Interactive Topological Exploration of Particle Ensembles

Harsh Bhatia*, Pavol Klacansky†, Shusen Liu†, Wathsala Widanagamaachchi†,
Attila Gyulassy†, Valerio Pascucci†, Peer-Timo Bremer*
* Lawrence Livermore National Laboratory; † Scientific Computing & Imaging Institute, The University of Utah

ABSTRACT

We present an integrated system of tools for exploring ensembles of time-varying particular simulations of salt diffusion. In particular, we allow scientists to interactively explore the nature and temporal evolution of *viscous fingers* – the primary feature of interest. Using a topological analysis of the salt concentration, our framework provides multiple inter-linked visualization and analysis tools, enabling interactive exploration of feature evolution in both time and parameter-space. Our pipeline connects spatial visualization of fingers with the visualization and analysis of multidimensional statistics, and allows the scientist to iteractively perform a detailed exploration of the entire ensemble.

1 INTRODUCTION

In this submission, we present our exploration and analysis of ensembles of particle-based simulation of salt diffusion and transport. In particular, we provide tools to interactively identify, visualize, and characterize *viscous fingers* of high salt concentration. Viscous fingers are an important type of instability created by the difference in viscosity and/or density of two mixing fluids. The given data captures the concentration of salt in water, which in this case is a good stand-in for fluid density, and hence, is a good choice to represent fingers. As a result, the task of identifying viscous fingers requires exploration of features in the data that correspond to high salt concentration regions.

The primary goal of this work is to create a pipeline of tools and techniques to enable interactive exploration of the simulation — both individual runs as well as the ensemble as a whole, with respect to a variety of parameters in order to provide the domain scientist an in-depth understanding of the simulations. For the purpose of this demonstration, we choose certain parameters, e.g., concentration value, without loss of generality; while the correct choice of parameters remains a research question for the scientists, our goal is to create a pipeline that does not depend upon our choices, but, in fact, allows exploring the simulation with respect to varying parameters.

The main challenges in the exploration arise due to the large scale and stochastic nature of the simulation. Physically meaningful insights must be derived by simultaneously exploring the ensembles with respect to the given parameters. Therefore, we present tools for statistical, visual, and topological exploration of the given data. By computing topological properties in preprocessing step, our tools support interactive visual and statistical queries on viscous fingers and various associated properties with respect to multiple simulation parameters and time.

2 PIPELINE AND TOOLS

According to the problem description, viscous fingers are defined as regions of high salt concentration. As a result, it is natural to represent fingers as the super-level sets of the salt concentration. However, rather than choosing a single threshold parameter we compute the merge tree [1] of concentration to simultaneously encode fingers for a wide range of thresholds. Merge tree represents the super-level set topology of the underlying function, and characterizes resulting features with respect to varying function value. Our topological and statistical analysis of particle ensembles requires preprocessing the data and conversion into a topological format, which allows interactive exploration through our tools.

Preprocessing. The computation of merge tree requires the definition of a neighborhood among particles or in other words connectivity information. To this end, we compute the Delaunay triangulation of the particle data using CGAL [3] though other strategies, e.g., connecting to all neighbors within the smoothing length, could easily be substituted. One challenge with the generic definition of fingers as regions of high salt concentration is the presence of an unlimited supply of salt at the top of the cylindrical domain. Technically, all fingers still connected to the top of the domain are part of the same region of high concentration which severely distorts the analysis. We compensate for this issue by filtering for $z > 9.8$ units (in a cylinder of height range $z \in [0, 10]$ units), effectively cutting off the infinite salt supply for the purpose of analysis. Finally, we compute the merge trees and the corresponding segmentations for individual timesteps, as well as temporal tracking of features computed by merge trees, and store them in optimized file formats allowing us to interactively explore the topology and statistics of the data for various parameter and time values.

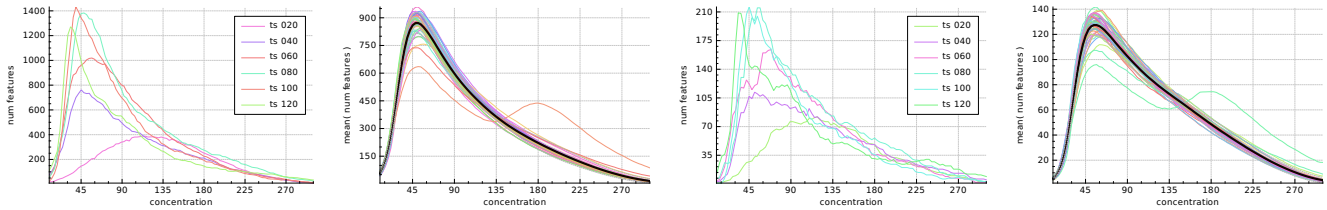
Interactive Exploration. Interactive exploration is performed using three inter-linked tools, all of which use the precomputed merge tree with segmentation and temporal-tracking information.

TALASS [5] is an interactive tool, which enables viewing segmentation and tracking for different thresholds. Illustrated in Fig. 3, TALASS also allows understanding temporal evolution of features by tracking merge tree features across time. This enables our analysis to understand the break-down (and recombination) of fingers as the simulation evolves. High-quality interactive rendering of particle data is achieved by utilizing the ***OSPRay*** [4] ray-tracing engine on CPU. OSPRay renderer was integrated with TALASS visualization, which provides a single visualization system for exploring the fingers with respect to space, time, and parameter.

An interactive ***statistical visualization*** tool was developed to explore the feature characteristics with respect to different parameters, e.g., time, concentration, etc., computed using TALASS, and assists choosing the suitable parameters for the analysis. Finally, we integrate TALASS with a high-dimensional data visualization tool, ***DataExplorerHD*** [2], which allows exploring the multivariate parameter space in the simulation and can reveal important patterns (cf. Fig. 4(a)). The multivariate statistical analysis is a valuable tool for examining ensemble behaviors (cf. Fig. 4(b)).

3 EXPLORATION OF VISCOUS FINGERS

Parameter exploration. As discussed above, viscous fingers are defined as regions of high salt concentration. However, it is not clear what salt concentration should be considered high. Intuitively, the choice of concentration, λ, is a trade-off between "sensitivity" and "specificity". If the threshold is very low, fingers are getting detected very early in their evolution (at very low concentration). However, as the simulation progresses, fingers defined using low thresholds will quickly merge into a single global feature. On the other hand, for high thresholds, one may see only few fingers that are small in size and typically disconnected from each other. Furthermore, as in any other simulated data, one can expect some noise in the system; certain small and isolated regions of high concentration may exist, which are not physically-meaningful fingers. In order to choose a meaningful value of λ, we must first explore the parameter space to understand the evolution of features with respect to concentration.



(a) (b) (c) (d)

Figure 1: Variation of the number of fingers with respect to concentration. (a) For a single run, as the time increases, higher number of fingers are detected for lower values of concentration. (b) The mean number of features peak around 35–45 concentration and remains relatively consistent across runs (different colors); black curve shows the mean of all runs. (c) and (d) The overall behavior remains the same when small fingers are considered noise and ignored.

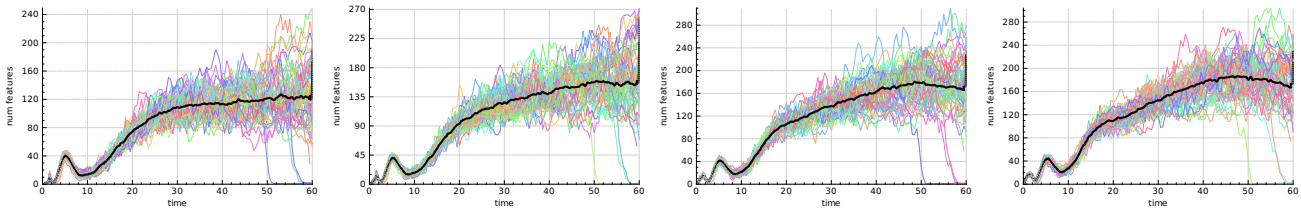


Figure 2: Variation of the number of fingers with respect to time. Different runs are shown in different colors, with black curve showing the mean of all runs. Left to right: concentration values 35, 40, 45, 50.

Using the merge tree of concentration, the plots in Figure 1 describe a unimodal behavior of the feature count with respect to increasing λ. In particular, Figure 1(a) shows that for earlier time-steps, the number of features remains largely stable with a peak for higher λ, indicating that not many fingers are formed. As the time evolves, larger number of fingers are observed for smaller $\lambda (\approx 40)$, reflecting the fact that the fluid is much more turbulent. Figure 1(c) shows similar trends even when smaller features are considered noise and discarded. Nevertheless, there is still noticeable amount of noise in the plots. Figures 1(b) and 1(d) show the number of features averaged over time for different runs of the simulation. It is noted that the unimodal trend is consistent across runs except one outlier — possibly a run with extreme initial conditions. The time-averaged number of features generally peaks between $35 < \lambda < 50$, making this a reasonable choice for defining fingers.

Temporal exploration. In order to further refine the parameter choice, we study how the number of fingers vary with time within a single run. Figure 2 shows this statistic for different λ, where different runs are shown in different colors. Although a lot of variance is observed across runs (primarily due to different initial conditions and exaggerated by turbulent behavior of the simulation), we note that the mean curve (across runs) shows stable behavior. In particular, the mean curve for $\lambda = 35$ shows most-stable ("flat") behavior across time. As a result, we choose 35 to be the defining concentration for fingers.

We use TALASS (cf. Figure 3) to track features for all concentration values in time, using a nearest neighbor correlation within the smoothing length radius to connect features. The initial tracking graph is convoluted and obscure, but TALASS also allows filtering features by size and to simplify the merge tree by persistence. These simplifications reduce the size and complexity of the tracking graph and allows meaningful visual exploration.

High-dimensional exploration. One of the main advantages in using merge trees to encode fingers is that it allows one to precompute an arbitrary set of descriptive statistics on a per-feature basis. In this case, e.g., we compute the size, maximal and mean salt concentration, mean z-velocity, etc. for each feature. The integration of TALASS and DataExplorerHD allows the synchronization of these features in high-dimensional parameter space with those in simulation domain. For example, Figure 4(a) highlights largest features in both the statistical plot and the particle rendering. The DataExplorerHD has the capability to apply various dimension-reduction and clustering algorithms to high-dimensional data, which further enhances the ability to explore the importance of various parameters for feature evolution.

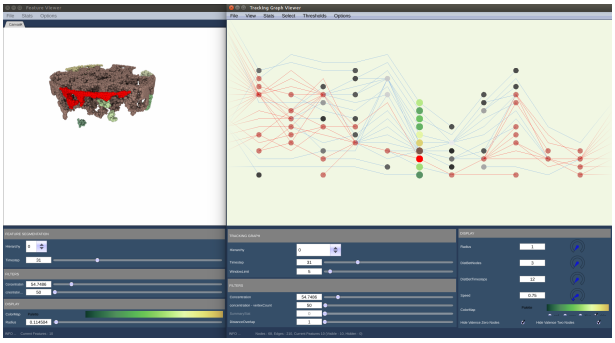
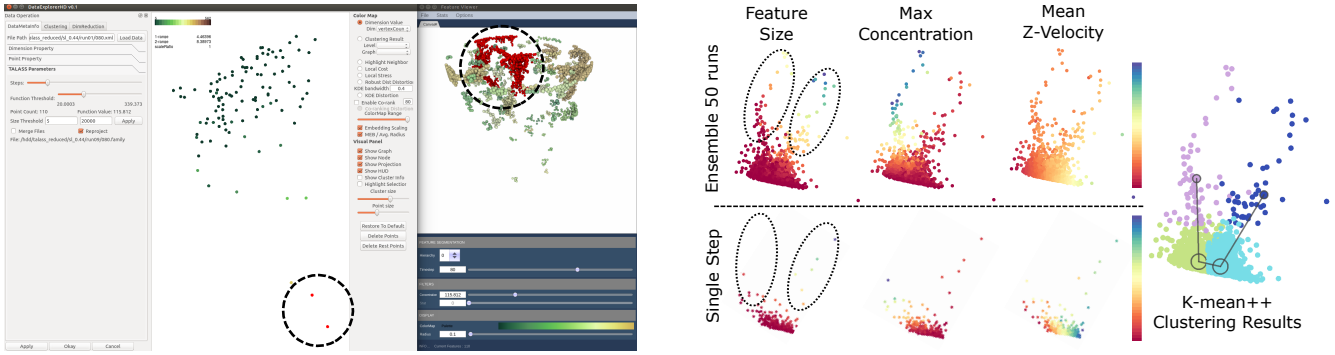


Figure 3: TALASS framework with linked views. On the left, segmentation view showing spatial domain, with particles rendered as spheres (using OSPRay). On the right, tracking graph view which shows feature tracking over time. The selected red feature is highlighted and its track is shown.

The spatial rendering is not suitable for understanding ensemble behaviors since different simulation runs have very different particle movement patterns. DataExplorerHD, on the other hand, provides a viable approach for analyzing the ensembles by visualizing features corresponding to the entire ensemble in a single statistic plot, and a direct cue of (dis)similarities between different runs. As shown in Fig. 4(b), the PCA (Principal Component Analysis) projection of ensemble feature statistic reveals an interesting branching structure. In this analysis, we aim to understand the relationship between feature size, maximum concentration, as well as mean z-velocity of features, of one time-step across multiple simulation runs. We can further the analysis by applying a clustering in the original multivariate space: here we can see the two branches corresponding to the purple and blue clusters. In this plot, each point corresponds to a feature. Based on the value displayed for each field, we can see the largest feature size and maximum concentration seem to correspond to the two "branches". This indicates that the largest feature usually does not have the highest



(a) (b)

Figure 4: (a) Integrated application of TALASS and DataExplorerHD allows highlighting features in the spatial domain along with their projection in the parameter space, respectively. (b) Ensemble of the statistics compared to single step. Points are colored according to the column label. The ensemble reveals the interesting branching structure in the data, which is barely visible in single step projection.

concentration. For the mid-sized or large features (the points in the two branches, the purple and blue clusters), the downward velocity roughly corresponds to features with high concentration, and larger feature size corresponds to upward velocity. In our exploration, such patterns appear only toward the end of simulation as the features take up more space in the simulation domain and some features are reflected upward by the bottom of the cylinder. One interpretation of such pattern is that relatively small features with high concentration will descend and grow in size as they become more diluted. However, as they become larger the features may be influenced more strongly by turbulent effects as well as the bottom of the domain.

4 Results

Our exploration pipeline and tools explained in Section 2 allow interactive visualization and exploration of the point data, viscous fingers, and associated statistics. This section discusses the application of these tools to explore the given data and respond to the various tasks of the contest.

Task 1. OSPRay rendering integrated with TALASS enables high-quality rendering of point data as spheres. Whereas the interface allows changing the radii of the spheres to avoid cluttering, we believe the appropriate choice of the radius is proportional to the smoothing length of the simulation as it defines how many particles are present in a unit volume. Finally, in order to reduce rendering of redundant structures and avoid clutter, we choose to not render the domain directly, as the evolution of fingers indirectly shows the walls of the cylindrical domain.

Task 2. As discussed in Section 3, we define viscous fingers using a carefully chosen value of the salt concentration in the domain. The pre-computation of topological information allows interacting with the data in real time and understanding the evolution of fingers with respect to concentration (and time). For the purpose of this exploration, the spheres are colored by finger identifier, i.e., all points in a finger are shown in the same color, which gives a quick overview of the shapes and sizes of different fingers.

Task 3. The temporal evolution of fingers is captured as a time-tracking graph in TALASS. The linked-view framework allows interacting with the data in time through selection of fingers and their tracks over time. Some of the interactive queries that the system supports include visualizing the breakdown and recombination of fingers across time and the change of volume, mean vertical velocity, center of mass, etc., of fingers across time. Statistical queries about finger properties can also be made interactively: evolution and distribution of finger count and various finger characteristics mentioned above etc., can be visualized with respect to both parameter and time.

Task 4. We provide two different tools to interactively explore the ensemble of runs. First, the interactive statistics allow users to create a wide variety of statistical summaries of the entire ensemble as well as of individual time steps (but multiple runs) or individual simulations. Second, the DataExplorerHD provides scientists with a full range of high-dimensional analysis tools including dimension reduction, clustering, etc., using the multi-variate feature attributes of interest. As the DataExplorerHD is directly linked to TALASS, both tools in combination provide a seamless integration between exploring summary statistics of any part of the ensemble and a detailed spatio-temporal analysis of individual simulations.

Task 5. Our system is entirely driven by the ability to interactively summarize arbitrary subsets of the ensemble using a wide range of statistics. We envision users to start by formulating broad questions and hypothesis, like those for picking the thresholds, which are subsequently refined according to the results. At any time the user can select features or simulations of interest to explore in more detail, e.g., to examine outliers or otherwise unexpected behaviors. We are dealing directly with the full ensemble and even the most expensive ensemble-wide statistics (the parameter plots like those of Figure 1) take on the order of seconds to complete. Furthermore, to be as flexible as possible we currently consider all possible thresholds above 0.9. Subsetting this range to around the threshold of interest, i.e., $[20-60]$, would reduce file sizes drastically and speed up the exploration even further. As shown in the accompanying video, the rendering and spatio-temporal exploration are fully interactive, which includes changes in concentration threshold. The pre-processing times range between seconds per time step for the smaller simulations to multiple minutes but can be performed data parallel and is required only once.

References

[1] P.-T. Bremer, G. Weber, V. Pascucci, M. Day, and J. Bell. Analyzing and tracking burning structures in lean premixed hydrogen flames. *IEEE Trans. Vis. Comp. Graph.*, 16(2):248—260, Mar 2010.

[2] S. Liu, B. Wang, J. J. Thiagarajan, P.-T. Bremer, and V. Pascucci. Visual exploration of high-dimensional data through subspace analysis and dynamic projections. In *Computer Graphics Forum*, volume 34, pages 271–280. Wiley Online Library, 2015.

[3] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.8.1 edition, 2016.

[4] I. Wald, A. Knoll, G. P. Johnson, W. Usher, V. Pascucci, and M. E. Papka. Cpu ray tracing large particle data with balanced pkd trees. In *2015 IEEE Scientific Visualization Conference (SciVis)*, pages 57–64. IEEE, 2015.

[5] W. Widanagamaachchi, C. Christensen, V. Pascucci, and P.-T. Bremer. Interactive exploration of large-scale time-varying data using dynamic tracking graphs. In *Large Data Analysis and Visualization (LDAV), 2012 IEEE Symposium on*, pages 9 –17, oct. 2012.