

VAPOR: Visual, Statistical, and Structural Analysis of Astrophysical Flows

John Clyne

National Center for Atmospheric Research, Boulder, Colorado USA

Kenny Gruchalla

National Renewable Energy Laboratory, Golden, Colorado USA

Mark Rast

University of Colorado, Boulder, Colorado USA

Abstract. In this paper we discuss recent developments in the capabilities of VAPOR: a desktop application that leverages today's powerful CPUs and GPUs to enable visualization and analysis of terascale data sets using only a commodity PC or laptop. We review VAPOR's current capabilities, highlighting support for Adaptive Mesh Refinement (AMR) grids, and present new developments in interactive feature-based visualization and statistical analysis.

1. Introduction

Interactive visualization and analysis of data from astrophysical flow simulations faces increasing challenges with the ever increasing size of those calculations. The mismatch between visualization/analysis and computational resources means that some form of data reduction must be employed to maintain interactivity in the visualization/analysis process. To date VAPOR has addressed this challenge via multiresolution access and Cartesian-volume region-of-interest (ROI) extraction (Clyne & Rast 2005). We briefly summarize here key elements of the VAPOR visual data analysis environment before discussing at some length VAPOR's support for Adaptive Mesh Refinement (AMR) grids and recent developments in iterative feature based visualization and analysis. These developments extend the region-of-interest concept to volumes defined by the solution properties and field variable correlations, loosely termed structures or features.

2. VAPOR¹

In previous work we have described in detail many of the capabilities of the VAPOR package (Clyne & Rast 2005; Clyne et al. 2007; Rast & Clyne 2008; Mininni et al. 2008b). Three key components distinguish VAPOR from other advanced visualization packages that the authors are aware of:

¹VAPOR open source available at <http://www.vapor.ucar.edu>

- a wavelet based multiresolution data model enables interactive data browsing of high-resolution simulation outputs using only modest computing resources (e.g., a conventional desktop or laptop)
- a feature set that is targeted toward the specialized analysis needs of the astrophysical and geophysical computational fluid dynamics communities
- a close coupling between VAPOR's highly interactive exploratory visualization capabilities and ITT's *fourth-generation* scientific data processing language, IDL

2.1. Multiresolution

VAPOR utilizes a hierarchical data representation as a strategy to approach the challenges of interactively analyzing large-scale data volumes. The simulation outputs are stored hierarchically, with each level in that hierarchy providing a coarsened approximation of the data at the preceding level. This approach exploits the fact that many visualization and analysis operations can tolerate a level of information loss by retrieving only the level of fidelity that is required for the current operations. For analysis operations that require access to the data at full resolution, this approach still allows the original data to be accessed in their entirety, without loss of information.

The hierarchical multiresolution access is accomplished through a wavelet decomposition and reconstruction scheme (Clyne 2003). The data are stored as hierarchy of successively coarser wavelet coefficients, with each level representing a halving of the data resolution along each spatial axis, resulting in an eight-fold reduction in the size of the data volume, and the corresponding reduction in required visualization and analysis resources. Storing the data hierarchy as wavelet coefficients avoids the penalty of keeping multiple data copies. A three-dimensional Haar wavelet (Haar 1910) is currently being used for this transformation. The computational cost of the forward and inverse transforms are negligible compared to those incurred by reading or writing the data, allowing the reconstruction of the data at factor-of-two resolutions with only minimal overhead.

This hierarchical data access scheme allows an investigator to control the fidelity of data in accordance with the available resources, the desired interactivity, and the requirements of the analysis. This forms the basis for an iterative analysis process, where the investigator can interactively browse coarsened representations of the data across the global spatiotemporal domain to identify features of interest. Once identified, the analysis domain can be restricted to these features, increasing the level of data resolution that can be handled interactively. Often both visual inspection and numerical analysis are fairly insensitive to considerable data coarsening (Clyne & Rast 2005), providing substantial savings in computational costs and input/output overhead during the early exploratory stages of investigation when interactivity is most crucial. Of course, subsequent verification of the analysis results can be accomplished less interactively at full resolution if necessary.

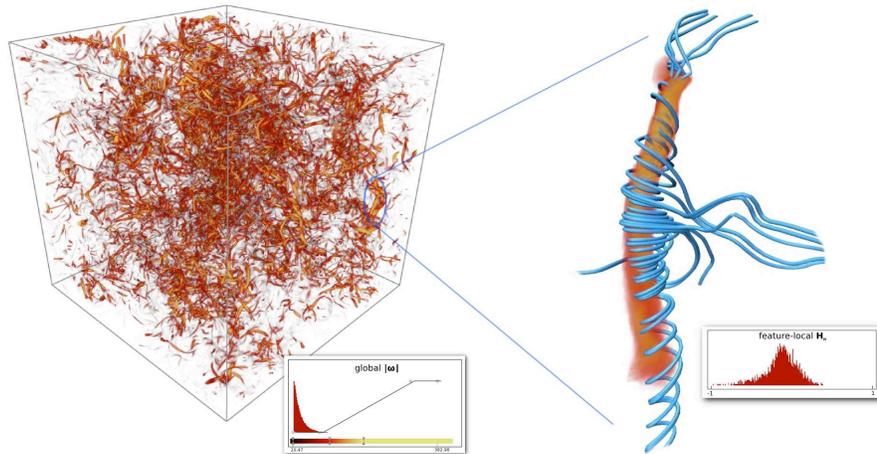


Figure 1. Tens of thousands of vortical structures (*left*) in a high-resolution simulation of Taylor-Green forced turbulence (Mininni, Alexakis, & Pouquet 2008a). A vortical feature identified and extracted from a data volume using high global values of vorticity and low absolute values of feature-local helicity (*right*), shown with streamlines seeded in the velocity field. These types of non-Cartesian regions of interest can be isolated as structures and examined both visually and quantitatively.

2.2. Targeted features

While VAPOR supports numerous general purpose visualization algorithms. It also provides capabilities tailored towards astrophysical and geophysical CFD needs. One example, discussed in detail in Mininni et al. (2008b), is the integration and display of magnetic field lines advected by a velocity field. Other specialized algorithms, not reported elsewhere, include methods for visually guided placement of streamline and pathline seed points based on physical properties of the flow such as local field maxima or minima. Interactive seeding is facilitated by cutting planes at arbitrary orientations in the volume and interactive probing of the data values.

A recent development (Gruchalla et al. 2009) has focused on broadening the concept of a ROI beyond Cartesian sub-volumes to the coherent structures by combining multivariate volume visualization techniques (Kniss, Kindlmann, & Hansen 2002; Doleisch, Gasser, & Hauser 2003) with a connected component analysis (Suzuki, Horibia, & Sugie 2003). Structures can be broadly and iteratively defined by multivariate transfer functions, and can thus represent *local* regions of correlation or anticorrelation between field variables as well as those identified by more traditional thresholding of a single measure. The algorithm executes a connected component analysis of the volume, based on user defined opacity values in the transfer function, to label individual structures. Structure dependent histograms of the original or other derived variables can then be displayed, and structure statistics can be used to guide further selection, definition, and identification in an iterative refinement loop. For example, from the tens of thousands of vortical structures in a recent simulation of Taylor-Green forced turbulence (Mininni, Alexakis, & Pouquet 2008a) those regions with both high

vorticity and low helicity can be readily identified and extracted (Figure 1) and compared to other highly vortical but more helical regions. Such non-Cartesian ROI extraction can significantly reduce data volumes, with the coordinates of the voxels contained in the structures readily output for use in subsequent analysis.

2.3. Coupling visualization with quantitative data analysis

VAPOR seamlessly interfaces with ITT's *fourth-generation* language IDL, allowing investigators to perform rigorous quantitative analyses guided by VAPOR's intrinsic 2D and 3D visualization capabilities. The integration of IDL and VAPOR is facilitated by metadata exchange defining the attributes and resolution of the data. A library of data-access routines allows IDL to read and write data in VAPOR's wavelet-encoded representation—an approach that is readily generalizable to other analysis packages. In typical usage, the investigator will maintain simultaneously active VAPOR and IDL sessions, visually identifying ROIs with VAPOR and exporting them to IDL for further study. Interactivity is maintained if the ROI is sufficiently small or if the operation is sufficiently well-behaved over coarsened approximations of the data (Clyne & Rast 2005). Quantities derived in the IDL session are imported back into the existing VAPOR session for continued visual investigation. Through the iteration of this process, large-scale data sets can be interactively explored, visualized, and analyzed without the usual delays caused by reading, writing, and operating on the data arrays in full.

The primary benefit of coupling visual data investigation with an high-level data analysis language is the ability to target expensive calculations of derived quantities to specific ROIs. The memory and computing requirement for calculating such variables in advance, across the entire domain, can require exorbitant resources, delaying or preventing further analysis. Moreover, the computation of some analysis quantities requires prior knowledge of the solution. This is particularly true if they are defined by field values (e.g. regions of maximum or minimum measure) or correlations between the flow variables. The new interactive non-Cartesian-volume feature based ROI capabilities of VAPOR allow, via a multivariate transfer function, precise definition of ROIs based on solution properties, and can thus focus analysis on highly reduced sub-volumes.

3. AMR

VAPOR supports a form of the block structured AMR grid that is most closely described by MacNeice et al. (2000), implemented in the PARAMESH package, and presently employed by the FLASH astrophysical thermonuclear flash code². The computational domain is covered by a base level, regular, Cartesian grid with uniform sampling. The base grid is partitioned into uniformly sized, non-overlapping blocks: each block contains the same number of samples and covers the same size physical space. Individual *parent* blocks may be refined by subdividing them into eight *child* octants. This refinement may be performed recursively creating an octree hierarchy, with varying levels of refinement. A

²<http://flash.uchicago.edu/website/home/>

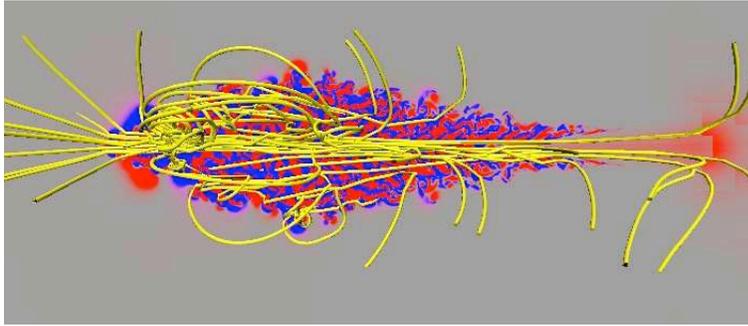


Figure 2. A 3D plume simulation computed with the FLASH astrophysical thermonuclear flash code. With on-the-fly resampling to a uniform grid, the full range of VAPOR capabilities are directly applicable to the AMR gridded data.

maximum depth of 10 or 20 levels is not uncommon. All blocks in the hierarchy contain the same number of uniformly distributed samples. VAPOR supports a somewhat less restrictive AMR mesh structure than that of PARAMESH, not requiring that adjacent blocks differ by no more than one level of refinement.

Direct visualization of AMR grids is a complex task. Only recently have practical algorithms been published for such routine visualization algorithms as direct volume rendering or isosurface construction (Weber et al. 2001b,a). VAPOR supports numerous fundamental visualization algorithms as well as novel visualization methods not found in other packages. To avoid the onerous task of generalizing all of VAPOR's principal visualization methods to support both regular rectilinear grids and a variety of AMR strategies, the approach taken by VAPOR is to resample AMR grids onto a uniformly sampled Cartesian grid. The resampling is performed on the fly, as needed, and at the resolution selected by the user. The user controlled sampling frequency matches a refinement level in the AMR grid, with blocks in the ROI possessing a coarser sampling than the user specified desired sampling refined through interpolation, and blocks of finer AMR sampling coarsened. This treatment is analogous to the wavelet based coarsening and refining that underlies VAPOR operations on regular rectilinear grids. The computational cost of this regridding is fairly modest and is ameliorated somewhat by VAPOR's extensive employment of caching.

At present the only AMR file format that VAPOR's interactive analysis tool, "vaporgui", is capable of reading is VAPOR's own custom format. Preparing an AMR data set for analysis with VAPOR requires first translating the data into this format. Command line utilities are provided for translation of FLASH data sets stored in the HDF5 file format ³. Further, example codes are provided that may be customized for use with other AMR encodings.

³<http://www.hdfgroup.org>

4. Conclusion

VAPOR continues to evolve to meet the visualization and analysis challenges facing computational astrophysical and geophysical fluid dynamicists as we near petascale compute capabilities. The focus remains on providing a flexible and useful tool for use in interactive analysis. Both adaptive mesh refinement and data volume reduction in post-batch analysis will be essential to interactivity in the petascale environment, and efficient algorithmic merging of these remains an ongoing challenge.

Acknowledgments. This work was funded in part by the National Science Foundation under grant ITR-0325934. The AMR plume data was provided by Matthias Rempel, and that of Taylor-Green turbulence by Pablo Mininni.

References

- Clyne, J. 2003, in Proceedings of Visualization, Imaging, and Image Processing '03, 152–157
- Clyne, J., Mininni, P., Norton, A., & Rast, M. 2007, *New J of Physics*, 9, 301
- Clyne, J., & Rast, M. 2005, in SPIE-IS&T Electronic Imaging, Vol. 5669, 284–294
- Doleisch, H., Gasser, M., & Hauser, H. 2003, in Proceedings of the symposium on Data visualisation 2003 (Grenoble, France: Eurographics Association), 239–248
- Gruchalla, K., Rast, M., Bradley, E., Clyne, J., & Mininni, P. 2009, in IDA 2009, Lecture Notes in Computer Science, Vol. 5772, 321–332
- Haar, A. 1910, *Mathematische Annalen*, 69, 331
- Kniss, J., Kindlmann, G., & Hansen, C. 2002, *IEEE Transactions on Visualization and Computer Graphics*, 8, 270
- MacNeice, P., Olson, K. M., Mobarry, C., de Fainchtein, R., & Packer, C. 2000, *Computer Physics Communications*, 126, 330
- Mininni, P., Alexakis, A., & Pouquet, A. 2008a, *Physical Review E*, 77, 036306
- Mininni, P., Lee, E., Norton, A., & Clyne, J. 2008b, *New Journal of Physics*, 10, 125007
- Rast, M., & Clyne, J. 2008, in Numerical Modeling of Space Plasma Flows/ Astronom-2007, ASP Conference Series, Vol. 385, 299–308
- Suzuki, K., Horibia, I., & Sugie, N. 2003, *Computer Vision and Image Understanding*, 89, 1
- Weber, G. H., Kreylos, O., Ligocki, T. J., Shalf, J., Hagen, H., Hamann, B., Joy, K. I., & Ma, K.-L. 2001a, in VMV '01: Proceedings of the Vision Modeling and Visualization Conference 2001 (Aka GmbH), 121–128
- Weber, G. H., Kreylos, O., Ligocki, T. J., Shalf, J. M., Hagen, H., Hamann, B., & Joy, K. 2001b, in *Data Visualization 2001* (Springer Verlag), 25–34