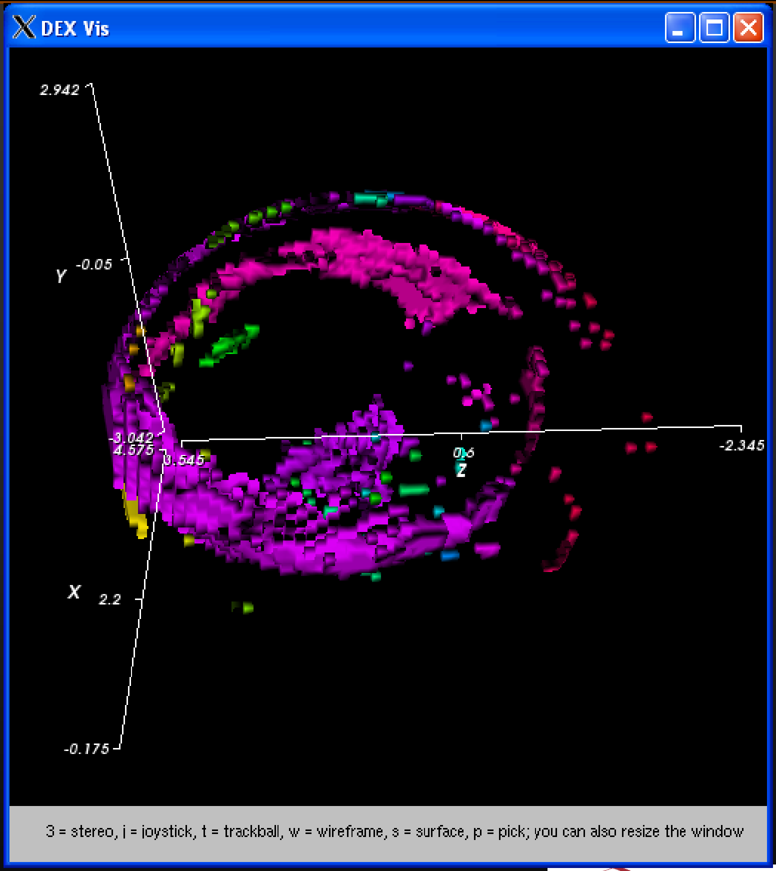
- **Fungible technologies for creating visualization applications.**
  - Components, data/application adapters, vis-centric network transport, resource discovery/allocation, dynamic application construction, decoupling UI from vis/analysis “engine,” decoupling execution control from component architecture.
- **Community-centric data models.**
- **Multiresolution and progressive analysis/vis.**

Fundamental Remote and Distributed Visualization Research Challenges, ctd.

- More interactions with other communities: science applications, data management and data analysis.
- Long-term deployment and maintenance strategy.
- Community and programmatic focus on technology interoperability.

Query-Driven Visualization (Dex)

- Combine Visualization with SDM technology to accelerate visualization and analysis.
- Select data based upon boolean queries.
- Only visualize/analyze data that meets query criteria.



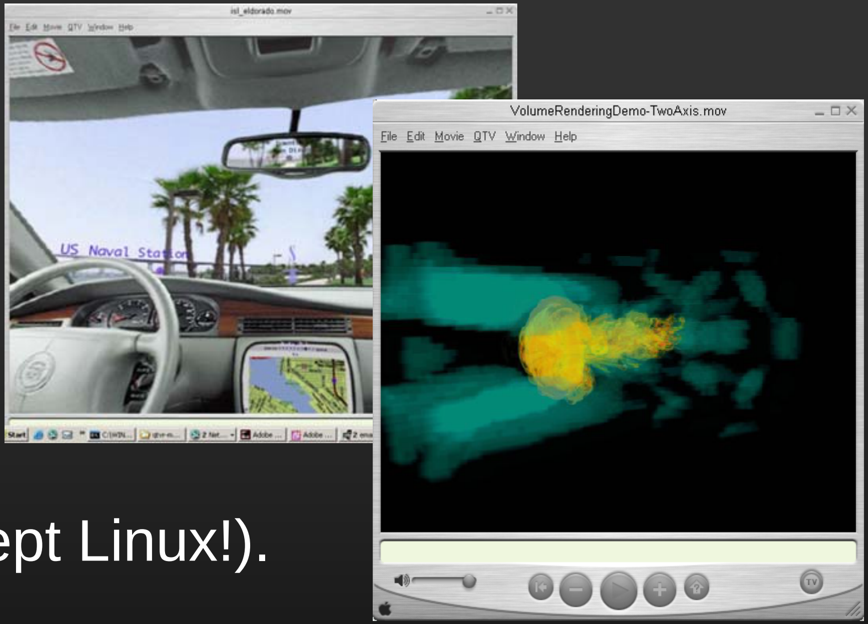
Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

Remote Desktop Delivery – Thin Client

- **QuickTime VR**
  - Panorama Movies
  - Object Movies
    - Two axis, time-varying.
- **QTVR:**
  - Industry standard
  - Freely available players (except Linux!).
- **LBNL Contribution**
  - Object-movie encoder.
  - Current research – multiresolution-capable.



Visualization Pipeline Optimization

- Context: many heterogenous, distributed resources.
- Goal: user wants to take advantage of distributed resources to solve a problem.
- Problem(s): need to select a set of resources to meet the task at hand.

Visualization Pipeline Optimization

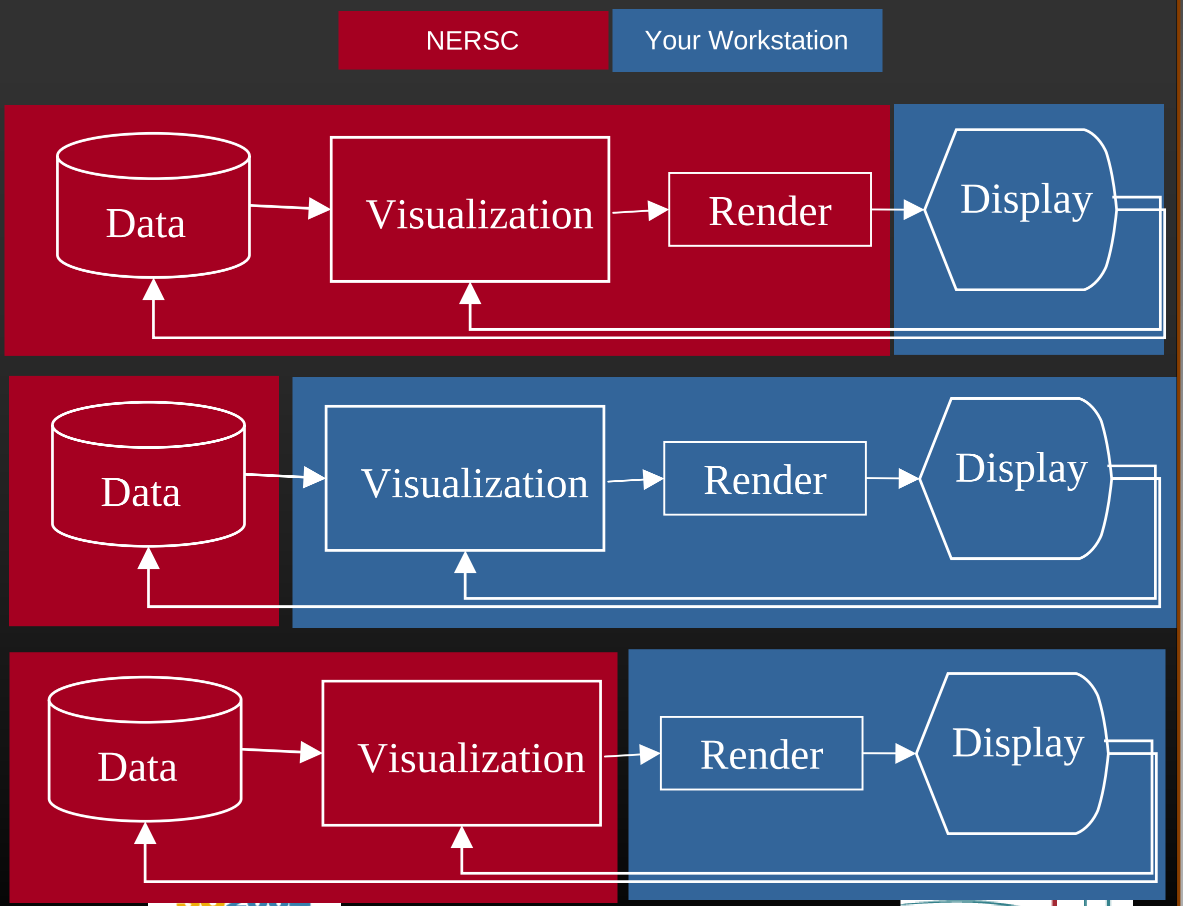
- Problem: component placement on distributed resources changes as a function of both performance target and specific data.
- Problem: distributed applications launched “by hand,” resource placement performed “by hand.”

Performance Modeling and Pipeline Optimization

- Approach: model performance of individual components, optimize placement as a function of performance target.
- Goal: automate the process of placing components on distribute resources.
- Results: quadratic order algorithm, high degree of accuracy.

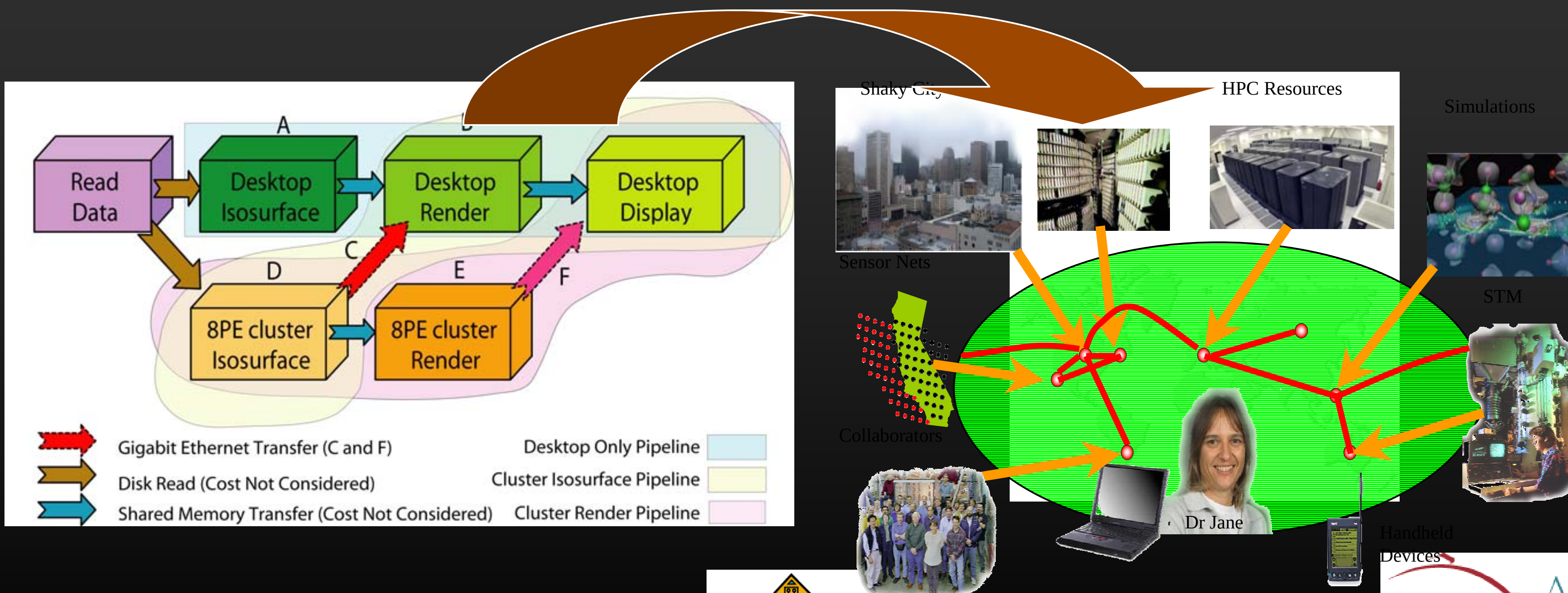
Performance Modeling and Pipeline Optimization

- **Render Remote**
  - Move images:
    - setenv DISPLAY
    - SGI's Vizserver
  - Data too big to move.
- **Render Local**
  - Move data
    - ftp, scp
    - Logistical networking
- **Hybrid approaches**
  - Move "vis results" for local rendering
  - CEI's Ensight, Visapult



Pipeline Optimization – User View

- Goal: simplify use of distributed visualization resources.

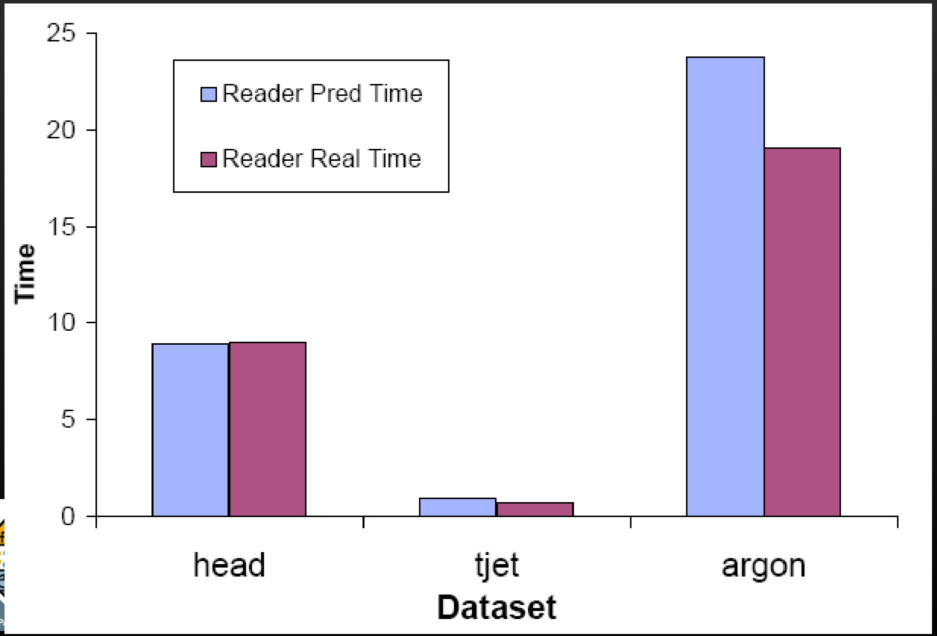


Visualization Pipeline Optimization – Overview

- **Obtain/derive performance measurements for pipeline components.**
  - Load data, compute isosurface, render & display.
- **Automatically select placement of tasks on distributed resources to meet performance objectives.**

Performance Modeling and Pipeline Optimization

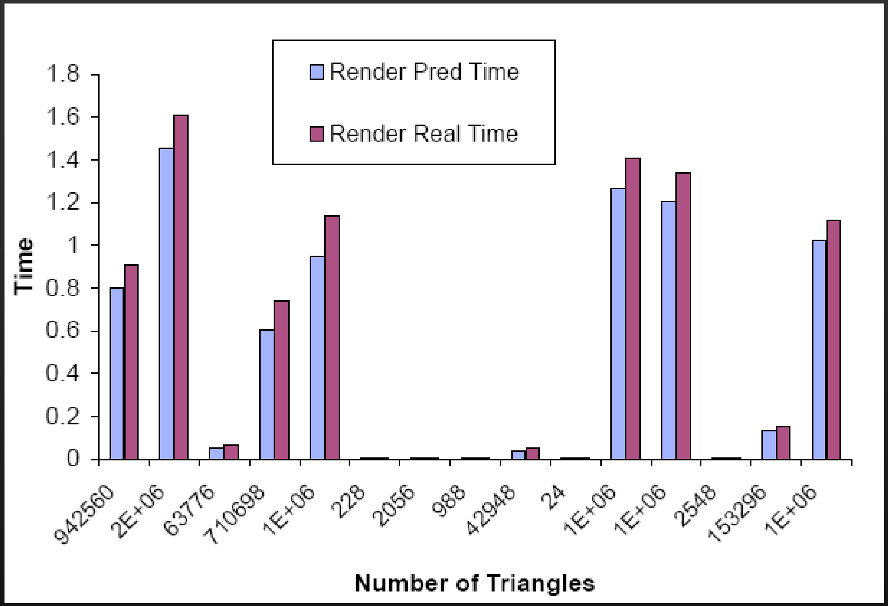
- **Single workflow:**
  - Reader -> Isosurface -> Render
- **Reader performance:**
  - Function of:
    - Data Size
    - Machine constant
  - $T_{reader}(n_v) = n_v * C_{reader}$



Office of Science
U.S. Department of Energy

Performance Modeling and Pipeline Optimization

- **Render Performance:**
  - Function of:
    - Number of triangles,
    - Machine constant.



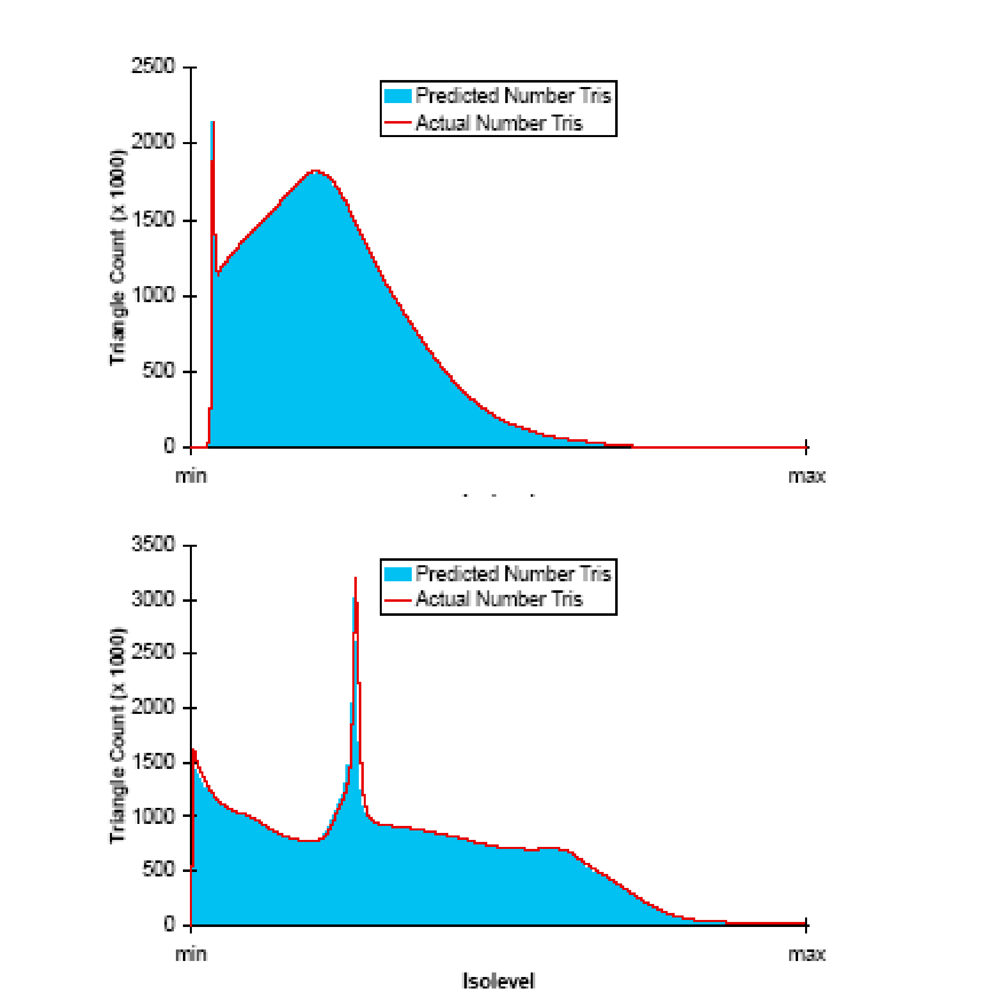
  - $T_{render} = n_t * C_{render} + T_{readback}$

Performance Modeling and Pipeline Optimization

- **Isosurface Performance:**
  - Function of:
    - Data set size,
    - Number of triangles generated (determined by combination of dataset and isocontour level).
  - Dominated number of triangles generated!
  - $T_{iso}(n_t,n_v) = n_v * C_{base} + n_t * C_{iso}$

Performance Modeling and Pipeline Optimization

- Precompute histogram of data values.
- Use histogram to estimate number of triangles as a function of iso level.



Office of Science
U.S. DEPARTMENT OF ENERGY

Performance Modeling and Pipeline Optimization

- **Performance targets:**
  - Optimize for interactive transformation.
  - Optimize for changing isocontour level.
  - Optimize for data throughput.

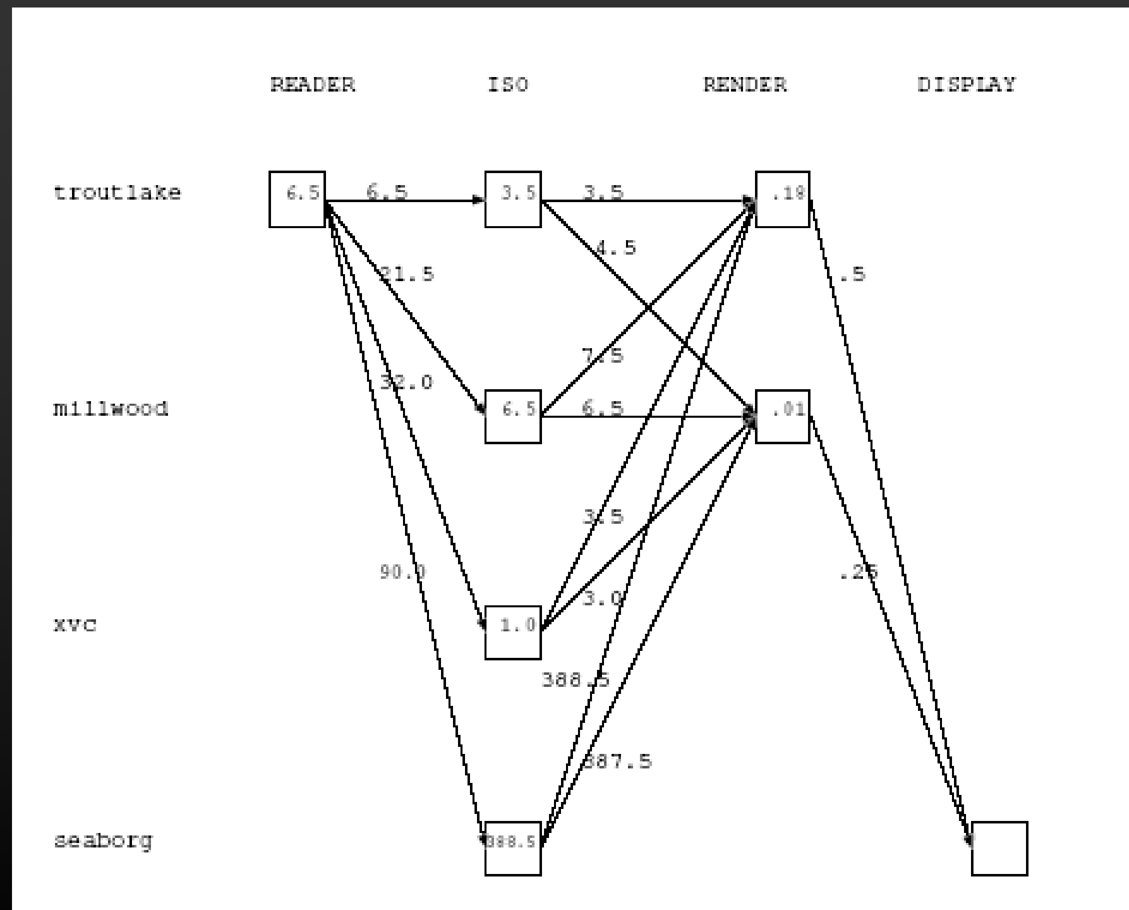
Performance Modeling and Pipeline Optimization

- **Pipeline Configurations:**
  - Render local – send data to workstation.
  - Render remote – send images to workstation.
  - Hybrid – send triangles to workstation.

Performance Modeling and Pipeline Optimization

- **Optimize placement using Djikstra's shortest path algorithm.**
- **Edge weights assigned based upon performance target.**
- **Low-cost algorithm:**

  O(E + VlogV)



Performance Modeling and Pipeline Optimization - Conclusions

- **"Microbenchmarks" to estimate individual component performance.**
  - Per-dataset statistics can be precomputed and saved with the dataset.
- **Low-cost (N*log*N) workflow-to-resource placement algorithm.**
- **Optimizes pipeline performance for an specific interaction target – relieves users from task of manual resource selection.**

Outline

- **Science-driven Visualization Challenges.**
- **LBNL Visualization Research**
  - Remote, Distributed and High Performance Visualization – Introduction and Approach.
  - Domain-specific solutions for scientific research.
  - Computer Science research.
- **Conclusion and Future Directions**

Conclusions

- **Close collaboration with applications produces usable, focused visualization technologies.**
- **Such collaborations are long-term relationships.**
  - How to formalize and sustain such relationships?

Conclusions

- **Component-based development holds much promise (see HECRTF).**
- **Underpinnings:**
  - Community-centric data models.
  - Interactive, parallel, distributed execution framework.

Conclusions

- Opportunity to move towards technology sharing and reuse, especially for visualization community.
- Produce usable, long-lived visualization technology for applications.
- Need for cross-program bridges – one form is stable infrastructure underpinnings based upon common component interfaces and community centric data models.

Summary

LBNL has a world-class Visualization R&D program that has a balanced and effective having an emphasis upon remote, distributed and high performance visualization, and meeting the needs of science.

Visit us on the web at http://vis.lbl.gov/

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Office of Science U.S. DEPARTMENT OF ENERGY

SC2004 PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB

The End

Office of Science
U.S. DEPARTMENT OF ENERGY

SC2004
PITTSBURGH PA NOVEMBER 6 - 12

BERKELEY LAB