In Situ MPAS-Ocean Image-based Visualization

James Ahrens
Los Alamos National Laboratory
Los Alamos, NM 87545
Email: ahrens@lanl.gov

Sébastien Jourdain
Kitware Inc
Santa Fe, NM 87505
Email: sebastien.jourdain@kitware.com

Patrick O'Leary
Kitware Inc
Santa Fe, NM 87505
Email: patrick.oleary@kitware.com

John Patchett
Los Alamos National Laboratory
Los Alamos, NM 87545
Email: patchett@lanl.gov

David H. Rogers
Los Alamos National Laboratory
Los Alamos, NM 87545
Email: dhr@lanl.gov

Patricia Fasel
Los Alamos National Laboratory
Los Alamos, NM 87545
Email: pkf@lanl.gov

Andrew Bauer
Kitware Inc
Clifton Park, NY 12065
Email: andy.bauer@kitware.com

Mark Petersen
Los Alamos National Laboratory
Los Alamos, NM 87545
Email: mpetersen@lanl.gov

Francesca Samsel
University of Texas at Austin
Austin, TX 78727
Email: figs@cat.utexas.edu

Benjamin Boeckel
Kitware Inc
Clifton Park, NY 12065
Email: ben.boeckel@kitware.com

***Abstract*—Due to power and I/O constraints associated with extreme scale scientific simulations, *in situ* visualization and analysis to will become a critical component to scientific discovery. The options for extreme scale visualization and analysis are often presented as a stark contrast: write files to disk for interactive, exploratory analysis, or perform *in situ* analysis to save data products about phenomena that a scientists knows about in advance. In this video demonstrating large-scale visualization of MPAS-Ocean simulations, we leveraged a third option based on ParaView Cinema, which is a novel framework for highly interactive, image-based *in situ* visualization and analysis that promotes exploration.**

I. Introduction

Extreme scale ($\geq 10^{15}$ FLOPS) supercomputing is here today or just over the horizon. These massive computing systems require scientist to change the manner in which they visualize and analyze simulation results. Until recently, data, at somewhat smaller scales, have been primarily stored or moved to another machine for post-processing. Moving forward, the standard post-processing techniques will not be able to effectively scale due to availability of storage and relatively slow increase in both storage and network bandwidth.

In situ techniques that visualize and analyze simulation data in simulation memory show a promising path forward. Up to now, *in situ* approaches typically operate on a predefined set of visualizations and analyses. In order for *in situ* visualization and analysis to meet scientist needs at extreme scale, it should: 1) preserve important elements of the simulations, 2) significantly reduce the data needed to preserve these elements, and 3) offer as much flexibility as possible for post-processing exploration.

In this video depicting a large-scale Model for Prediction Across Scales (MPAS) Ocean simulation, we leveraged a third option based on ParaView Cinema [1], which is a novel framework for highly interactive, image-based *in situ* visualization and analysis that promotes exploration. This interactive exploration - so important to scientific discovery - is supported in a meaningful manner (beyond simple animation of *in situ* products). ParaView Cinema effectively preserves the ability to interactively explore the same "operation space" defined at the start of the problem, so that data elements can be combined in much the same way they could in the original ParaView post-processing tool.

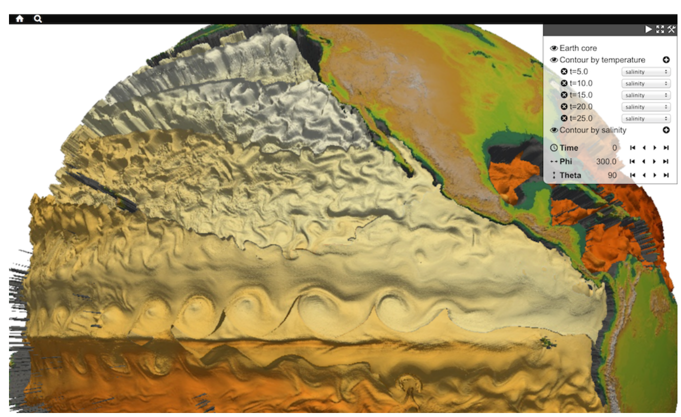

Fig. 1. MPAS-Ocean simulation indicating isosurfaces that represent the locations of water masses within the ocean of temperature colored by salinity.

Imagery is on the order of 10^6 in size, whereas extreme scale simulation data is on the order of $\geq 10^{15}$ in size.

This provides the scientist with nearly $\leq 10^9$ images to facilitate capture, curation, and exploration of important elements (features) in the simulation. ParaView Cinema is an exploration tool of a traditional *in situ* mode, but we sample the visualization and analysis parameter space, such as camera positions, operations, parameters to operations, etc. to produce a set of images stored in a data-intensive database.

The simulation depicted in the video is a run of MPAS-Ocean. MPAS is a new software framework for the rapid development of climate model components on unstructured grids. The grids may be quasi-uniform or variable density, on a spherical or rectangular domain, and may use quadrilateral cells, triangle cells, or Voronoi tessellations. MPAS variable density grids are particularly well suited to regional climate simulations. MPAS is developed cooperatively by Los Alamos National Laboratory and the National Center for Atmospheric Research for the purpose of decadal to century-long climate change research, as well as short-term weather forecasting.

In particular, we are demonstrating results from the MPAS-Ocean model component, which was publicly released in 2013 [2]. The visualization shows a high-resolution global ocean simulation with realistic topography, using 1.8 million horizontal 15 km-sized grid cells and 40 vertical levels. Ocean currents and eddy activity compare well with observations [3]. These simulations are typically run on 6000 processors, and achieve two simulated years per wall clock day. At this resolution, file output sizes present difficulties for traditional post-processing analysis and visualization workflows. In addition, using ParaView Cinema, the visualizations and analyses are derived from full-resolution data with high accuracy.

In Figure 1, the isosurfaces of temperature and salinity indicate the locations of water masses within the ocean. Water masses, with names like North Atlantic Deep Water and Antarctic Bottom Water, occur within specific ranges of temperature and salinity. These visualizations allow oceanographers to view the pathways and extents of these water masses, and compare them to observed climatology. Meandering ocean jets and eddies are visible as perturbations to these visualization objects.

II. Related Work

ParaView Cinema has a number of contributions, while building on a vast amount of previously disseminated results. Therefore, we quickly review related work for interactive exploration databases, creating new visualizations and metadata and content searching.

Interactive Exploration Database. The large image collection is produced from a structured sampling of time steps, visualization operators and camera positions. One option for the management of this large image collection is to compress them into movies. Chen et al [4] and Kageyama and Yamada [5] used this approach. ParaView Cinema extends these approaches by support compositing of images to create new visualizations as well as metadata and content searching. Tikhonova et al in [6], [7], [8] represent data as a collection of proxy images. Combining the approaches would support additional data compression, flexibility and exploration possibilities.

Creation of New Visualizations. Our approach supports the creation of new visualization by combining images using depth ordering information. Tikhonova et al [6] also supports the creation of new visualizations from their database using interpolation. When compositing images with opaque geometry results will be pixel accurate whereas with image interpolation some loss is expected. The idea for compositing visualization results evolved from the long history of parallel compositing techniques that enable scalable interactive visualization [9].

Metadata and Content Searching. Recent work by Subrahmaniam [10] identifies issues and future research directions for multimedia databases. Responding to ParaView Cinema's unique access patterns we created our own image database. When we create imagery *in situ* we save camera positions, time steps, details about the visualization operators, and statistics about the data. Directly storing analysis results and how they are created is important and encompasses our connection to provenance systems, such as VisTrails [11].

Image content queries support querying about the visual weight of the objects in the generated visualization. There are many approaches to calculating the statistics of the 2D projection of a set of 3D objects [12]. Related to our approach, Jun Tao et al [13] computes a collection streamline images, applies an image quality metric to select an optimal viewpoint. Our approach extends this work by virtue of being *in situ* and our ability to change our evaluation metric dynamically with a scientist generated query.

III. Approach

Exploratory, interactive post-processing visualization and analysis will still be essential for scientific discovery at extreme scale. First, the debugging use-case while developing new simulators requires exploratory visualization and analysis support. Second, the collaboration use-case for colleagues without equal computing resources need exploratory visualization and analysis support. Finally, the transition use-case for existing simulators demands a transition path from terascale/petascale ($10^{12}/10^{15}$ FLOPS) to extreme scale.

A. From Simulation Data To Image Database

ParaView Cinema utilizes ParaView [14], a modern visualization and analysis tool used around the world in post-processing for advanced modeling and simulation workflows of extremely large data sets leveraging distributed memory computing resources.

Fig. 2. The advanced selection interface enables the scientist to adjust visualization and analysis operators and how to sample the parameters space.

When starting an analysis based on ParaView Cinema, the scientist will define a desired set of visualization and analysis operators using a test data set and ParaView. Next, the scientist uses the advanced selection interface, shown in Figure 2, to make sophisticated prioritized decisions for the production of analysis output.

Using the ***Exploration settings*** section, the scientists determines: how often to perform *in situ* analysis, what visualization and analysis objects to create, and how to sample the visualization object parameter space. The ***Camera settings*** section allows the scientist to describe how to sample the camera space by defining a camera manager, and making appropriate selections for θ and ϕ sampling. In the ***Image settings*** section, the scientists select the image sampling (resolution) and the the image type (raw data TIFF format or compressed lossless data PNG format). The results of these choices are constantly updated in the ***Cost estimate*** section. The costs are reported as the number of images, the image size, the collection size, and the additional computational time.

As simulations progress towards extreme scale, scientists will operate in a constrained storage and bandwidth environment. Hence, the sampling of the parameter and camera space will require prioritization by the scientist to fit within a storage/bandwidth budget. ParaView Cinema was designed to operate within these constraints.

The output of the advanced selection interface, see Figure 2, is an *in situ* visualization and analysis `Python` script that implements the defined selections. For the *in situ* visualization and analysis, we leverage ParaView Catalyst, an open-source *in situ* (and other use cases) data visualization and analysis optimized `C++` library, which includes an API for `C`, `FORTRAN`, `Python` and `C++`. Catalyst is designed to be tightly coupled with simulation codes, and can be directly embedded into parallel simulation codes to perform scalable *in situ* data visualization and analysis at run time.

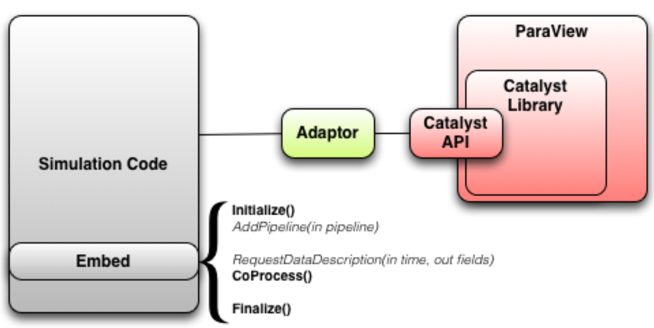

Fig. 3. ParaView Catalyst generalizes to many simulation codes through the use of adaptors. Adaptors translate the data structures of the simulation into VTK-based data structures the Catalyst library can process.

Finally, as the simulation runs, the image results are ingested into the database. By this we mean, that the metadata, image provenance (i.e. a search-able description of how the image was created – the simulation, the input deck, and which operators were applied), and the image uniform resource locator (URL). There is potentially no need to move the potentially large image data around in bulk.

B. New Visualizations and Content Querying

ParaView Cinema utilizes real-time compositing to create an experience reminiscent of interactively exploring the simulation resulting data itself. The visualization and analysis object compositing allows the scientist to reason about their simulation results from visualization space as opposed to the explorations offered from other image space rendering and sampling tools. But it's the user experience that significant enhances ParaView Cinema capabilities seemingly beyond those available in traditional post-processing tools. Complicated, computationally intensive visualization and analysis algorithms run in constant time, from a end-user perspective, using the precomputed images and analysis products.

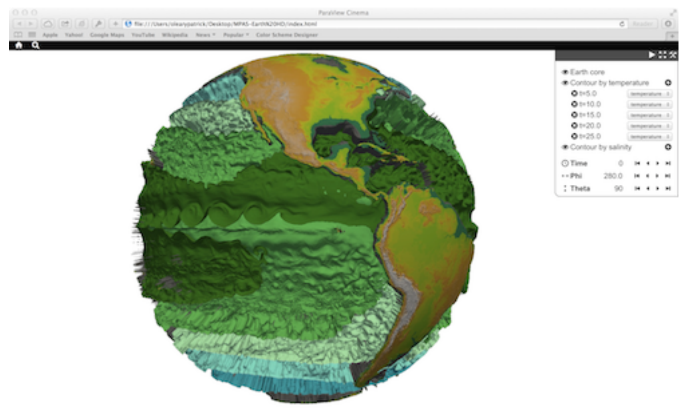

Fig. 4. The interactive exploration, enabled by the ParaView Cinema graphical user interface (GUI), pulling resulting image and analysis products from the interactive exploration database.

The interactive exploration database currently supports three elements and two modes of interaction, depicted in Figure 4. The three elements of interaction are *time*, (visualization and analysis) *objects*, and *camera*. The modes of interaction are: *animation*; where the interaction sequence through time, objects and camera; and *selection*, where the scientist select or query time, objects, and camera.

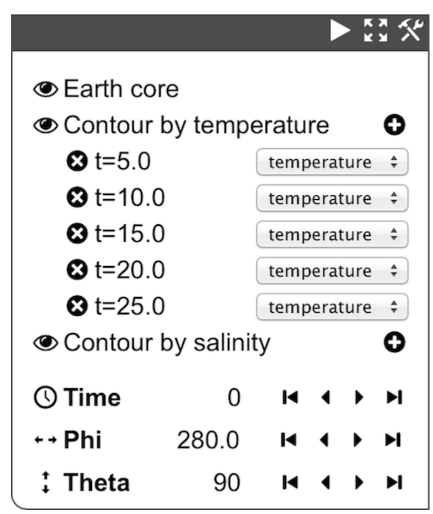

Fig. 5. The user interface of the scientist defined visualization pipeline for visualization object (Earth core, temperature contours, and salinity contours) compositing.

We can automatically display multiple objects from visualization space by selecting the associated (time, objects, and camera) image and analysis products from the database and composing them together. ParaView Cinema does not require the scientist to do this manually through a database query. Instead, the scientist uses a GUI, depicted in Figure 5, that emulates an application like ParaView to create an experience similar to exploring raw simulation data.

In Figure 5: the *Eyes* indicate visualization and analysis objects that can be interactively turned on and off; the *Plus* allows for selection and coloring of specific contour objects; and the *Time*, *Phi* and *Theta* controls allow for either selection or animation of these elements.

In addition, by leveraging the metadata associated with the image and analysis products in the interactive exploration database, ParaView Cinema allows the scientist to execute metadata queries (or browse visualization and analysis objects) to produce a prioritized sequence of matching results, as demonstrated in Figure 6. The metadata, produced by the *in situ* visualization and analysis `Python` script, includes data properties of the simulation data, such as histograms, as well as image properties.

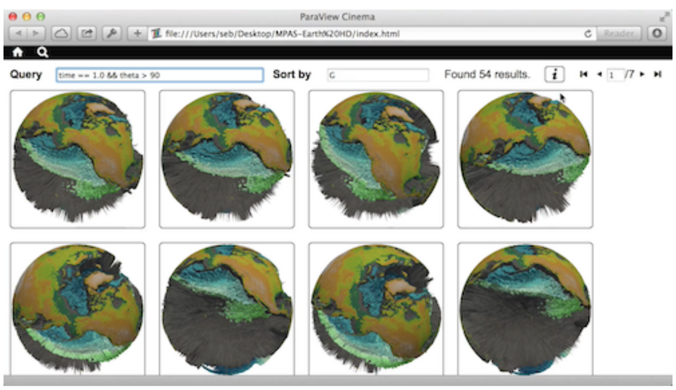

Fig. 6. The user interface of the scientist defined search pipeline for visualization object compositing.

The compositing infrastructure also makes it possible to perform queries that search on the content of the image in the database. A query could be formulated that matches on the quality of the view of a particular isosurface value. For example, in Figure 6 we queried visualization space for salinity isosurfaces of `34.0` colored by temperature at `time == 1.0`. We further specified results from the Northern Hemisphere (`theta > 90`). Finally, we sorted the results, in decreasing order, by the percentage of pixel coverage representing the salinity contour.

IV. Conclusion

ParaView Cinema is a novel framework for an image-based approach to extreme scale *in situ* visualization and analysis It couples visualization and analysis outputs with an image database query method to enable interactive exploration and metadata browsing. As implemented, ParaView Cinema meets the overarching goals for *in situ* visualization and analysis systems: 1) to preserve important elements of the simulations, 2) to significantly reduce the data needed to preserve these elements, and 3) to offer as much flexibility as possible for post-processing exploration.

Acknowledgment

This work was funded by Dr. Lucy Nowell, ASCR Program, Office of Science and ASC, Department of Energy (DOE). Patrick O'Leary and Sébastien Jourdain were also funded by a DOE Office of Nuclear Energy Fast Track SBIR award DE-SC0010119.

References

[1] J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D. Rogers, and M. Petersen, "An image-based approach to extreme scale in situ visualization and analysis," in *IEEE Supercomputing 2014 (SC'14)*. IEEE Computer Society, 2014, p. to appear.

[2] "MPAS," Online, June 2014. [Online]. Available: http://mpas-dev.github.io

[3] T. Ringler, M. Petersen, R. Higdon, D. Jacobsen, P. Jones, and M. Maltrud, "A multi-resolution approach to global ocean modeling," *Ocean Modelling*, vol. 69, pp. 211–232, 2013.

[4] J. Chen, I. Yoon, and W. Bethel, "Interactive, internet delivery of visualization via structured prerendered multiresolution imagery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 302–312, Mar. 2008.

[5] A. Kageyama and T. Yamada, "An approach to exascale visualization: Interactive viewing of in-situ visualization," *Computer Physics Communications*, vol. 185, no. 1, pp. 79–85, 2014.

[6] A. Tikhonova, C. D. Correa, and K.-L. Ma, "Visualization by proxy: A novel framework for deferred interaction with volume data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1551–1559, 2010.

[7] A. Tikhonova, H. Yu, C. D. Correa, J. H. Chen, and K.-L. Ma, "A preview and exploratory technique for large-scale scientific simulations," in *Proceedings of the 11th Eurographics Conference on Parallel Graphics and Visualization*, ser. EG PGV'11. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011, pp. 111–120.

[8] A. Tikhonova, C. D. Correa, and K.-L. Ma, "An exploratory technique for coherent visualization of time-varying volume data," *Computer Graphics Forum*, vol. 29, no. 3, pp. 783–792, 2010.

[9] K. Moreland, W. Kendall, T. Peterka, and J. Huang, "An image compositing solution at scale," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '11. New York, NY, USA: ACM, 2011, pp. 25:1–25:10.

[10] V. S. Subrahmanian and S. Jajodia, *Multimedia Database Systems: Issues and Research Directions*, 1st ed. Springer Publishing Company, Incorporated, 2012.

[11] J. Freire, D. Koop, F. Chirigati, and C. Silva, "Reproducibility using vistrails," *Implementing Reproducible Computational Research*, 2014.

[12] N. Elmqvist and P. Tsigas, "A taxonomy of 3d occlusion management for visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 5, pp. 1095–1109, 2008.

[13] J. Tao, J. Ma, C. Wang, and C.-K. Shene, "A unified approach to streamline selection and viewpoint selection for 3d flow visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 3, pp. 393–406, 2013.

[14] A. H. Squillacote, D. E. DeMarle, J. Ahrens, C. Law, B. Geveci, K. Moreland, and B. King, *ParaView Guide*, 1st ed. Kitware, Incorporated, 2007.