Agenda for Combustion Visualization

Principal Investigators: W. Bethel, LBNL/ICSD; Terry Ligocki, NERSC

Requested Funding: $100K

Preface

Bethel and Ligocki were previously awarded an LDRD for FY99. That project, titled Visualization Tools Interoperability, was assigned project number 3659-37. Due to circumstances beyond control, that project was judged unlikely to succeed, and the project described herein outlines a suitable, if not better, replacement project.

1.0 Objective

There are two overall goals for this project. We propose a process of discovery for the purpose of identifying and defining a research and development agenda for combustion visualization. The discovery process will involve close collaboration with the Combustion program at NERSC/LBNL, and will dovetail with the effort proposed in a related FY99 LDRD (Bell, Brown and Colella). With such an agenda, we also propose to deliver by fiscal year's end a set of software tools that address some of the high-priority visualization needs of the NERSC/LBNL Combustion program. These include tools that are capable of embedded boundary geometry, as well as tools that are capable of processing large data sets.

2.0 Approach

The combustion program at LBNL/NERSC consists of two groups. The Applied Numerical Analysis Group (ANAG) focuses on combustion modeling with embedded boundary conditions. The Center for Computational Science and Engineering (CCSE) focuses on turbulence modeling. Both groups use box-structured, adaptive mesh refinement (AMR) tools for computation and visualization. Each group has substantial visualization needs that are not met by commercially available software.

Before undertaking development, a careful assessment of combustion-specific visualization needs will be defined. Such an assessment involves close cooperation with the combustion program. With appropriate personnel from the combustion program, we will jointly prepare a detailed assessment of the short and long term visualization needs, as well as to define priorities that are appropriate for the program and are consistent with the time and budgetary constraints of this proposal. The assessment is expected to be an evolving list of requirements, and is subject to revision as dictated by the needs of the program.

Based upon the assessment, a set of visualization tools will be produced that are specific to each group's needs. While both ANAG and CCSE share much in common, specificities of their research domains require a problem-specific set of visualization tools. Obviously, it will be beneficial to both programs if such a set of tools overlaps to a large degree. ANAG's embedded boundary conditions give rise to specific needs that are not present in CCSE (at this time). Likewise, CCSE visualization tools are intended to be deployed on large distributed memory architectures. This environmental concern is not a high priority for ANAG (at this time).

3.0 Research Plan

The central themes of our effort are to define the short and long term visualization needs of the LBNL/NERSC combustion programs, and to deliver an initial set of software tools that address those needs.

In order to define the visualization needs, ongoing collaborative exchanges between the Visualization and Combustion programs will serve to broaden understanding of the problem scope. Our approach will be to follow a traditional "systems analysis" model. Such a model involves "the customer" throughout the problem definition, design, testing and delivery phases of the project. Appendix A of this proposal contains detailed technical information specific to each of the ANAG and CCSE problem domains.

From our perspective, central to CCSE visualization needs is the theme of remote visualization of large datasets. These data sets typically consist of a small number of variables per grid cell, box-structured adaptive grids, time-varying data, and data that is "too large" to fit onto a desktop workstation for the purposes of visualization. Most of CCSE's large-scale computations are performed on large distributed memory architectures, such as the T3E at NERSC. Therefore, CCSE will derive benefit from visualization tools that can be deployed on distributed memory architectures and can display the resulting images onto desktop workstations.

In contrast, ANAG's focus is on adaptive meshes with embedded boundaries. In addition to the usual array of visualization techniques for scalar and vector data, visualization of the embedded boundary and related information presents special challenges. In some instances, it is possible to represent the boundaries using a polygonal mesh. In other instances, a procedural model is evaluated at simulation time. Often, the boundary changes as the simulation evolves. We intend to assess a number of different approaches for embedded boundary visualization. These will include a polygonal representation, visualization-time procedural evaluation, and image-composition/layering techniques where appropriate. During the process of needs assessment, we expect that additional techniques will be proposed for evaluation. Included with the embedded boundary representation are additional variables that include flux, surface area and grid topology.

Given the short duration of the proposed work, our top priority is to provide a clear assessment of the needs for combustion visualization at LBNL/NERSC. As that process unfolds, software tools that can be used for ANAG embedded boundary visualization will be prototyped; ANAG has urgent needs for such tools. Similarly, software tools that implement remote visualization of CCSE data will be prototyped. We view this as an exploratory effort that will result in a long-term research and development agenda for combustion visualization.

4.0 Potential Results and Significance

In the field of visualization, Berkeley Lab has an established reputation as being a leader in what could be called "practical" scientific visualization. Our emphasis in visualization over the years has been to foster close relationships between the scientific researchers and members of the LBL Visualization Group. The close working relationship promotes synergy, often resulting in novel but always practical tools. For the purpose of longer-term proposals, such as SSI, it is crucial to propose visualization research that is couched in the context of ongoing scientific research. Interestingly, in the pre-SSI proposals related to combustion (e.g., Argonne's NGI proposal) we find a marked absence to combustion-specific visualization, such as embedded boundary visualization. Our proposal is oriented specifically towards facilitating scientific research, and has been directly shaped by close contact between the LBL/NERSC Visualization Group and representatives from both ANAG and CCSE. The medium-term roadmap for visualization research for combustion we propose to define will have direct bearing upon our contribution to SSI and related initiatives.

The visualization needs of the combustion programs at LBNL/NERSC, particularly ANAG, are urgent and immediate. These groups cannot afford to wait for NGI or SSI funding. Our proposed effort is designed to accommodate a short term delivery of a base set of tools. The tools will be part of a longer term roadmap for a visualization effort for the combustion program.

The terms "remote visualization" and "large data visualization" remain stubbornly positioned on visualization research topic lists. The work we propose addresses both these topic areas. The CCSE data sets are projected to scale up to the petabyte range in the near future. Not only does data of this size qualify as "large," the visualization will by necessity be performed remotely, then delivered to the desktop. Such an effort is consistent with, if not exemplary of the type of service provided by NERSC as a national facility.

5.0 Relationship to Other Berkeley Lab Projects

As indicated in the preface, the work described in this proposal supplants that awarded as LDRD project 3659-37.

FY99 LDRD by Bell, Brown and Colella titled Computational Modeling of Turbulent Combustion Dynamics.

6.0 Proposed Organization

Dan Graves will serve as the primary contact between Bethel/Ligocki and the ANAG group. Graves' effort is provided for under separate funding. Likewise, effort for the CCSE contact, Vince Beckner, is provided for under separate funding.

Bethel and Ligocki will lead the effort to discover and define the visualization research agenda, and will jointly develop visualization tools for both CCSE and ANAG.

7.0 Budget

7.1 Effort

Approximately 1/2 FTE, split evenly between Bethel & Ligocki.

7.2 Facilities

This project will make use of existing computer hardware and software at LBNL. No new equipment nor software purchases are required.

7.3 Travel

The original proposal requested $10K for travel to ASCI DMF meetings. We would like to keep up to date with that group, even though we are not actively working on that project. In addition, there is an upcoming meeting in Salt Lake City, sponsored by DOE and NSF, that addresses the topic of large data visualization. We will attend that meeting, and use funds from the existing travel budget from 3659-37 for that purpose.

8.0 Appendix A

This Appendix provides technical details for the combustion visualization effort.

8.1 ANAG Priorities

One priority cited by ANAG is the issue of data file formats. ANAG would like to adopt a file format that supports data produced by their simulations, but that is “supported” by a third-party. The ANAG modifications to CCSE’s BOXLIB are not third-party supported. We have proposed evaluating some common and supported self-describing formats, such as NetCDF and HDF. These alternatives are appealing as these formats are supported by ongoing efforts (NetCDF at NCAR, and HDF at NCSA). HDF development is being funded in part by the ASCI DMF committee. The newer version of HDF will be used as one of several storage formats for the ASCI DMF Vector Bundles presentation layer. In addition, a version of HDF that is layered upon MPI-IO has been developed by NCSA, and a preliminary version of “parallel HDF” has been placed on the NERSC T3E (Cray recently delivered MPI-IO for the T3E). The fundamental question to be answered is “can one of these formats support the needs imposed by adaptive meshes and embedded boundary conditions, and if so, at what cost?”

CCSE has developed a visualization tool called *AMRvis* used for the visualization of combustion data from adaptive grids. Extending AMRvis to accommodate embedded boundary visualization has been deemed impractical. Among other reasons, AMRvis is a production visualization tool that has been highly optimized for use with the types of grids generated by CCSE. Nonetheless, ANAG has stated that they would like to have “visualization capabilities similar to those in AMRvis.” ANAG requirements above and beyond what is available in AMRvis include visualization of “embedded boundaries” in adaptive grids. In some instances, these boundaries are known prior to the simulation run, and can be represented explicitly with geometric models. In other instances, the boundaries are procedural, or computed, and can be derived from the simulation data. ANAG has tools that can extract geometric embedded boundary information from the adaptive mesh simulation results. Combining the display of embedded boundary information with basic visualization capabilities is a high priority.

A core set of features for a visualization environment suitable for use by ANAG/CCSE include essentially two components. The first component is a framework for deploying visualization algorithms. These include things like false-coloring of data, volume rendering with user-defined opacity and color transfer functions, and so forth. These are the tools that change raw data into images. The second component of a visualization environment is the notion of data access and manipulation.

Additional open issues include:

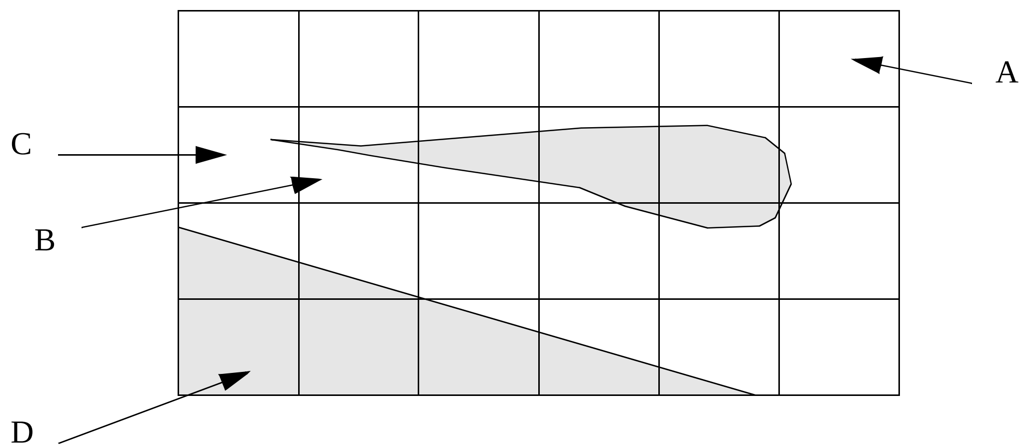
- Does the visualization tool run on a local workstation, or remotely?

- Are images delivered using Web-based technology, or do we restrict our consideration only to the case of creating and displaying images with local graphics hardware?
- What is the best design and implementation strategy that promotes the exchange of software and data between ANAG and other groups, including industry and academic research partners?

8.1.1 ANAG Specific Visualization Effort

The following list outlines a number of visualization techniques that are commonly used or have been identified by ANAG as being high priority for their work:

- Visualization of "VOFs". A "VOF" is a "volume of fluid." Grid cells may contain zero or more VOFs (see the figures below). "Regular cells" contain zero or one VOFs, and no embedded boundary conditions. "Irregular cells" contain one or more VOFs, along with embedded boundary conditions. The problems posed by VOF visualization doesn't readily lend itself to a quick solution with other visualization techniques, for it is not entirely clear how to display VOF information in a way that is meaningful.



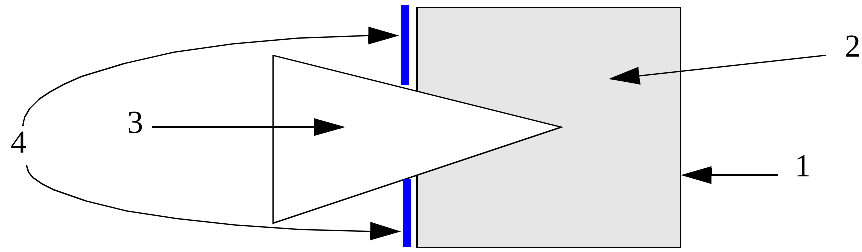
A - A *regular* cell, contains one VOF and no embedded boundary.

B - An *irregular* cell. Contains two VOFs and two embedded boundaries.

C - An *irregular* cell. Contains one VOF and two embedded boundaries.

D - A region in which no data exists.

Expanding one of the cells in the previous image, we have the following:



1. The grid cell.
2. The area/volume fraction of the cell occupied by data
3. The region of the cell that contains no data. Note that this void region may extend into adjacent cells.
4. Percent of cell boundary (face) that fluid can move through into adjacent cells.

- Related to VOF visualization is the display of embedded boundaries.In some instances, the boundary information may lend itself to a pre-simulation polygonal representation. In other instances, the embedded boundary may be pseudo-procedural or computed along with the simulation. In the former case, it is a fairly straightforward task to integrate the display of raw geometry data in conjunction with display of visualization output. In the latter case, it will be necessary to integrate pre-existing tools within ANAG to generate geometry from the adaptive grid decomposition. It is important to provide some "wiggle room" in this area to allow for exploration of alternative methods of embedded boundary visualization.
- Grid visualization. Seeing the geometry of the adaptive grids is an integral part of the combustion research process. The adaptive grid geometry is evolved over time by the simulation itself. Seeing the grid geometry can be an aid in debugging adaptive mesh refinement models.
- False coloring raw data. Raw data is transformed into color (and opacity) with a color table and a transfer function.Two dimensional slices of data are transformed into pixel maps; three dimensional volumes of data are transformed into semi-transparent "clouds."
- Contouring, both in two and three dimensions.
- Glyph-based visualization. For each grid cell, data point, or other feature of interest, generate and place an icon in space. This type of visualization is used for direct vector field visualization (shows vector magnitude and direction), tensor fields (add twist), and, interestingly enough, information visualization (economics, demographics, and other forms of non-intrinsically-spatial data).
- Raw data display. Researchers from the combustion program want direct access to the underlying raw data. The word "spreadsheet interface" has been used in this context.

A non-trivial stumbling block we face in deploying these techniques is the fact that all such visualization and surface representation techniques must work directly with the adaptive grid representation. In the past, we have avoided this problem by "flattening" the grid to its highest resolution. While that approach is tractable for small data sets, and enables data visualization using commonly-available tools, it fails in two regards. First, all notion of the embedded boundary condition is lost. Second, as data set sizes increase, it will be impractical to explode an already large dataset to something larger by an order of magnitude.

8.1.2 ANAG Data Manipulation

Hand in hand with data visualization is data access and manipulation. The success of previous visualization packages, such as AVS and Khoros, lies in the fact that those systems provide a framework in which it is a straightforward and enjoyable matter to “plug in” user code implementing a new visualization technique, while using the infrastructure of the environment to access and manipulate data. We have experimented in the past with the use of these packages for the visualization of ANAG data. The limitations discussed in the previous section have proven to be an insurmountable obstacle, thereby rendering the use of these commercial packages impractical as well as of limited use.

The following list provides an overview of the types of data manipulation features desired:

- Slicing. Given n-dimensional data, extract an m-dimensional hyperslice, where $m < n$. Simple slicing of data implies taking complete sections of data along orthogonal axes, either in space or in time, or both.
- Subsetting. In contrast to slicing, which is a dimension-reducing operation, subsetting preserves dimension but reduces size. Subsetting may be spatial - “give me everything inside this region,” or may be temporal - “give me everything between the beginning and *t=100*” or may be some combination of the two.

These data reduction operators are further complicated by the adaptive nature of the underlying grid. We expect that novel methods of data manipulation will be suggested and evaluated during the course of the effort.

8.2 CCSE Priorities

Remote visualization of large data sets using volume rendering techniques is the top priority for CCSE. By “remote visualization,” we mean the process of computing an image on one machine (or machines), and image display on another machine. Remote visualization is one solution to the difficulties presented by ever-increasing data set sizes. In some cases, it is impractical to transfer the entire data set to the local workstation. To state the problem differently, we have the choice of moving the entire data set, O(*N**3*), or an image of the projected volume, O(*N**2*). The difficult problem, from the visualization perspective, is can interactive rendering rates be achieved on a distributed memory architecture. From a networking perspective, the difficult problem is moving 10 or more frames-per-second of image data across a WAN from the rendering source to the desktop workstation. That issue forms the basis for the LBNL-NGI proposal.

The deployment plan for CCSE includes R&D of visualization tools that are capable of rendering images of large data sets at interactive rates. The implementation plan calls for providing suitable interfaces so that these tools can be integrated with AMRvis. For this reason, the data manipulation issues cited above for ANAG do not apply in quite the same way here since AMRvis implements many of these data manipulation tools.