

# Framework Technologies & Methods for Large Data Visualization

*W T Hewitt, University of Manchester  
I Curington, AVS Inc*

## Abstract

The presentation will address large data visualization issues in the context of commercial visualization tool development. A review of techniques for multidimensional data visualization will be followed by case studies from CEM, CFD, VLSI, Medicine, and Geophysics.

An artifact of this type of visualization is that the visualization task itself becomes a consumer of HPC resources. The second part of the tutorial is concerned with the issues of implementing these techniques in a multiprocessor environment, and improving the performance of current visualization systems.

A range of technical areas will be discussed, including experimental research and production algorithm development.

Both current research and future challenges facing visualization system vendors will be discussed.

Attendees at the tutorial will gain an understanding of the issues underlying visualization in a parallel and distributed environment including:

- ❑ Familiarity with domain decomposition methods and parallelisation techniques
- ❑ Knowledge of the principles of volume, flow, and multidimensional visualization
- ❑ Ability to use distributed computation to enable accurate and timely visualization of large complex datasets
- ❑ Familiarity with the latest developments in visualization and HPC systems

## Detailed Description

The visualization of large datasets poses two major problems: understanding the available techniques and realizing these techniques, as this task itself requires HPC resources. This tutorial aims to attack these two issues.

In the first part we will review multidimensional data visualization through the following topics:

- ❑ What is multidimensional data

- A look at some problems and examples
- Techniques and their uses
  - Applying traditional techniques
  - New techniques: glyphs, textures, tables and stacked plots, scatterplots, Andrews curves, permutation matrices, parallel coordinates data sonification and virtual reality.
- Some visualization systems which cater for multidimensional data

This will be followed by case studies **selected from** the following areas

- **Computational Fluid Dynamics:** will be used to show techniques for visualizing vector and tensors, and how computer based techniques match experimental techniques, and how computer based techniques can do much more than experimental methods.
- **Computational Electromagnetics:** is an exciting new area, which is lacking in a broad range of visualization techniques. We will present some work which has looked at a variety of ways of visualizing this type of data.
- **Medical Data:** The large data sets produced by a variety of scanners are a rich source of data. We will present a number of examples where visualization has made a significant impact on the clinical processes used, and the some volume visualizations of a mummy.
- **VLSI Mask Layout:** Complex VLSI designs for IC mask fabrication use deeply hierarchical geometric layout formats. With even modest file size, nested structure references can easily generate many millions of geometric primitives. The custom render method facility in the AVS/Express graphics pipeline is used to traverse the layout hierarchy at display time, using the native VLSI design data structure. Using this technique, 26 million graphics primitives have been viewed in a single view on a standard laptop PC.
- **Geophysics :** Interactive geophysical analysis applications often push the boundaries of available computing environments. To view the results of an oil reservoir analysis covering many years of oil field production, the data structures must accommodate a large number of reservoir properties, a large and topologically complex mesh structure, and a large number of time-varying state properties. Visualization software architectures for this class of application must use mixed in-core/out-of-core techniques, topological structure extraction, and mesh optimization methods.

The second part of the tutorial will focus on two areas: implementing some of these algorithms, particularly volume visualization and surface extraction, in multiprocessor environments, and secondly on improving the performance of visualization systems themselves.

Inevitably we must briefly cover general material: issues of parallel computing, summary of parallel architectures, taxonomy of parallel visualization decompositions.

The main material in this section covers parallel surface extraction and parallel direct methods on SIMD, DM-MIMD and VSM-MIMD machines. Next we cover parallel compositing.

We next look at issues of integrating parallelism into visualization systems, by describing the Visualization in Parallel (VIPAR) project we have carried out. VIPAR provides an environment for the development of parallel visualization modules, through three phases:

- An automatic parallel module generator
- A dataflow network editor for building managing parallel visualization modules
- A development environment for (remote) computational steering.

We will present the system architecture, its working methods, and some applications of the work

In AVS/Express, a very flexible graphics pipeline architecture is exploited to minimize memory requirements and data structure conversions during rendering for large data problems. A lean, thin layer OpenGL renderer provides a display path, without dependence on scene tree, dynamic scene creation, or any requirement to hold the scene content in virtual memory. For small data sets and high-speed interaction, stored scene tree mechanisms are used, but not required. Data source methods, and visualization methods are able register user-defined draw methods providing a “chunking” mechanism to build up the scene using a multi-pass or model segmentation model. Visualization procedures are only executed as needed during the incremental updates, and do not need access to the entire model. In addition, this architecture allows procedural rather than stored data structure sources.

**Coupling Disk-mapped I/O to Visualization:** Extending the data reference and event handling model, a set of file-access modules have been constructed to map file content structures to memory structures used in visualization. These manage the file access on demand from the visualization modules, so that standard visualization modules can be used on files larger than virtual memory. A recent animation won an award at the ASP/DFD meeting created from 193 GBytes of data using this method.

**Data Dependent Surface Reduction:** Two geometric surface reduction systems have been developed to aid visualization of complex objects. The first is a fast interactive tool, to aid in maneuvering around complex shapes. The second provides precise error control, so the reduction result has measurable deviations from the original, as required in medical applications. Further enhancements include the ability to analyze data gradients on the original mesh and preserve critical data features in the reduced mesh.

**Multi-Channel/Multi-Pipe Onyx:** A design for interactive visualization for the Silicon Graphics Multi-Pipe Onyx will be discussed. The extensions to the standard visualization-rendering pipeline, using the Multi Pipe Utility (MPU) will be described, with shared memory and parallel rendering process architecture.

In summary we will present techniques for visualizing large datasets, how these techniques might be realized in a HPC environment, and how current commercial visualization systems are developing to accommodate the large datasets.

### ***What Will Attendees Learn?***

The tutorial is aimed at intermediate level and is for attendees who have some knowledge of visualization and/or parallel computing and who wish to learn more about visualization in a HPC environment.

Attendees at the tutorial will gain an understanding of the issues underlying visualization in a parallel and distributed environment including:

- ❑ Familiarity with domain decomposition methods and parallelisation techniques
- ❑ Knowledge of the principles of volume, flow, and multidimensional visualization
- ❑ Ability to use distributed computation to enable accurate and timely visualization of large complex datasets
- ❑ Familiarity with the latest developments in visualization and HPC systems

### **Course Schedule**

<b>Time</b>	<b>Topic</b>
30 mins	Who are we and what are we going to talk about
60 mins	Multidimensional Data Visualization Techniques
30 mins	Volume visualization
60 mins	Case studies
30 mins	Parallel Strategies: issues of parallel computing, summary of parallel architectures taxonomy of parallel visualization decompositions
60 mins	Parallel volume visualization algorithms, VIPAR
60 mins	Optimizing Visualization systems: disk mapping, data dependent surface reductions, multi-channel/multi-pipe Onyx
30 mins	Future developments, Q&A and wrap up