Integrated Multi-field Visualization

Ken Joy
Director, Institute for Data Analysis and Visualization
Computer Science Department
University of California, Davis
What am I going to talk about?

- We work on “scale” problems all the time!

![Top500 extrapolation graph](image1)

(Sean Ahern)

![3D visualization](image2)

(Wes Bethel)
What am I going to talk about?

- We work on “applications” all the time!
  - Biomedical visualization, climate, ...

(Hari Krishnan)

(Valerio Pascucci)
What am I going to talk about?

- We work on “improving” our visualization “suite” of algorithms
  - Marching methods, topology methods, volume rendering, vector fields ...
Challenges of Petascale DNS: Mountains of Data

- Data size:
  - O(3/4 PB) raw field particle data on Jaguar
  - 25 GB/s ADIOS

- Data complexity:
  - Data is multi-variate
  - Turbulence chaotic
  - Wide range of scales
  - High intermittent matter!
  - Time-varying
  - Organized collectives
  - Non-locality in space and time
  - Temporal correlation vectors

Data Volumes

<table>
<thead>
<tr>
<th>Code</th>
<th># Variables</th>
<th>Resolution</th>
<th># Dumps</th>
<th>Total Volume</th>
<th>Runtime</th>
<th>Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHIMERA 1.0</td>
<td>~ 200</td>
<td>576x96x192</td>
<td>3000</td>
<td>~50 TB</td>
<td>~3 Months</td>
<td>1 PF</td>
</tr>
<tr>
<td>CHIMERA 2.0</td>
<td>~ 350 (expanded nuclear network)</td>
<td>576x96x192</td>
<td>3000</td>
<td>~90 TB</td>
<td>~3 Months</td>
<td>20 PF</td>
</tr>
<tr>
<td>GenASiS</td>
<td>~5000</td>
<td>512x512x512</td>
<td>3000</td>
<td>~30 PB</td>
<td>?</td>
<td>1 EF</td>
</tr>
</tbody>
</table>

GenASiS data get large quickly as we move from moments of the distribution function to the distribution function itself. Otherwise, data sizes increase modestly. The exascale machine will be a "strong scaling" platform (relatively dearth of memory).

Houston Texas, February 22-23, 2011

(Courtesy, Sean Ahern)
The Opportunity

### Data Volumes

<table>
<thead>
<tr>
<th>Code</th>
<th># Variables</th>
<th>Resolution</th>
<th># Dumps</th>
<th>Total Volume</th>
<th>Runtime</th>
<th>Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHIMERA 1.0</td>
<td>~ 200</td>
<td>576X96X192</td>
<td>3000</td>
<td>~50 TB</td>
<td>~ 3 Months</td>
<td>1 PF</td>
</tr>
<tr>
<td>CHIMERA 2.0</td>
<td>~ 350</td>
<td>576X96X192</td>
<td>3000</td>
<td>~90 TB</td>
<td>~ 3 Months</td>
<td>20 PF</td>
</tr>
<tr>
<td>GenASiS</td>
<td>~5000</td>
<td>512X512X512</td>
<td>3000</td>
<td>~30 PB</td>
<td>?</td>
<td>1 EF</td>
</tr>
</tbody>
</table>

- GenASiS data get large quickly as we move from moments of the distribution function to the distribution function itself.
- Otherwise, data sizes increase modestly.
  - The exascale machine will be a "strong scaling" platform (relative dearth of memory).
The Opportunity

### Data Volumes

<table>
<thead>
<tr>
<th>Code</th>
<th># Variables</th>
<th>Resolution</th>
<th># Dumps</th>
<th>Total Volume</th>
<th>Runtime</th>
<th>Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHIMERA 1.0</td>
<td>~ 200</td>
<td>576X96X192</td>
<td>3000</td>
<td>~50 TB</td>
<td>~ 3 Months</td>
<td>1 PF</td>
</tr>
<tr>
<td>CHIMERA 2.0</td>
<td>~ 350</td>
<td>576X96X192</td>
<td>3000</td>
<td>~90 TB</td>
<td>~ 3 Months</td>
<td>20 PF</td>
</tr>
<tr>
<td>GenASiS</td>
<td>~5000</td>
<td>512X512X512</td>
<td>3000</td>
<td>~30 PB</td>
<td>?</td>
<td>1 EF</td>
</tr>
</tbody>
</table>

- GenASiS data get large quickly as we move from moments of the distribution function to the distribution function itself.
- Otherwise, data sizes increase modestly
  - The exascale machine will be a “strong scaling” platform (relative dearth of memory)
### The Opportunity

**Data Volumes**

<table>
<thead>
<tr>
<th>Code</th>
<th># Variables</th>
<th>Resolution</th>
<th># Dumps</th>
<th>Total Volume</th>
<th>Runtime</th>
<th>Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHIMERA 1.0</td>
<td>~ 200</td>
<td>576x96x192</td>
<td>3000</td>
<td>~50 TB</td>
<td>~3 Months</td>
<td>1 PF</td>
</tr>
<tr>
<td>CHIMERA 2.0</td>
<td>~ 350 (expanded nuclear network of 10-150 species)</td>
<td>576x96x192</td>
<td>3000</td>
<td>~90 TB</td>
<td>~3 Months</td>
<td>20 PF</td>
</tr>
<tr>
<td>GenASiS</td>
<td>~5000</td>
<td>512x512x512</td>
<td>3000</td>
<td>~30 PB</td>
<td>?</td>
<td>1 EF</td>
</tr>
</tbody>
</table>

- GenASiS data get large quickly as we move from moments of the distribution function to the distribution function itself.
- Otherwise, data sizes increase modestly.
  - The exascale machine will be a "strong scaling" platform (relative dearth of memory).
Challenges in Multi-Field Visualization

- The number of variables studied in large-scale simulations will continue to increase.
  - ...interactions/correlations between the variables explodes
- So -- there are “extreme” opportunities in multi-field visualization
“Fields” will become more complex...

- Scalar fields
- Vector fields
- Tensor fields
- Function Fields
- Material Fractions
- Uncertainty
- Ensembles
- Distribution Fields
- Others...
“Fields” will become more complex...

- Scalar fields
- Vector fields
- Tensor fields
- Function Fields
- Cell-based Fields (e.g. Material Fractions)
- Uncertainty
- Ensembles
- Distribution Fields
- Others...

--- Worked on this forever!
--- Really good results in the last few years
--- Some results here
“Fields” will become more complex...

• Scalar fields
• Vector fields
• Tensor fields
• Function Fields
• Cell-based fields (e.g., Material Fractions)
• Uncertainty
• Ensembles
• Distribution Fields
• Others...

<--- Worked on this forever!
<--- Really good results in the last few years
<--- Some results here

Here are “future” opportunities in data exploration, analysis, and visualization
Cell-based Volume Fractions

“Volume of Fluid Methods”, mid-1970s

Greg Nielson and Dick Franke, Visualization Seminar, Dagstuhl 1997

Visualization Research on this topic since about 1997
Embedded Boundary/Material Interfaces

(John Anderson)
Material Interface Methods

Classic approximations

Higher-order Approximations
Multi-Material Works Now

Three-Material Example

(Iuri Prilepov)
Reproduces Fine Detail Better

Elongated Bubble

Three-Material Example

(Iuri Prilepov)
Function Fields

- At each vertex of a mesh, one is given a “function”

Air-pollution data from the San Joaquin Valley, CA. Each vertex has an associated function [particle size by number of particles]. This frame is from a large-scale 24-hour simulation of the air quality in the valley.
We turn functions fields into scalar fields

- Measure a functions “distance” from a given “exemplar” function

\[
d(f, g) = \left( \int_I w(t)(f(t) - g(t))^2 \right)^{\frac{1}{2}}
\]

\[
d(f, g) = \left( \sum_{1}^{n} w_i (f_i - g_i)^2 \right)^{\frac{1}{2}}
\]
Function Fields

(John Anderson)
A Solution for Multifield Visualization?

- Query-driven Data Exploration
  
  - The data is so large, and so complex, that we may as well treat it as a data base.
  
  - “Mine” previous work on data bases
  
  - The terms “Uncertain data bases”, “Fuzzy Data Bases”, and “Uncertain Queries” fits with Chris’ Talk (next)
Ensembles?

Studying 25 Rayleigh-Taylor Instability calculations (all at 10us)
Two “knobs”: turbulent viscosity coefficient, buoyancy coefficient
Five values for each knob, 25 pairs total

Average Speed over all 25

Max Speed over all 25

Min Speed over all 25

Biggest difference over all 25

(Hank Childs)
The Message

- Multi-Field Visualization is a very important component of our “future” data exploration and analysis tools.

- Lots of opportunities here!!!
  - How does one visualize various “field” types?

- ...

- How does one integrate fuzzy database theory into visualization methods?
Thanks...

• Contributors

• Funding Agencies
  • National Science Foundation, Lawrence Livermore National Lab, Lawrence Berkeley National Laboratory, Department of Energy (SciDAC Program), UC Davis, Siemens Corp
Thank You!

Ken Joy
UC Davis

joy@cs.ucdavis.edu