NUMERICAL MODELING OF SPACE PLASMA FLOWS / ASTRONUM-2007
ASP Conference Series, Vol. 385, ©2008
Nikolai V. Pogorelov, Edouard Audit, and Gary P. Zank, eds.

Coupled Analysis and Visualization of High Resolution Astrophysical Simulations

Mark Rast

University of Colorado, LASP, 1234 Innovation Dr., Boulder, CO 80303, U.S.A.

John Clyne

National Center for Atmospheric Research, 1850 Table Mesa Dr., Boulder, CO 80303, U.S.A.

Abstract. Computational physics has benefited from on-going microprocessor innovations, which have enabled larger and larger numerical simulations. One consequence of these technological advancements has been an explosion in the amount of data generated. For many modelers, available software tools and computing resources are proving inadequate for investigation of high-resolution numerical outputs. In this paper we discuss the general problems associated with very large data visualization and analysis and our work on a particular solution to those through the development of VAPOR (open source, available at http://www.vapor.ucar.edu): 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 briefly illustrate VAPOR's utility through the exploration of a high-resolution simulation aimed at understanding the effects of hydrogen ionization on convective dynamics in stellar envelopes.

1. Introduction

Continual advancements in microprocessor technology have in the last decades dramatically increased the capability of the supercomputers available for computational science. Systems with *teraflop*, and in the near future *petaflop*, performance have enabled, and will continue to enable, numerical models of extraordinary resolution and scale, with consequent enormous data output. The ability to manage, analyze, and study the numerical output of scientific computations has not kept pace with the capacity to generate it. For many numerical modelers the greatest challenge in the scientific discovery process begins once the simulation has completed and analysis commences.

Even within the discipline of computational fluid dynamics (our particular focus), the magnitude of this challenge varies with the scientific goals. Operational and forecast models often reduce very large three-dimensional volumetric data to two-dimensional maps (e.g., surface temperatures or winds) before output, thus dramatically diminishing difficulties with data manipulation. In research focused strictly on statistical or spectral properties of flows, the dimensionality of the solution output may be similarly reduced. This is not true of efforts which aim to understand the local dynamics, thermodynamics, or physical stability of the solution. These studies typically focus on force and energy balance at precise locations within the domain. The locations of interest, the nature of the analysis to be undertaken, and the relevant secondary quantities to be derived, are often not known or determinable before the solution is computed. In such cases, visualization and analysis on interactive time scales becomes essential, and the scientific return realized from the initial computational effort invested can be limited by the availability, or lack there-of, of appropriate post-processing resources, both software and hardware.

Many factors have contributed to an imbalance between the size of the problems currently computable and the size of the solutions interactively analyzable. Perhaps the most important is the contrasting natures of numerical simulation, which is well-suited to *batch* job submission, and data analysis, which is inherently *interactive*. This contrast is exacerbated by the emphasis that computing centers typically place on the delivery of batch computing cycles, often to the detriment of other computing needs. Most batch compute systems consist of many hundreds or thousands, and soon tens of thousands, of processing units, while analysis and visualization resources rarely exceed tens of CPUs. The problem also reflects a disparity in the rate of advancement of other essential computing technologies. For example, microprocessor performance, the main driver behind supercomputer advancements, doubles roughly every 18 months in accordance with "Moore's law" (Moore 1965). IO interconnect bandwidth performance, key to the interactive processing of very large data sets, is on a much more modest improvement curve. From 1977 to present, CPU performance has increased by nearly seven orders of magnitude while disk data transfer rates have increased by only two (Ross et al. 2005). Other computing technologies exhibit similar disparities, and these trends are expected to continue into the foreseeable future.

How then does one enable interactive interrogation of terascale data sets? Without an unrealistic leveling of computing technology advancement or a fundamental change in the provisioning of high performance computing resources, some form of data reduction is necessary. Extraction of local subregions from the global spatial-temporal domain, based on the occurrence of interesting events, offers one possibility. Another is the global approximation of the discrete solution using fewer samples than the original computation. The challenge in the first case is locating the regions of interest (ROIs), and in the second, achieving sufficient accuracy to maintain confidence in the coarsened approximation.

We have implemented both forms of data reduction in the VAPOR package (an open source desktop application available at http://www.vapor.ucar.edu) and have demonstrated success investigating numerous very large scale simulation outputs using only modest computing resources (Clyne et al. 2007). Multiresolution data access, advanced visualization algorithms enabled by today's powerful graphics processing units (GPUs), and an intuitive GUI, allow the researcher to quickly browse the space-time domain of their data to gain qualitative understanding and rapidly identify significant features. Once an ROI is identified, quantitative analysis, often far more computationally intensive than visualization, can be seamlessly undertaken across the much smaller sub-domain and at a resolution specified by the user to maintain interactivity. In this paper we briefly describe VAPOR, focusing on its data handling capabilities, present some of our experiences using it in the context of a relevant and challenging problem in astrophysical fluid dynamics, and discuss preliminary work on more aggressive data reduction techniques aimed at supporting future petascale applications.

2. VAPOR

While numerous *freeware* and commercial applications exist for analyzing and visualizing large, time-varying gridded data sets, they all have shortcomings that significantly curtail their usefulness in the exploration of high resolution simulation data. Open source visualization applications, such as Paraview (Ahrens et al. 2001) and Visit (Childs et al. 2005), and commercial applications such as CEI's Ensight, support advanced visualization algorithms appropriate for computational data sets, but are generally lacking in quantitative analysis capabilities. Their utility is, in the authors' opinion, targeted more towards visualization experts than scientific end-users, and most significantly, they demand specialized parallel computing resources to handle large data sets, resources that are often unavailable to the individual researcher. High-level data languages such as ITT's IDL and Mathwork's Matlab were designed with the scientific end-user in mind, support a rich set of mathematical operators suitable for quantitative data analysis, but offer only limited visualization capability and only minimal scalability, restricting their use to moderate sized problems.

The VAPOR software environment attempts to address these shortcomings, while at the same time providing an intuitive user interface and feature set. The design of the VAPOR GUI and functionality is guided by a committee of computational physicists to ensure that the tool meets the needs of the user community. VAPOR's advanced visualization capabilities enable the rapid identification of features or spatial-temporal ROIs in both scalar and vector data. Its seamless coupling to high-level analysis languages facilitates rigorous quantitative investigation and data manipulation on the reduced domain identified, and its hierarchical data representation scheme permits the investigator to make effective speed/quality trade-offs in order to maintain a high degree of interactivity. A more complete description of VAPOR's capabilities may be found in Clyne and Rast (2005) and Clyne et al. (2007).

2.1. Hierarchical data representation

The size of high-resolution data sets poses perhaps the greatest challenge to interactive analysis. Storing all the data on rotating media for random access is often unfeasible, necessitating time-consuming shuffling of the data between archival and disk-based storage. Even when sufficient on-line storage capacity exists, IO bandwidths are generally inadequate to support interactive processing without costly parallel IO systems. Constraints on the physical memory size of both the CPU and GPU, form the next bottleneck in the data movement hierarchy. Even if these issues were addressed, the processing capabilities of the CPU and GPU place limitations on the amount of data that can be processed interactively.

The strategy employed by VAPOR to overcome these difficulties is based on the assumption that, while some analysis may require access to the data at full resolution, many visualization and analysis operations are less sensitive to information loss. This suggests the use of hierarchical data representation. Simula-

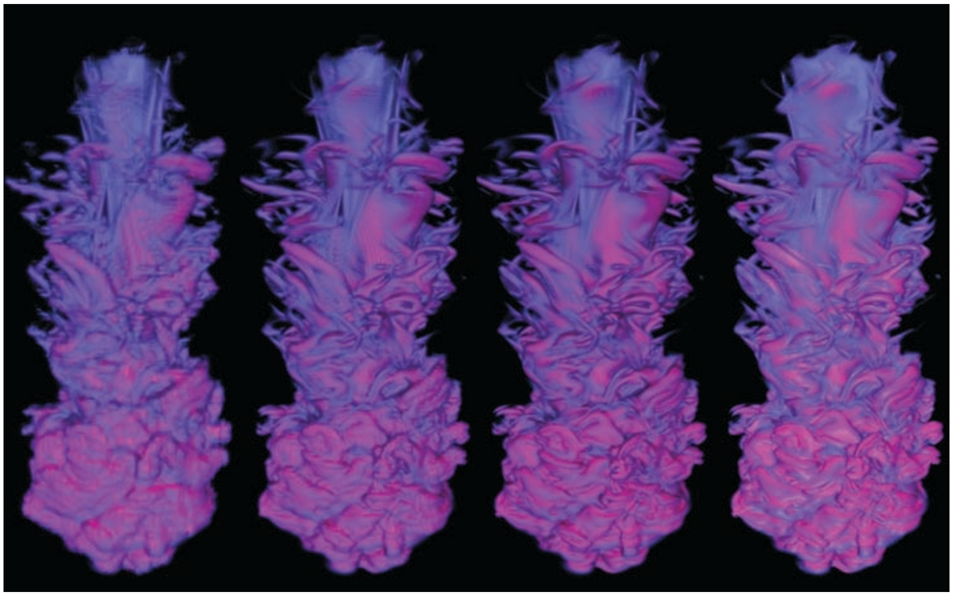

Figure 1. Compressible down flowing thermal plume, from Clyne et al. (2007), at progressively increasing resolution. Shown from left to right at $63^2 \times 256$, $126^2 \times 512$, $252^2 \times 1024$, and $504^2 \times 2048$, respectively. The individual plume images were all generated with a software ray caster using the same transfer function. Note how visual identification of even very fine structures is fairly insensitive to resolution degradation.

tion outputs are stored hierarchically, with each level in that hierarchy providing a coarsened approximation of the data at the preceding level. The original data may be accessed in their entirety, without loss of information, or an approximation of the original, sampled on a coarser grid, may be retrieved. At each level, coarsening halves the spatial resolution, yielding, for a three-dimensional data set, an eight-fold reduction in the data volume and a corresponding reduction in analysis and visualization resource demands. Grid coarsening currently employs a three-dimensional Haar wavelet transformation (see §4 for a discussion of future plans). Storing the Haar wavelet coefficients (Haar 1910) avoids the penalty of keeping multiple data copies, and allows reconstruction of the data at factor-of-two resolutions with only minimal overhead. The computational cost of the forward and inverse transforms are negligible compared to those incurred by reading or writing the data. Importantly, this hierarchical data access scheme permits the investigator to throttle the flow of data in accordance with the resources available, and thus control the level of interactivity. Users can browse coarsened representations of the data across the global spatial-temporal domain to identify features of interest. Once identified, the reduced domain may be examined at any level of detail up to the original resolution. Often both visual inspection (e.g. Figure 1) and numerical analysis are fairly insensitive to rather dramatic data coarsening (Clyne and Rast 2005), allowing considerable savings in computational overhead during the early exploratory stages of investigation when interactivity is most crucial. Subsequent verification of analysis results can be accomplished less interactively at full resolution if necessary.

2.2. Coupling visualization with quantitative data analysis

Acquiring a scientific understanding of large numerical data generally requires an interplay between highly-interactive, but qualitative visual-examination, and quantitative numerical analysis. Visualization can be used to identify salient features of the data, which then may give rise to hypotheses that can be validated or rejected by further study. Conversely, quantitative analysis of the data requires visualization to illuminate essential geometries and physical properties. Combining qualitative and quantitative data access is most effective when the process is seamless, enabling users to quickly transition back and forth between the two. For very large data sets the challenge is to maintain a sufficient level of interactivity throughout the process. VAPOR facilitates the interplay between visualization and analysis by combining data culling, through a combination of ROI isolation and hierarchical representation as described above, with seamless coupling to already existent analysis packages. The goal is to provide a tool for interactive analysis and visualization that does not require the user to learn a new analysis language.

VAPOR visualization is performed with intrinsic algorithms that are accelerated using the hardware GPUs, while numerical analysis currently employs ITT's *fourth-generation* language IDL. The coupling between IDL and VAPOR is facilitated by a library of data access routines, which allow IDL read and write access to VAPOR's wavelet-encoded data representation. This data access library could be employed by other scientific data processing utilities as well, and so the approach is readily generalizable to other analysis packages. The user can simultaneously maintain active VAPOR and IDL (or other analysis) sessions, visually identifying ROIs with VAPOR and exporting them to IDL for further study. Interactivity within IDL is maintained if the ROI is sufficiently small or if the operation is sufficiently well-behaved over coarsened approximations of the data (Clyne and Rast 2005). Newly calculated quantities can then be imported back into the existing VAPOR session for continued visual investigation. By repeating this process, very large 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 main advantage of coupling visual data investigation with an array-based data analysis language is the ability to defer expensive calculations of derived quantities until they are needed and then perform them only over sub-domains of interest. The time and space required for computing such variables in advance and across the entire domain can easily overwhelm the resources available, delaying or preventing altogether further analysis. Moreover, some quantities can only be computed with reference to the location of a flow structure and are therefore not in principle a priori computable (see §3). The coupling between VAPOR and IDL facilitates the calculation of derived quantities as needed over sub-regions of the domain, realizing considerable savings in storage space and processing time.

3. Example Application

Solar and stellar envelope convection occurs in the presence of ionization and recombination of the principal plasma constituents, notably hydrogen. These

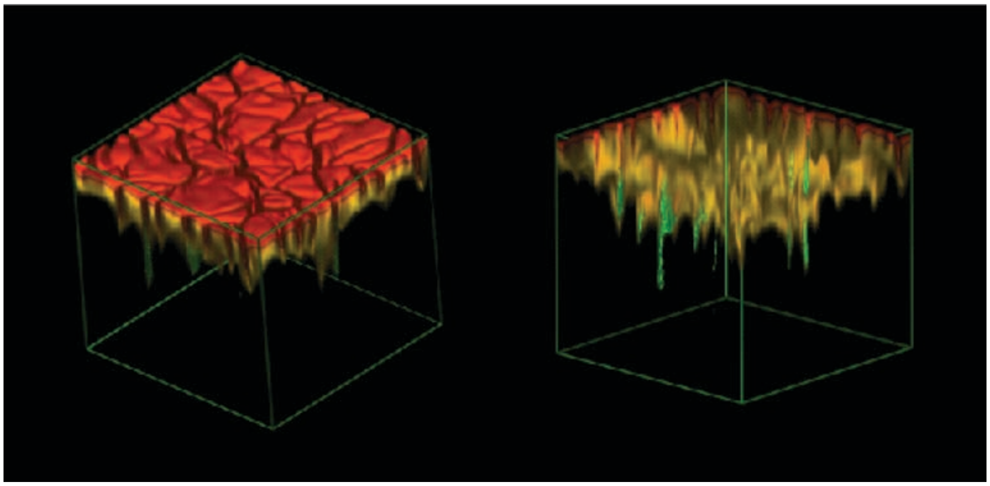

Figure 2. Two views of a single snapshot of the hydrogen ionization state in a simulation of compressible ionizing convection. Supersonic downflows initiate in the partially ionized region (orange/yellow color tones) of the domain. Sites of supersonic downflow are indicated with green color tones.

processes significantly influence both the transport properties and dynamics of the convective flows by modifying the particle number density, specific heat, and internal energy content of the gas (Rast & Toomre 1993). In particular, strong temperature fluctuations and enhanced buoyancy forces develop locally in the fluid wherever rapid changes in ionization state occur. The spatial and temporal scales of motion decrease and the flow velocities increase as the fraction of fluid participating in ionization increases. Buoyantly accelerated supersonic downflows, not seen in simulations of ideal-gas convection, result. We have examined these effects utilizing a series of idealized three-dimensional numerical experiments, which simulate compressible convection in a plane-parallel layer of fluid, a specified fraction of which is reactive (ionizing and recombining) single-atomic-level hydrogen. The fully nonlinear three-dimensional solutions were computed using a hybrid-pseudospectral code, employing spectral decomposition in the two horizontal directions and second-order finite-difference derivative approximations in the vertical. With the exception of the thermal diffusion term, treated implicitly, the solution was time advanced using an explicit two-level Adams-Bashforth scheme. (This differs from the three-dimensional compressible plume model of Figures 1 and 4 which employed a fully explicit second-order finite-difference code with third-order Runge-Kutta time stepping.) Figure 2 presents a snapshot from one of these simulations. Shown, from two perspectives, is the ionization state of hydrogen (orange/yellow tones) overlain by the sites of supersonic downflows (green). As the reactive fraction of hydrogen increases, the frequency of occurrence and vigor of the supersonic downflows also increases.

We used VAPOR's visualization capabilities to identify sites of supersonic downflow, and IDL to quantitatively examine their dynamics. As illustrated in Figure 3, sites of supersonic downflow are also those of very high vertical vorticity. The cores of the vortex tubes are evacuated, with centripetal acceleration locally balancing acceleration due to the radially (inwardly) directed pressure

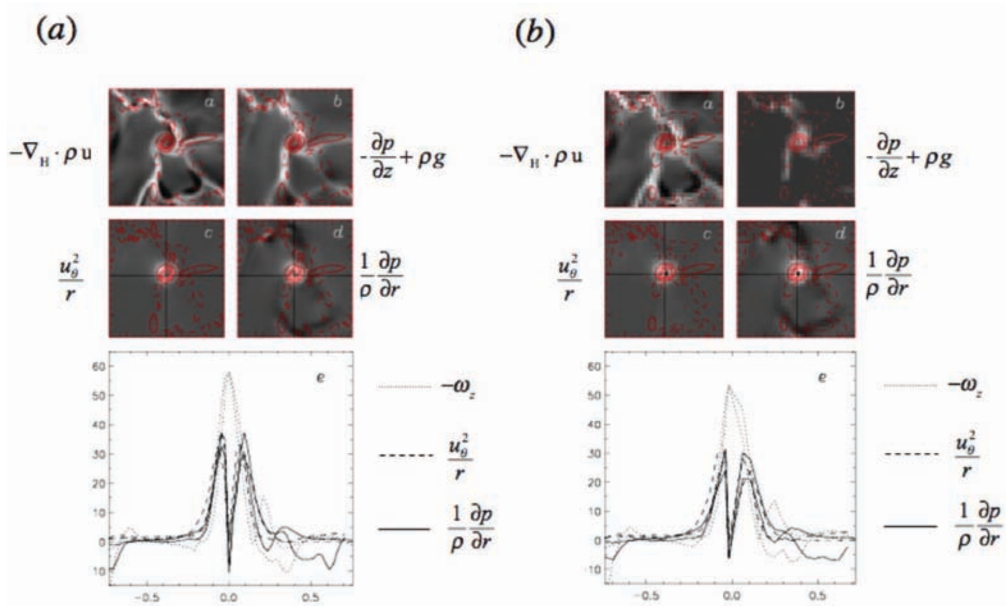

Figure 3. Force balance at a site of supersonic downflow showing the results of analysis undertaken at full (a) and half (b) resolutions. Vertical vorticity contours overlay grey scale images of key dynamical quantities: horizontal mass flux convergence, buoyancy force, centripetal acceleration, and radial pressure gradient. Plots below lie along vertical and horizontal cuts through the plume center at the depth of the grey scaled images.

gradient. Buoyancy forces that accelerate the fluid to supersonic speeds are maximum on the periphery of the tube where inflowing material accumulates. We note that in this case, not only is the ROI unknown before analysis begins, but at least two of the physical quantities of interest, the centripetal acceleration and the radial pressure gradient, can not be calculated globally as part of the batch simulation even if their need were anticipated. Computation of these quantities requires identification of the vortex center, itself part of the solution. Finally, comparison between Figures 3a and b illustrates that, while there are small quantitative differences, the same physical interpretation of these flows follows when conducting the above analysis at half spatial resolution (a factor of eight in data reduction). This is not unusual for both visual and analytic operations (Clyne and Rast 2005) since little information in most hydrodynamic solutions resides at the Nyquist frequency.

In this example application, a mechanistic not statistical understanding of the solution was desired. It is for this type of investigation that VAPOR is particularly well suited. With it, local flow properties, force balance, and stability can be interactively interrogated. Since the sites of interest are not a priori known, their interactive selection from the vast spatial-temporal data volume is essential. For very large data volumes, interactivity can be maintained only through a combination of ROI selection and multi-resolution access. In the example above, these together enabled a factor of 128 reduction in the data volume of a single snapshot cube before the analysis of an individual downflow site was undertaken. Much greater savings were achieved if one considers the cost of the interactive browsing required to identify individual events in the full data time series. Data reduction translates directly into enhanced interactivity and consequent scientific productivity.

4. Towards the petascale

The methods we have described and illustrated above have been shown to be highly valuable in the exploration of terascale data sets using only modest computing resources. We have successfully employed them on data from simulations with up to 1536^3 grid points (Clyne et al. 2007). However, interrogating data from forthcoming petascale applications will, we believe, require even more aggressive data reduction methods, and we again turn toward wavelet representations (Daubechies 1988; Mallat 1989; Sweldens 1995).

With a suitable choice of basis function, u, and without loss of information, we can represent a large class of functions, f, containing N samples as

$$f(t) = \sum_{n=0}^{N-1} a_n u_n(t). \tag{1}$$

If $u_n(t)$ are complex exponentials, the expansion is the discrete Fourier series. If $u_n(t)$ are wavelet functions, the summation is a wavelet series typically represented as a two-parameter expansion:

$$f(t) = \sum_{j=0}^{log_2 N-1} \sum_{k=0}^{j} d_{j,k} \psi_{j,k}(t), \tag{2}$$

where $d_{j,k}$ are the real-valued coefficients, j represents scale and k translation of the wavelet basis functions ψ.

We can compress our representation of f by creating a partial sum, reducing the number of terms in the expansion. Thus an approximation of f, $\hat{f}$, may be given by

$$\hat{f}(t) = \sum_{m=0}^{M-1} a_m u_m(t), \tag{3}$$

where $(M < N)$. The L^2 error between f and $\hat{f}$ is then

$$L^2 = \|f(t) - \hat{f}(t)\|_2^2 \tag{4}$$

For a wavelet expansion we can compress f by truncating j, limiting the scale (frequency) of f. This is precisely how VAPOR currently creates a hierarchical data representation. A nice property of frequency truncation compression is that the surviving coefficient parameters, j and k, are implicit and we do not need to store their values. However, better approximations may be achieved if we exploit the fact that for orthonormal basis functions, a property many wavelet families posses, the L^2 error of an approximation is given by:

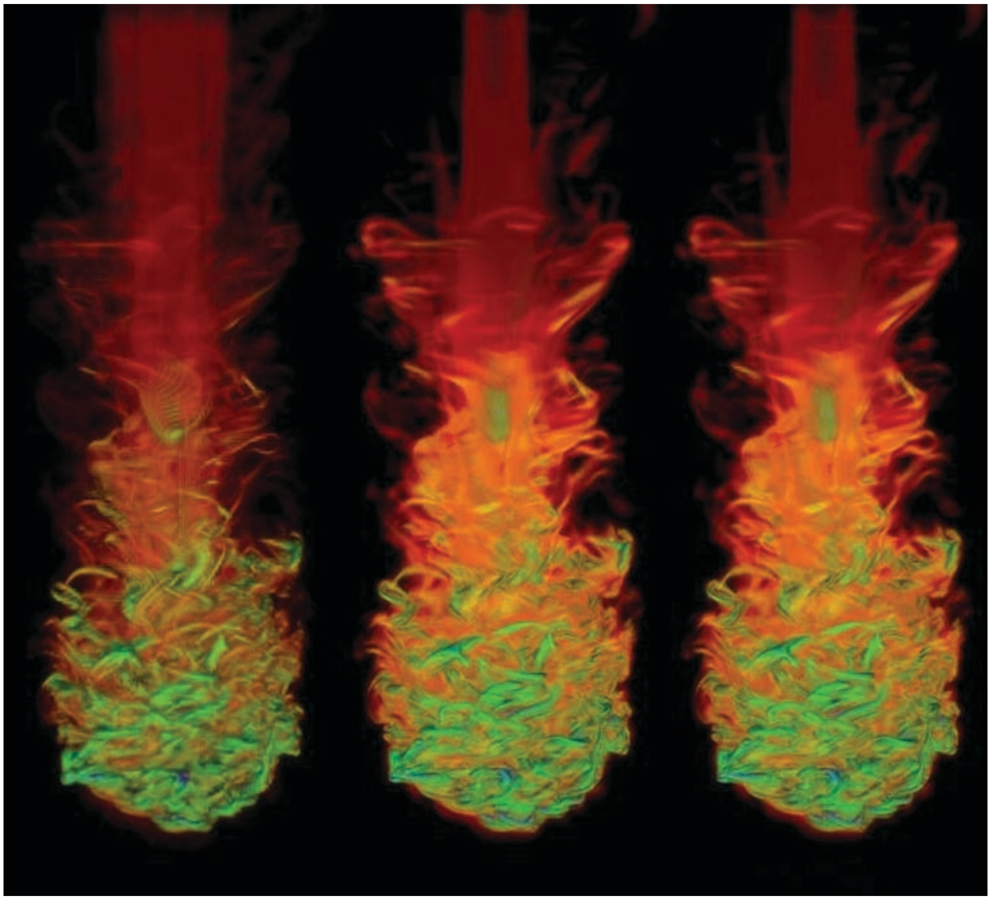

Figure 4. Comparison of plume data set (Clyne et al. 2007) compressed 64:1. Shown from left to right are: data compressed using frequency truncation; the original data; data compressed by prioritizing wavelet coefficients to minimize the L^2 error. The individual plume images were all interactively rendered on a GPU using the same transfer function. Note, this GPU based rendering process, used by VAPOR, differs from the early non-interactive software-based approach used to make Figure 1.

$$L^2 = \|f(t) - \hat{f}(t)\|_2^2 = \sum_{i=M}^{N-1} (a_i)^2 \ , \tag{5}$$

where a_i are the discarded coefficients. Thus an optimal compression, minimizing the L^2 error, can be obtained simply by discarding those coefficients with smallest absolute value. If the coefficients are zero, no error is introduced. It is this property that we can exploit to provide a more accurate representation of f for a given byte budget, or achieve comparable fidelity with scale truncation using a smaller byte budget.

Figure 4 compares images of the same data set but separately employing scale truncation and L^2 error minimization. The improved fidelity and finer grained control achieved when using the latter approximation does not come without a cost. Unlike the frequency truncation method currently employed, the coefficient prioritization method (L^2 error minimization) requires explicit storage of the location (i, j parameters) of the surviving coefficients, $d_{i,j}$. Thus a non-negligible storage overhead is introduced. Furthermore, identifying these coefficients can be a computationally expensive task. We are presently exploring solutions to both of these challenges.

Acknowledgments. This work was funded in part by the U.S. National Science Foundation under grant number ITR-0325934.

References

Ahrens, J., Brislawn, K., Martin, K., Geveci, B., Law, C., & Papka, M. 2001, IEEE Computer Graphics and Applications, 21(4), 34

Childs, H., Brugger, E., Bonnell, K., Meredith, J., Miller, M., Whitlock, B., & Max, N. 2005, in Proceedings of IEEE Visualization 2005, 191

Clyne, J., & Rast, M. 2005, in SPIE-IS&T Electronic Imaging, 5669, 284

Clyne, J., Mininni, P., Norton, A., & Rast, M. 2007, New Journal of Physics, 9, 301

Daubechies, I. 1988, Commun. Pure Appl. Math., 41, 909

Haar, A. 1910, Mathematische Annalen, 69, 331

Mallat, S. 1989, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (7), 674

Moore, G. 1965, Electronics, 38(8)

Rast, M/ P., & Toomre, J. 1993, ApJ, 419, 224

Ross, R., Felix, E., Loewe, B., Ward, L., Nunez, J., Bent, J., Grider, G., Salmon, E., & Bancroft, M. 2005, Technical report, NITRD

Sweldens, W. 1995, Proceedings SPIE, 2569, 68