An Analytical Framework for Particle and Volume Data of Large-Scale Combustion Simulations

Franz Sauer
University of California, Davis
1 Shields Avenue
Davis, California 95616
fasauer@ucdavis.edu

Hongfeng Yu
University of Nebraska, Lincoln
1400 R Street
Lincoln, Nebraska 68508
yu@cse.unl.edu

Kwan-Liu Ma
University of California, Davis
1 Shields Avenue
Davis, California 95616
ma@cs.ucdavis.edu

ABSTRACT
This paper presents a framework to enable parallel data analyses and visualizations that combine both Lagrangian particle data and Eulerian field data of large-scale combustion simulations. Our framework is characterized by a new range query based design that facilitates mutual queries between particles and volumetric segments. Scientists can extract complex features, such as vortical structures based on vector field classifications, and obtain detailed statistical information from the corresponding particle data. This framework also works in reverse as it can extract vector field information based on particle range queries. The effectiveness of our approach has been demonstrated by an experimental study on vector field data and particle data from a large-scale direct numerical simulation of a turbulent lifted ethylene jet flame. Our approach provides a foundation for scalable heterogeneous data analytics of large scientific applications.

Categories and Subject Descriptors
E.m [**Data**]: Miscellaneous – *Particle Data, Volume Data*; J.2 [**Computer Applications**]: Physical Sciences and Engineering – *Chemistry, Physics*.

General Terms
Management, Performance.

Keywords
Feature extraction and tracking, data transformation and representation, scalability issues.

1. INTRODUCTION
Detailed combustion simulations are essential in developing next generation engines of high efficiency. Direct numerical simulations (DNS) can capture and describe the key turbulent combustion chemistry interactions. Sandia National Laboratories scientists developed S3D, a massively parallel solver, to solve the DNS governing equations [1]. The DNS solution from S3D is obtained for field data on a fixed spatial grid, and is also recorded in terms of particle data. The field data, corresponding to the Eulerian specification of the flow, focuses on the spatial locations through which the fluid flows over time [2, 3]. The particle data, corresponding to the Lagrangian specification of the flow, records the trajectory of massless tracer particles through space due to advection only [4]. Studying turbulent combustion from both the Eulerian and Lagrangian viewpoints can lead to new understandings of important processes, such as autoignition, for improving efficiency in energy conversion.

Researchers have developed several algorithms and systems to facilitate combustion scientists in exploring and analyzing the field and particle data from the turbulent combustion simulations. Examples include volume visualization [5], topological analysis [6], statistical analysis [7], and trajectory visual analytics [8]. These techniques have successfully helped scientists extract important statistical and structural information from large combustion data, and gain new insights into complex physical and chemical processes.

Given these sophisticated analytic tools, scientists also desire to further collate the studies from both the Eulerian and Lagrangian viewpoints to possibly obtain a deeper understanding of fundamental combustion processes. In a typical scenario, scientists can first identify volumetric features using segmentation and/or classification methods. Feature selection usually implies certain statistical or topological characterization of the field data. Based on spatial identified features, scientists may also want to examine particles within the volumetric regions of the features. The statistical information, such as the conditional average, can be collected for the selected particles to describe the relationship between variables. In addition, the selected particles can be assembled into a meaningful time series for possibly connecting volumetric features over time, given that feature tracking is conventionally difficult to achieve without high temporal sampling rate of field data. Hence, this combined capability can potentially become a powerful tool that enables scientists to investigate the interplay between knowledge and hypotheses developed from field data and particle data.

A scalable realization of analysis tasks involving both data types is very challenging. First, the field data and the particle data have completely different data representations deployed by simulations. The Eulerian field data records the values at the fixed spatial locations, while the Lagrangian particles are advected spatially with the flow velocities. Second, given a distributed environment, it is non-trivial to select particles within volumetric regions, and vice versa. It is difficult to design a uniform data partitioning and distribution scheme to favor the

UltraVis'13, November 17, 2013, Denver, CO, USA

http://dx.doi.org/10.1145/2535571.2535590

operations for both particle and field data. Third, the analysis tasks on both data are typically based on functions of primary variables, and we need to evaluate these functions at runtime. However, the output of these functions is not known *a priori*, which makes it hard to apply indexing techniques [9, 10] to preprocess the data and determine access patterns. Thus, due to the heterogeneous nature of the data and a priori unknown functions, it is fundamentally difficult to design a unified and scalable scheme favorable to both Eulerian and Lagrangian data.

In this work, we study the issue of analyzing heterogeneous combustion data and make the following contributions:

- We introduce a parallel framework to support analytics for both field data and particle data while minimizing computation overhead.
- We enhance particle queries and make it possible to select particles within irregular spatial regions in an efficient way.
- We develop a new parallel segmentation method that can efficiently extract sub-features from highly interconnected topological classifications.

We demonstrate our framework with real-world combustion data and show that the resulting analyses and visualizations can provide scientists with new insights into their large data in an efficient manner.

2. RELATED WORK

A set of techniques have been developed to enable analysis and visualization of large particle data. Ma et al. [11] presented hardware-assisted rendering techniques for interactive visualization of the data generated from particle accelerators. They generated compact volume representations for the large scale particles with high density, and used fast volume rendering to depict different regions of the particle data by adjusting the transfer functions. Jones et al. [12] presented a data exploration system that supports visual exploration of particle data from gyrokinetic simulations. They developed a user interface that enables interactive selection of particles with multiple ranges through parallel coordinates. A similar system has also been developed by Rübel et al. [13] for studying particle data from a laser wakefield accelerator simulation. Woodring et al. [14] proposed a simulation-time random sampling of a large-scale particle cosmological simulation, which can amortize the cost of post-processing analysis and visualization. Wei et al. [8] developed a visual analysis method based on time-series curve clustering to understand particle behaviors in both the phase and physical spaces for combustion simulations. They also developed a parallel line clustering method to process large particle trajectory data using multiple GPUs and CPUs [15].

Researchers have also developed feature-based analytics frameworks to study field data of large scientific simulations. Bennett et al. [7] presented a framework for feature-based statistical analysis of large-scale combustion simulations. The framework first extracts volumetric features using the topological analysis method and then computes conditional statistics per feature to study correlations with scalar quantities. Glatter et al. [16] developed a textual pattern matching approach for specifying and identifying general temporal patterns. They provided a high-level language for querying multivariate patterns from large-scale data. Kendall et al. [17] created a framework for quantitative analysis of flow features using geometric and other derived attributes for interactively exploring large-scale flow data sets. Gaither et al. [18] developed a system for automatic feature detection, extraction, and classification, which facilities the characterization of coherent structures over time for large simulations.

Both particle and feature-based methods are critical to describing and understanding large scientific data. However, little effort has been put into the enhancement of analysis and visualization using both data types. This requires us to re-examine the characteristics of particle and feature-based algorithms and develop scalable methods to support both particle and field data.

3. OUR APPROACH

In this section we describe our framework and its various components. We start by discussing an overview of the framework followed by the classification and segmentation of topological features from the vector field data. Next, we discuss the extraction and analysis of particles from the Lagrangian data. Lastly, we discuss how the framework can be utilized both for particle extraction based on segmented features and for feature segmentation based on queried particles.

3.1 Overview

Our framework focuses on combining information from both the vector field data and particle data into useful analytical results. Processing the vector field data consists of two steps in our combustion application. First we classify the vector field data into various topological types. The field data is now represented as a set of voxels, where each voxel has a topological type associated with it. Next, we use the classified voxel data and perform a region growing step to segment features from the voxel data. On the other hand, like in most existing work, the particle data only goes through one processing phase, which extracts a subset of particles according to some criteria.

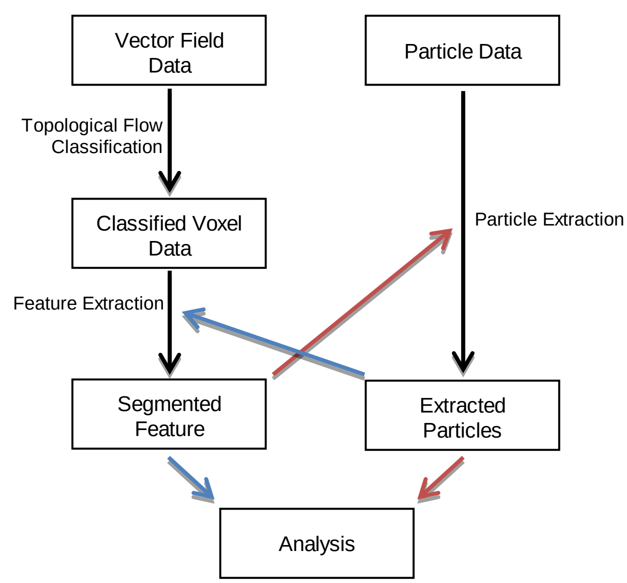

Figure 1. Workflow of our framework; the black arrows represent processing done only using one type of data, the red arrows represent feature-based particle query, and the blue arrows represent particle-based volume feature query.

Our framework combines the processing steps of these two data formats using two possible variations. The first is feature-based particle query, in which we use a segmented feature from volume data as an input to the particle extraction step and only extract particles found within the feature. Analysis can then be done on the extracted set of particles to gain a better understanding of the properties of the segmented feature.

The second variation is particle-based volume feature query. In this method a set of particles are extracted using a range query on one or more of its primary variables (or functions computed at runtime). These particles are then used as an input to extract features that correspond to a particular range query and encompass the particles.

Figure 1 shows the workflow for each variation, where the conventional operations are shown in black and new operations enabled by our framework are shown in red and blue. We elaborate on each portion in the sections that follow.

3.2 Topological Feature Extraction

Turbulent flow plays an essential role in modulating combustion processes. Topological feature extraction can facilitate scientists studying the interaction between the flow structures and the flame, and test hypotheses of important combustion processes, such as flame stabilization.

3.2.1 Topological Flow Classification

We use the classic method proposed by Chong et al. [19] to classify 3D flow fields. Given a 3D flow from a DNS, we compute the rate-of-deformation tensor in a local reference frame at each Cartesian grid point. By analyzing the invariants of the tensor, we can categorize the local flow structure into one of 27 fundamental types. With the classification of each grid point with respect to its topological type, a 3D flow field can be segmented into the regions of different flow patterns. Scientists find that only a few dominated flow patterns are presented in the flow fields of DNS [20]. These are shown in Table 1 with their description by Chong et al. [19].

Table 1. Dominated Flow Topological Classifications

Classification	Topological Description
2	Node / node / node, unstable (NNN/U)
11	Node / saddle / saddle, stable (NSS/S)
12	Node / saddle / saddle, unstable (NSS/U)
18	Focus / stretching, stable (FS/S)
19	Focus / stretching, unstable (FS/U)
20	Focusing / compressing, stable (FC/S)
21	Focusing / compressing, unstable (FC/U)

Among these flow classifications, scientists are particularly interested in vortical structures that have a coherent FC/U region at its core and are surrounded by an arm shaped FS/S region. This structure is presumed to have a strong influence on the deformation of local flamelets, and is critical for developing phenomenological models. Figure 2 shows an overview of such vortical structures in a combustion flow field. It is desirable to select and extract individual vortices at different locations and study the corresponding local properties.

Although the topological classification is well defined at each grid point of field data, selecting and extracting 3D vortical structures from a combustion data presents unique challenges. Unlike a laminar flow field, a combustion flow field is typically

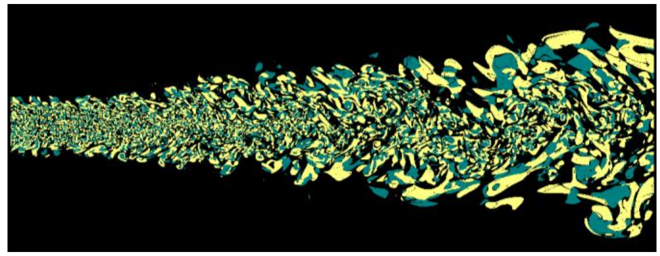

Figure 2. An overview of the classification of FC/U regions (green) and FS/S regions (yellow) from a slice of a 3D combustion flow field.

characterized by high turbulence and heavily interwoven vortical structures. Substantial clutter present in the combustion flow prevents scientists from effectively perceiving detailed patterns and makes isolating individual structures very difficult. These issues are addressed in the region growing portion of our framework.

3.2.2 Feature Selection and Extraction

In our design, we use a simple interface to assist users to explore complex vortical structures and extract features of interest. Users are presented with a 2D slice of the jet showing the topological classification of each point. Each classification is colored differently to let users distinguish features in the jet. Users are allowed to change the depth of the slice in order to view all portions of the jet. These means can significantly lower the visual complexity of showing turbulent vortical structures.

Users can click on the 2D slice to select a region with a certain topological flow type. The point clicked is chosen as a seed point to generate the corresponding full 3D structure using a modification of a region growing algorithm [21]. The region growing process is performed in a breadth-first manner, in that a queue is maintained for searching and classifying the voxels. The queue is initialized using the seed point. For each voxel in the queue, we check topological flow type of its neighbors. If its type matches the selected topological type, the voxel is marked as being part of the region, and its neighbors that have not been visited are entered into the queue. We iterate this process until the queue is empty.

Due to the turbulent nature of the combustion flow field we study, vortices can be interwoven in a vicinity area. The standard region growing algorithm does not sufficiently isolate structures which are weakly connected by only a few voxels. To address this issue, we add a threshold value to detect and remove the weak connections. Before adding a voxel to the region, we count the number of similar neighbors in 3D space to determine how strongly the voxel connects to the rest of the region. A voxel is then only added if its connectivity strength is above a user defined threshold. In this way, users are able to easily explore and extract the desired features from a turbulent flow. Figure 3 shows how different shapes can be extracted from the same seed point by simply adjusting the threshold value. Figure 4 shows a 3D representation of such an extracted feature.

To account for the large data set sizes that are produced from DNS combustion simulations, we parallelize this process using a master-slave paradigm. The user views a slice of the data on the master processor. When a user selects a seed point, the master processor serially grows the region on the 2D slice. All of the voxels in this 2D grown region are treated as seed points to be used in 3D growing, and distributed to worker processors, which

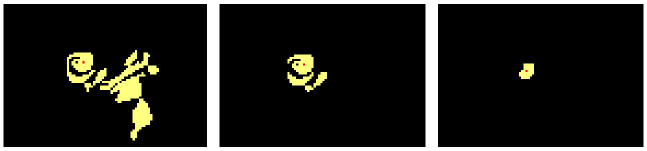

Figure 3. 2D slice of different shapes extracted from a constant seed point (red) by adjusting the threshold value; the images from left to right represent an increasing threshold.

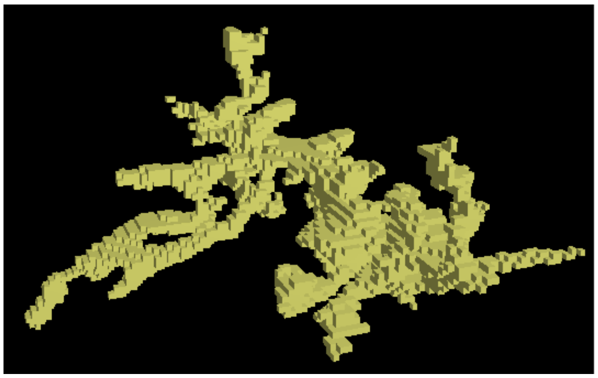

Figure 4. A 3D representation of an extracted FS/S feature.

each see a portion of the entire 3D domain (split evenly among workers). Any worker that receives one or more seed points from the master processer starts growing in its local 3D domain. If a voxel that lies on the boundary between two processors is added to the region, the processor sends a message to its neighbor along that boundary so that region growing can potentially continue across boundaries. This continues until all workers have finished growing such that their respective queues are empty.

3.2.3 Feature Sub-classification and Extraction

While the region growing algorithm with the user defined threshold is an effective way to isolate weakly connected features, it is still limited in the types of shapes that can be extracted. Furthermore, using high thresholds can result in region shrinkage as exterior voxels may not be considered strongly connected enough to be added to the region. To give the user another level of control, we modify the region growing algorithm.

Instead of using only 27 fundamental topological types, we classify each type further into 4 sub-topological types (e.g. 18-1, 18-2, 18-3, and 18-4) according to 4 levels of the same topological description [19]. Having more classifications can facilitate the reduction of the interconnectivity of like topological types in the data set and produce better results when the original region growing algorithm is used. Figure 5 shows the difference between the original topological classification and the sub-topological classification in a zoomed-in portion of the jet.

Furthermore, we modify the region growing algorithm to be able to merge together strongly connected regions of different sub-topological type but same topological type. For example, 18-1 can be merged with 18-3, but not with 21-3. When the region growing algorithm encounters a voxel with a topological type that could be merged, it places it in a separate queue. This separate region is grown and can be potentially merged if it exceeds a certain connectivity strength that is determined by counting voxels along the boundary of the two regions. This strength must exceed a user defined threshold in order to constitute a merge.

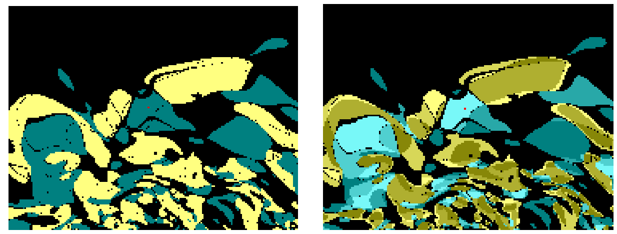

Figure 5. The left image shows a zoomed-in portion of the jet showing the standard topological classification. The right image shows the sub-topological classifications using different shades of the same color.

Figure 6. Extracting and merging different sub-structures from a constant seed point (red) by adjusting the threshold value; the images from left to right represent an increasing threshold.

Figure 6 shows how different sub-topological types can be merged together from the same seed point by simply adjusting the threshold value. This can be parallelized in a manner similar to the normal topological region growing as described in Section 3.2.2. Users can choose to extract features using either of the two methods to suit their needs.

3.3 Lagrangian Particle Query and Analysis

The particle query portion of our framework utilizes components from our previous work, specifically the parts of the COMPARED (Combined Particle Analysis, Reduction, Exploration, and Display) system which manages the large scale Lagrangian particle data [22]. This system is able to utilize multi-core parallelism to efficiently extract subsets of particles based on a range query of their inherent properties.

Our framework first supports conventional query-based particle analytics that are commonly used in various domains and applications. The framework can select particles using explicit range query of particles' inherent properties, such as position or temperature, or a set of derived variables that are computed at runtime, such as mixture fraction. Because of its embarrassingly parallel nature, the particle range query can be easily parallelized by partitioning the data among worker processors; the range query of each particle can be evaluated independently on a single processor.

This framework can then compute statistical information according to specific analysis requirements. Besides simple information, such as minimal/maximal/average values, scientists are also interested in computing more advanced statistics, such as the conditional average of one variable subject to another variable. This information is useful for exploring the relationship between primary variables within the selected subset of particles.

We can also visualize the selected particles using their spatial information, providing immediate feedback for exploring the particle data. Particles are rendered as Point Sprites and colored according to their primary variables. From visualization results, we can qualitatively assess the spatial interaction between variables.

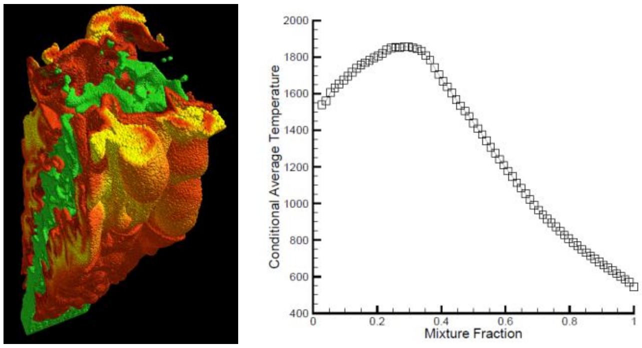

Figure 7. The left image shows particles in a raw fuel stream and a region where the flame reaction zone is located. It is colored by temperature showing a cold (green) fuel stream and hot (red) products of combustion. The jet originates from the bottom left corner and flows towards the top of the figure. The right image shows the conditional mean of temperature on mixture fraction for the queried particles.

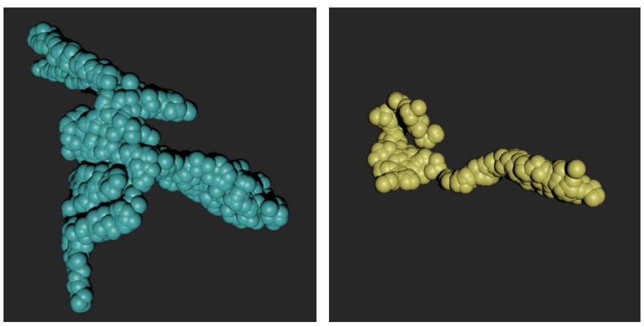

Figure 8. The left image shows an extracted set of particles representing an FC/U feature. The right image shows particles representing an FS/S feature.

Figure 7 shows the visualization and statistics of a subset of particles through certain range query. The ability of query, analysis, and visualization enables scientists to interactively sort through and conduct analysis on particle subsets of interest from the entire large particle data set.

3.4 Parallel Feature-based Particle Query

With the extracted topological features, scientists also desire to study their properties using the Lagrangian description. The key operation to enable this analysis is to identify the particles encapsulated in the spatial region of the features. The statistical information of the particles can then be computed to further identify flame behaviors and convey their correlation in multivariate space.

We extend the particle extraction phase to accept voxel data in the form of a 3D bitmask from the region growing algorithm and extract corresponding particles from the Lagrangian data set. A value of 1 in the 3D bitmask represents a voxel that is part of a grown region. By only using a 3D bitmask to represent the field data, we can minimize the amount of information that must be transferred to the particle extraction phase. The spatial coordinates of each particle are used to map each particle to the voxel coordinate space. If the particle corresponds to 1 in the region bitmask, it is extracted for later analysis. Each feature, as generated by the region growing algorithm, can contain several thousands of voxels allowing scientists to extract and analyze particles within very complex shapes. Figure 8 shows an extracted set of particles representing an FC/U feature and an FS/S feature. In this way, we can incorporate volumetric features into the query portion of our framework.

3.5 Parallel Particle-based Volume Feature Query

Query-based volume analysis is also supported by our framework. Since we can easily map particles to the voxel coordinate space, we generate a list of voxels in the form of a 3D bitmask that contain one or more queried particles. A value of 1 in this 3D bitmask represents a voxel that contains particles. This allows us to generate features that correspond to particle range queries by using these voxels as input into our region growing phase. If the voxel corresponds to 0 in the bitmask, it is ignored during the region growing process. We can then compare this region to the volume data to see which topological classifications dominate our feature. In this way, we can also incorporate particle queries into the feature extraction portion of our framework.

Another very useful application is trajectory-assisted feature tracking. Queried particles can be assembled into meaningful time-series curves as particle trajectories. These trajectories provide a possible way to identify a correspondence between features at different time steps. The information of these correspondences is essential to tracking features over time, which is traditionally difficult without sufficient temporal samples. With our framework, we can identify the features along one particular set of trajectories by performing particle-based volume feature query at each time step.

4. RESULTS

In this section we describe the results of performance testing and demonstrate a few examples of how both feature-based particle query and particle-based volume feature query can be used for analysis. Tests were performed on vector field data and particle data from a large-scale direct numerical simulation of a turbulent lifted ethylene jet flame. The data set, provided by scientists at Sandia National Laboratories, consists of vector field data on a 2025 by 1600 by 400 domain and particle data consisting of over 40 million particles.

4.1 Performance Results

Performance tests were done on Hopper, a 6,384 node Cray XE6 system, at the National Energy Research Scientific Computing Center (NERSC). Each node consists of two AMD 'MagnyCours' 2.1-GHz processors. We test the performance of the region growing and particle extraction phases separately as they each scale differently. Furthermore, these tests do not reflect any I/O times and assume that data has already been distributed to all compute nodes.

Since the region growing time depends on the size of the region, we run performance testing on a feature that is typically the size that scientists are interested in extracting. For this particular data set, this corresponds to a feature whose volumetric size is on the order of 10,000 voxels. This size was determined by consulting combustion scientists at Sandia National Laboratories.

The performance results of growing such a feature can be seen as blue curves in Figure 9. When dividing the total domain among 16 processors or less, this particular region still only resides on one processor and gets grown completely in serial. Between 32 and 128 processors, the region is partitioned among an increasing number of processors resulting in jumps in speedup. As region gets subdivided further, the communication cost increases since the number of voxels on a boundary between processors also increases. We see that when using more than 128 processors, the communication cost slowly becomes a dominant part of the growing process. This shows that there is an optimal number of processors that results in the most speedup. This can vary as the size and shape of the feature changes.

We also run performance tests on the particle extraction phase. Unlike the region growing phase, the time required to extract a subset of particles is independent of the feature size. This is because we must map all the particles in the data set to a voxel location before checking if that particular voxel is part of a segmented feature. However, this step is embarrassingly parallel as no communication is required between processes to do the mapping and bitmask comparison. As seen as red curves in Figure 9, the particle extraction stage of the framework displays a linear speedup with the number of processes used. We achieve 100% parallel efficiency from 8 to 512 processors.

We also compare each portion of our framework relative to one another. Using the number of optimal processors (approximately 128 in this test) based on the region growing performance data, we find that neither the region growing phase nor the particle extraction phase dominates the overall cost. Furthermore, the time of each phase is on the order of 10^{-2} seconds allowing scientists to explore the data at an interactive speed. With this nearly optimal processor number, we can achieve 74% parallel efficiency from 8 to 128 processors from both phases combined.

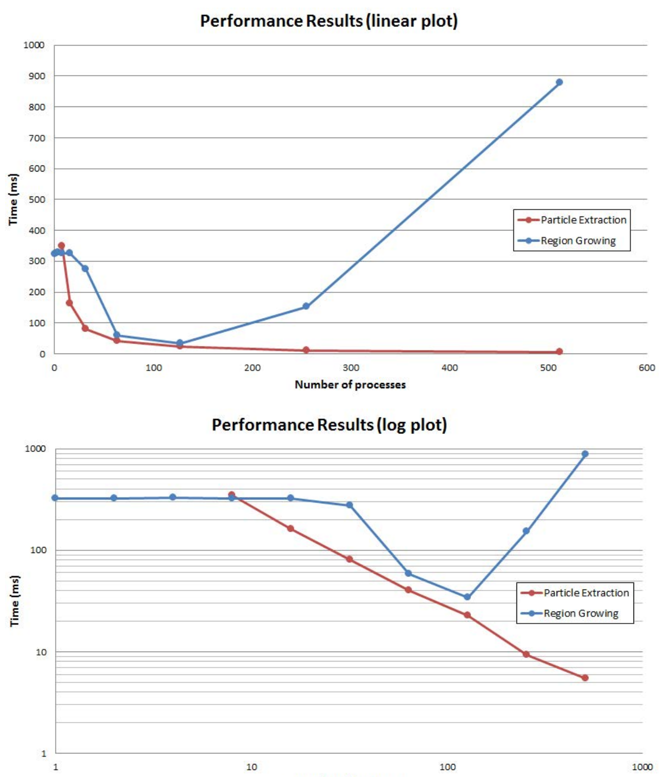

Figure 9. Performance results for both the region growing step and the particle extraction step on a linear plot (top) and a log-log plot (bottom)

4.2 Particle Analysis of Extracted Volume Features

This section describes a few feature-based particle analysis results, where we first extract particles within a segmented feature, and then look at the parameters of these particles to better understand our feature. The dataset features a non-premixed combustion jet, that is, the fuel and oxidizer are inserted into the jet separately before combustion occurs. The components of the jet correspond very closely to the mixture fraction between these two materials. We therefore base our example analysis largely around the mixture fraction variable.

We can differentiate between mixing and burning solutions by looking at a plot of temperature versus mixture fraction for a representation of the full particle data set, as shown in Figure 10. The mixing solution is represented by the linear correlation between temperature and mixture fraction, whereas the burning solution is represented by the non-linear correlation. In the burning solution, the temperature correlates positively with a low mixture fraction (too much air and not enough fuel). In the other extreme case (not enough air and too much fuel), the temperature correlates negatively with mixture fraction. The peak temperature on the curve represents the point where fuel and air are mixed in just the right proportions (the stoichiometric mixture fraction).

Figure 11 shows the data from the extracted sets of particles representing the two different topological features and Figure 12 shows their location in the jet. It is evident that feature A (shown in red) represents a part of the jet that is burning. On the other hand, feature B (shown in green) represents a part of the jet that is mixing. The shapes of the extracted particles are located in Figure 8 with feature A on the left and feature B on the right.

The mixing and burning branches are also evident in a plot of hydroxide (OH) mass fraction versus mixture fraction. When the mass fraction of hydroxide (a result of the burning) is zero, we can see a mixing between the fuel and the oxidizer because no chemical reactions are taking place. The burning solution in this case is described by the curve where the hydroxide mass fraction is non-zero. Figure 13 shows the hydroxide mass fraction of the extracted particles from the same two topological features. Once again, it is evident that feature A (shown in red) represents a part of the jet that is burning while feature B (shown in green) represents a part of the jet that is mixing.

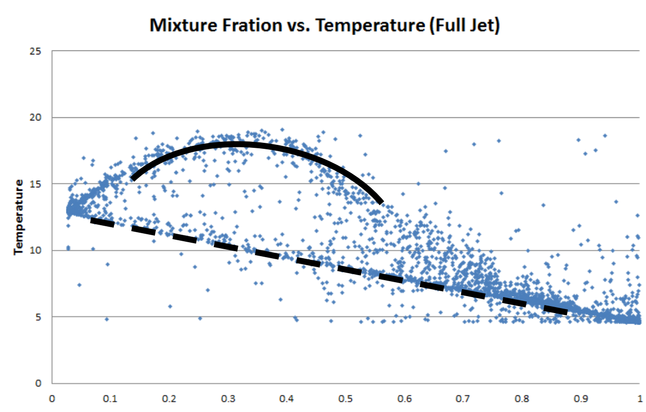

Figure 10. A scatter plot representative of the full particle data set highlighting the mixing solution (dashed line) and the burning solution (solid curve).

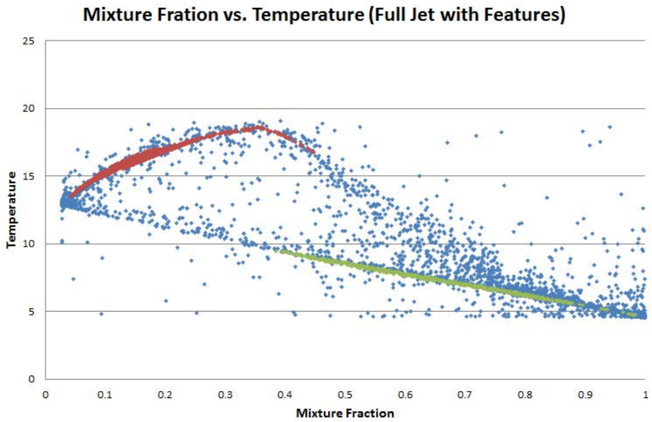

Figure 11. The extracted particle data for feature A (red) and feature B (green) overlaid on the particle data representing the full jet.

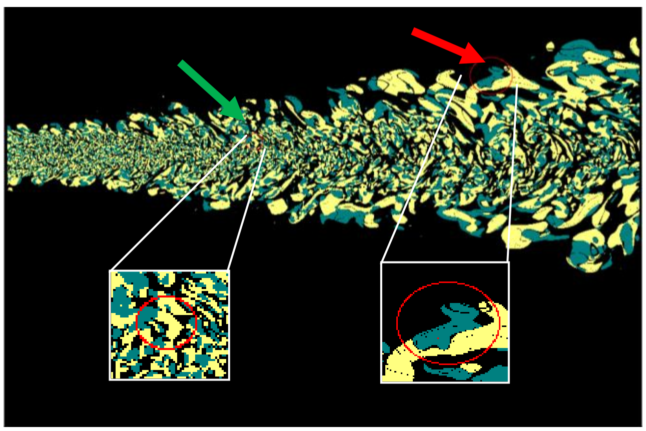

Figure 12. Two volumetric features from different parts of the jet; feature A (red arrow) and feature B (green arrow).

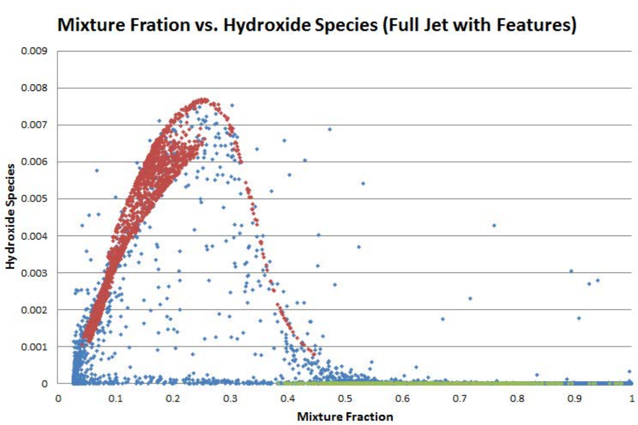

Figure 13. The extracted particle data for feature A (red) and feature B (green) overlaid on the particle data representing the full jet.

4.3 Volume Analysis of Extracted Particles

This section describes a particle-based volume feature query analysis result, in which we look at the flow classifications that correspond to particle range queries. In this example, we extract a set of particles according to one of their fundamental parameters, temperature. We use the range query, *Temperature > 15* to extract only high temperature particles (that we know are in a burning portion of the jet).

The spatial locations of these particles are mapped to the voxel coordinate space and used to generate a feature representing the hottest parts of the jet. Looking at the flow classifications within this feature shows that 35.9% consists of FS/S regions and 23.2% consist of FC/U regions. The remaining portion consists of unclassified areas and a combination of other much less dominant flow classifications. We can repeat this procedure using a different range query to extract the coldest portions of the jet. Looking at the flow classifications shows similar results: 32.6% consists of FS/S regions and 21.6% consists of FC/U regions. A similar breakdown occurs for the "warm" portions of the jet as well.

For our particular example we see that the FS/S regions dominate both the hottest and coldest portions of the jet. This indicates that it is unlikely that there is a correlation between dominant flow classification and temperature, although a more careful analysis may be necessary. Correlating the various flow patterns to the portions of the jet described by other primary variables or derived variables, such as mixture fraction, can be very useful to combustion scientists. Such an analysis is easily achieved through our framework.

5. CONCLUSIONS AND FUTURE WORK

In this paper we present a framework that can conduct parallel data analysis on a combination of Lagrangian particle data and Eulerian field data from large-scale combustion simulations. Our framework combines these dataset types through two variations: feature-based particle query, in which we extract particles within a segmented feature, and particle-based volume feature query, in which we look at the flow classifications that correspond to particle range queries. These methods can allow scientists to combine the information from the Eulerian and Lagrangian viewpoints and obtain a deeper understanding of fundamental combustion processes.

In addition, we provide scientists with two methods to extract features from a heavily interwoven turbulent flow field. The first is a modification of a standard region growing algorithm to include a user defined threshold value that can be used to separate weakly connected regions from the rest of the jet. The second uses a sub-classification technique to grow sub-regions which can be connected using a separate user defined threshold. We allow scientists to use either method in our framework.

Performance results on a large-scale direct numerical simulation of a turbulent lifted ethylene jet flame show that parallelizing the region growing and particle portions of our framework can lead to large speedups. While particle extraction portion is extremely scalable, the region growing is limited by communication overheads.

Future work will involve first improving the region growing step of our framework so that it scales similarly to the particle extraction step. This could be achieved by developing a scalable method of local growing and parallel graph connection to segment features. However, applying this method to combustion datasets, which are very turbulent and highly interconnected in nature, may prove challenging.

In addition, we plan to extend the particle-based volume feature query portion of our framework to track segmented features over time. By looking at the trajectories of particles we can identify a correspondence between features at different time steps. This will allow scientists to not only track the position of features, but also to identify how features merge and split apart over time.

Lastly, we plan to implement our framework into the combustion simulations directly to allow scientists to perform in situ analysis and visualization on their datasets. This is especially useful for minimizing storage and I/O overheads when dealing with extremely large datasets.

6. ACKNOWLEDGMENTS

This research has been sponsored in part by the National Science Foundation through grants OCI-0905008, OCI-0850566, OCI-0749227, and CCF-0811422, and also the Department of Energy through grants DEFC02-06ER25777, DE-CS0005334, and DE-FC02-12ER26072.

7. REFERENCES

[1] J. H. Chen, A. Choudhary, B. de Supinski, M. DeVries, E. R. Hawkes, S. Klasky, W. K. Liao, K. L. Ma, J. Mellor-Crummey, N. Podhorszki, R. Sankaran, S. Shende, and C. S. Yoo. Terascale Direct Numerical Simulations of Turbulent Combustion using S3D. *Comput. Sci. Disc.*, vol. 2, pp. 015001, 2009.

[2] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, Cambridge, 1967.

[3] S. H. Lamb. *Hydrodynamics (6th ed.)*. Cambridge University Press, Cambridge, 1994.

[4] P. K. Yeung. Lagrangian Investigations of Turbulence. *Annual Review of Fluid Mechanics*, Vol. 34, pp. 115-142, 2002.

[5] H. Akiba and K.-L. Ma. A Tri-Space Visualization Interface for Analyzing Time-Varying Multivariate Volume Data. *In Proceedings of Eurographics/IEEE VGTC Symposium on Visualization*, pp. 115-122, 2007.

[6] P.-T, Bremer, G. H. Weber, J. Tierny, V. Pascucci, M. S. Day, and J. B. Bell. Interactive Exploration and Analysis of Large-Scale Simulations Using Topology-Based Data Segmentation. *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, pp. 1307-1324, 2011.

[7] J. Bennett, V. Krishnamoorthy, S. Liu, R. Grout, E. Hawkes, J. Chen, J. Shepherd, V. Pascucci, and P.-T. Bremer. Feature-Based Statistical Analysis of Combustion Simulation Data. *IEEE Transactions on Visualization and Graphics*, vol. 17, pp. 1822-1831, 2011.

[8] J. Wei, H. Yu, R. W. Grout, J. H. Chen, and K.-L. Ma. Visual Analysis of Particle Behaviors to Understand Combustion Simulations. *IEEE Computer Graphics and Applications*, vol. 32(1), pp. 22-33, 2012.

[9] K. Wu. FastBit: An Efficient Indexing Technology for Accelerating Data-Intensive Science. *J. Phys.: Conf. Ser.*, pp. 556-560, 2005.

[10] L. J. Gosink, K. Wu, E. W. Bethel, J. D. Owens, and K. I. Joy. Bin-hash Indexing: A Parallel Method for Fast Query Processing. *Technical Report 729E-LBNL*, 2008.

[11] K.-L. Ma, G. Schussman, B. Wilson, K. Ko, J. Qiang, and R. Ryne. Advanced Visualization Technology for Terascale Particle Data. *In Proceedings of IEEE/ACM Supercomputing Conference*, 2002.

[12] C. Jones, K.-L. Ma, S. Ethier, and W.-L. Lee. An Integrated Exploration Approach to Visualizing Multivariate Particle Data. *Computing in Science & Engineering*, vol. 10(4), pp. 20-29, 2008.

[13] O. Rübel, Prabhat, K. Wu, H. Childs, J. Meredith, C.G.R. Geddes, E. Cormier-Michel, S. Ahern, G.H. Weber, P. Messmer, H. Hagen, B. Hamann, and E.W. Bethel. High Performance Multivariate Visual Data Exploration for Extremely Large Data. *In Proceedings of IEEE/ACM Supercomputing Conference*, 2008.

[14] J. Woodring, J. Ahrens, J. Figg, J. Wendelberger, S. Habib, and K. Heitmann. In-situ Sampling of a Large-Scale Particle Simulation for Interactive Visualization and Analysis. *In Proceedings of EuroVis*, 2011.

[15] J. Wei, H. Yu, K.-L. Ma, and J. H. Chen. Parallel Clustering for Visualizing Large Scientific Line Data. *In Proceedings of IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, 2011.

[16] M. Glatter, J. Huang, S. Ahern, J. Daniel, and A. Lu. Visualizing Temporal Patterns in Large Multivariate Data using Textual Pattern Matching. *IEEE Transactions on Visualization and Computer Graphics*, vol. 14(6), pp. 1467-1474, 2008.

[17] W. Kendall, J. Huang and T. Peterka. Geometric Quantification of Features in Large Flow Fields. *IEEE Computer Graphics and Applications (Special Issue on Extreme Scale Analytics)*, vol. 32(4), pp. 50-59, 2012.

[18] K. P. Gaither, H. Childs, K. W. Schulz, C.s Harrison, W. L. Barth, D. Donzis, and P.-K. Yeung. Visual Analytics for Finding Critical Structures in Massive Time-Varying Turbulent-Flow Simulations. *IEEE Computer Graphics and Applications*, vol. 32(4), pp. 34-45, 2012.

[19] M. S. Chong, A. E. Perry, and B. J. Cantwell. A General Classification of Three Dimensional Flow Fields. *Physics of Fluid*s, vol. 2, pp. 765-777, 1990.

[20] R.W. Grout, A. Gruber, C.S. Yoo, and J.H. Chen. Direct Numerical Simulation of Flame Stabilization Downstream of a Transverse Fuel Jet in Cross-flow. *In Proceedings of the Combustion Institute*, vol. 33, pp. 1629-1637, 2011.

[21] R. Huang and K.-L. Ma. RGVis: Region Growing Based Techniques for Volume Visualization. *In Proceedings. 11th Pacific Conference on Computer Graphics and Applications*, pp. 355-363, 2003.

[22] R. W. Grout, H. Yu, and J. H. Chen. A Parallel Framework for GPU-Enhanced Query and Analysis of Lagrangian Particle Data from Combustion Simulations. *IEEE/ACM Supercomputing Poster*, 2009.