VAPOR: A tool for interactive visualization of massive earth-science datasets

Alan Norton and John Clyne
National Center for Atmospheric Research
Boulder, CO USA
Presentation at AGU Dec. 13, 2010

This work is funded in part through U.S. National Science Foundation grants 03-25934 and 09-06379, and through a TeraGrid GIG award

Alan Norton (vapor@ucar.edu)
Outline

• Introduction:
  – Problems with visualizing and analyzing massive data
  – VAPOR project goals

• Overview of VAPOR’s capabilities
  – Support for browsing massive data
  – Basic visualization features
  – Unique visualization capabilities

• Three live demos:
  – Geo-referencing, tracking Hurricane Katrina in WRF output
  – Identification and analysis of P. Mininni’s ‘Current Roll’ in a large MHD dataset
  – Identification and tracking of vortices in a high-resolution hurricane dataset (Y. Chen)
Typical Analysis/Vis Workflow

Supercomputing \[\rightarrow\] Temp Disk \[\rightarrow\] Archive

Offline processing:

Analysis and Visualization \[\rightarrow\] Analysis Repository

Alan Norton (vapor@ucar.edu)
Archival is not keeping up

- Supercomputer sustained computation rate is doubling every 12-15 months
- Archive storage capacity is doubling every 25-26 months
- Fraction of data saved for analysis halves every ~2 years.

Alan Norton (vapor@ucar.edu)
Visualization and Analysis are limited by I/O

Performance improvements for I/O 1977-2005, compared with computation rate improvements

Source: DARPA HPCS I/O presentation

Alan Norton (vapor@ucar.edu)
Problems with Petascale Analysis/Vis Workflow

- Insufficient capacity, speed
- Only infrequent archival
- Takes days or weeks
- Only for small samples, statistics
- Insufficient speed for interactivity

Supercomputing → Temp Disk → Archive

Analysis and Visualization

Analysis Repository

Offline processing:
Implications for Visualization and Analysis of Petascale computations

Two serious problems:

- Smaller portion of data is available for analysis because of limited storage and archive capacity.
- Analysis and visualization of the available data becomes non-interactive due to limited I/O rates

Result: Loss of scientific productivity

[Numerical] models that can currently be run on typical supercomputing platforms produce data in amounts that make storage expensive, movement cumbersome, visualization difficult, and detailed analysis impossible. The result is a significantly reduced scientific return from the nation's largest computational efforts.

Mark Rast
University of Colorado, LASP

Alan Norton (vapor@ucar.edu)
VAPOR project overview

The VAPOR project is intended to address the problem of datasets that are becoming too big to analyze and visualize interactively

- **VAPOR** is the **Visualization and Analysis Platform for Oceanic, atmospheric and solar Research**
- **Goal**: Enable scientists to *interactively* analyze and visualize massive datasets resulting from fluid dynamics simulation
- **Domain focus**: 2D and 3D, gridded, time-varying turbulence datasets, especially earth-science simulation output.
- **Essential features**:
  - Multi-resolution data representation for accelerated data access
  - Exploits GPU for fast rendering
  - Interactive user interface for scientific visual data exploration
  - Available (free) on Mac, Windows, Linux
Wavelet transforms for 3D multiresolution data representation

• Some wavelet properties:
  – Data can be accessed at desired resolution and compression level
  – Lossless or Lossy (up to 500:1 compression)
  – Numerically efficient ($O(n)$)
• Forward and inverse transform
  – No additional storage cost
Wavelet compression enables 100:1 or better compression with minimal degradation

original
64-fold averaged with Haar
Smyth: salt sheet boundary simulation
64-fold compress biorth spline

Alan Norton (vapor@ucar.edu)
Petascale workflow using wavelet compression

- Data would be interactively analyzed and visualized, during and after simulation.
- Intermediate times available in compressed form.

Remote visualization sessions supported

Monitor wavelet-compressed results

Rapid retrieval of requested data

Frequent compressed saves

Infrequent full saves;

Supercomputing

Temp Disk

Archive

Wavelet Repository

Interactive Analysis and Visualization

Alan Norton (vapor@ucar.edu)
VAPOR capabilities (latest version: 2.0)

- All tools perform interactively, exploiting multi-resolution representation
- Wavelet compression enables up to 500:1 reduction of I/O reads
- GPU-accelerated interactive graphics
- Python calculation of derived variables
- Flow integration
  - Streamlines, particle traces
  - Field line advection
  - Image-based flow visualization
- Data probing and contour planes
- WRF-ARW terrain-following grids
  - Direct import of WRF output files
- Geo-referenced image support

Smyth, salt sheet boundary simulation

Mininni, Current roll

Alan Norton (vapor@ucar.edu)
How VAPOR differs from other visualization platforms

- Multi-resolution data representation with compression
  - To enable interactive display and analysis of peta-scale datasets
- Python and NumPy embedded support
- Intended to be used by scientists, not visualization engineers
  - Requirements defined by a steering committee of scientists
- Narrow focus: turbulence simulation on gridded domains
- *Not* built on existing visualization libraries (e.g. VTK)
- Emphasis on desktop and laptop platforms; no distributed implementation

Alan Norton (vapor@ucar.edu)
Demonstrations

• Visualization of WRF-ARW output: Hurricane Katrina with geo-referenced terrain images

• Identification and analysis of ‘Current Roll’ in a 1536$^3$ MHD simulation (Pablo Mininni)
  – Interactive browsing by controlling refinement level and compression
  – Flow seed positioning for magnetic field lines
  – Magnetic field line advection

• Identification and tracking of vortices in eye-wall of high-resolution hurricane simulation (Yongsheng Chen)
  – Image-based flow visualization identifies vortices
  – Field-line advection tracks them

Alan Norton (vapor@ucar.edu)
Visualize and track hurricane Katrina

- Visualization of WRF-ARW simulation with moving nest
- Apply terrain image and county maps obtained from Web Mapping Services
- Geo-referencing provides spatial context for streamlines, volume rendering, isosurfaces.

Alan Norton (vapor@ucar.edu)
Visualization Discovery Example

Small scale structures in MHD turbulence with high Reynolds number

- Data from Pablo Mininni, NCAR
- 1536x1536x1536 volume, 16 variables (216 GB per timestep)
- Scientific goal: understand MHD flow dynamics at high resolution and high Reynolds no.
- Analysis and visualization performed with VAPOR and IDL
- Resulted in discovery of intertwining current sheets ("current rolls")

Alan Norton (vapor@ucar.edu)
Multi-resolution data browsing

Wavelet data representation supports control of data resolution as well as compression level

- Interactively visualize full data at low resolution, high compression
- Zoom in, increase resolution, reduce compression for detailed understanding
Using multiple time-steps:
Track evolving structures with field line advection

Animates field lines in velocity field

- Useful for tracking evolution of geometric structures (e.g. vorticity field lines in tornado)
- Based on algorithm proposed by Aake Nordlund

Data provided by P. Mininni
Visualization was used to understand the nature of increased turbulence along eye-wall

- Unsteady flow shows overall wind dynamics
- VAPOR’s IBFV tool can identify horizontally oriented transient vortex tubes near the ocean surface
- Using VAPOR’s field line advection, these vortices can be tracked and animated over time

Results of Yongsheng Chen

Alan Norton (vapor@ucar.edu)
VAPOR plans

VAPOR’s steering committee and other users help prioritize features, with releases every 9-12 months

- Vapor 2.0.0 was just released

Some high priority features for upcoming releases:
- Parallel conversion of data during simulation
- Animation control
- Extensible architecture
- Iso-Lines
- Support for ocean modeling

Alan Norton (vapor@ucar.edu)
Summary

- VAPOR is designed to enable interactive visualization and analysis of massive datasets by exploiting the wavelet multi-scale representation.
- VAPOR supports a variety of useful interactive techniques for investigating and visualizing data, based on needs expressed by scientific users.
- Recent improvements to VAPOR are designed to enable interactive access to petabyte datasets and support anticipated peta-scale workflows.

Alan Norton (vapor@ucar.edu)
VAPOR Availability

• Version 2.0 just released
  – available on Website
• Runs on Linux, Windows, Mac
• System requirements:
  – a modern (nVidia or ATI) graphics card (available for about $200)
  – ~1GB of memory
• Executables, documentation available (free) at
  http://www.vapor.ucar.edu/
• Source code, feature requests, etc. at
  http://sourceforge.net/projects/vapor
• Contact: vapor@ucar.edu
Questions?

Alan Norton (vapor@ucar.edu)
Acknowledgements

- **Steering Committee**
  - Nic Brummell - CU
  - Yuhong Fan - NCAR, HAO
  - Aimé Fournier – NCAR, IMAGe
  - Pablo Mininni, NCAR, IMAGe
  - Aake Nordlund, University of Copenhagen
  - Helene Politano - Observatoire de la Cote d'Azur
  - Yannick Ponty - Observatoire de la Cote d'Azur
  - Annick Pouquet - NCAR, ESSL
  - Mark Rast - CU
  - Duane Rosenberg - NCAR, IMAGe
  - Matthias Rempel - NCAR, HAO
  - Geoff Vasil, CU
  - Thara Phrabhakaran, IITM
  - Gerry Creager, TAMU
  - Leigh Orf, Central Mich. U.

- **WRF consultation**
  - Wei Wang – NCAR, MMM
  - Cindy Bruyere – NCAR, MMM
  - Yongsheng Chen-NCAR, MMM
  - Wei Huang – NCAR, CISL

- **Developers**
  - John Clyne – NCAR, CISL
  - Alan Norton – NCAR, CISL
  - Kenny Gruchalla – CU
  - Victor Snyder – CSM
  - Rick Brownrigg – NCAR, CISL
  - Pam Gillman – NCAR, CISL

- **Research Collaborators**
  - Kwan-Liu Ma, U.C. Davis
  - Hiroshi Akiba, U.C. Davis
  - Han-Wei Shen, Ohio State
  - Liya Li, Ohio State

- **Systems Support**
  - Joey Mendoza, NCAR, CISL

Alan Norton (vapor@ucar.edu)