IMPACT DRIVEN RESEARCH: EVALUATING AND IMPROVING THE DE FACTO STANDARD FOR PARALLEL PARTICLE ADVECTION

Allocation: Innovation and Exploration/100 Knh
PI: Robert Sisneros[1]
Co-PI: David Pugmire[2]

[1]University of Illinois at Urbana-Champaign
[2]Oak Ridge National Laboratory

EXECUTIVE SUMMARY

Foundational visualization algorithms are central to the production of visualization tools running at computing centers around the world but consume tremendous amounts of limited resources. We believe that understanding the performance characteristics of these algorithms is critical in being good stewards of those unique resources. In this project, we investigate the foundational algorithm for parallel particle advection by focusing on a parameter sweep of the de facto standard algorithm "Parallelize Over Data" (POD). Our work has led to several discoveries: a glaring and resolvable issue in poor default values, identification of the best choices of parameters for future tests, and the existence of viable testing alternatives that save time and resources. Our work sets forth a framework for applying solutions to a practical and often overlooked area in computer visualization.

INTRODUCTION

The understanding of vector fields resulting from large scientific simulations is an important and often difficult task. The predominant analytical techniques are derived by calculating the paths that a set of weightless particles travel along when released in a flow. There are a rich set of direct and derived observable quantities that have resulted in several such methods: streamlines, pathlines, stream surfaces, and finite time Lyapunov exponents. However, there are significant computational challenges of these particle-based visualization approaches when they are applied to very large vector field data. Furthermore, the degree to which this negatively impacts performance is data-dependent. Finally, this dependency does not apply solely to the flow field, as many of the above mentioned techniques are highly sensitive to initial seed placement. Therefore, analysis typically includes a trial-and-error exploration where a single flow field for each instance of the calculation may have drastically different computational profiles.

As a result of these challenges, particle advection-based algorithms are susceptible to overload/starvation. Efficient computation requires careful balancing of computational demands placed on I/O, memory, communication, and processors. Algorithms for load balanced particle advection in small parallel settings and have been implemented for several of these components into a test branch of the source code of VisIt, an open source analysis tool. VisIt's particle advection framework provides access to current data decomposition and domain distribution specifics, as well as the capability for the on-demand loading of a domain to any processor.

METHODS & RESULTS

We performed an extensive parameter sweep of the POD algorithm using the implementation found in VisIt along with an updated version that performs communication asynchronously. We developed a framework that manages generating and executing arbitrary test configurations of parameter settings and performed tests on multiple datasets (see Figure). To aid in our analysis, we developed an imbalance metric to help unify comparability and created new visual metaphors for viewing the multiple axes of the study.

We have made significant findings upon analysis of our parameter sweeps to date. First, we find that optimal configurations are not the widely used default, and, in fact, we find that the default configurations to be quite poor. Second, we find that it is not only possible to predict proper configurations, but that the same configuration is likely applicable in both the asynchronous and synchronous versions of the algorithm. On any given architecture we expect that the default configuration may not be optimal, but through testing, this is correctable and an updated system-specific default will save computing resources. Furthermore, resources may be saved by testing, given that asynchronous tests are the only ones needed. Additional analyses made possible through the imbalance metric has allowed us to further pinpoint additional parameters to ignore in future testing. Moreover, we have also identified areas for directed future analysis.

NEXT GENERATION WORK

We want to bridge the gap between the state-of-the-art and the state-of-the-practice for all foundational visualization algorithms. Furthermore, we hope to promote the efficient use of world-class computing facilities through evangelizing such a practical, yet impactful research paradigm.

PUBLICATIONS AND DATA SETS

Sisneros, R., and D. Pugmire, Tuned to terrible: A study of parallel particle advection state of the practice. *High Performance Data Analysis and Visualization (HPDAV), IPDPS 2016*, Chicago, IL. May 2016.

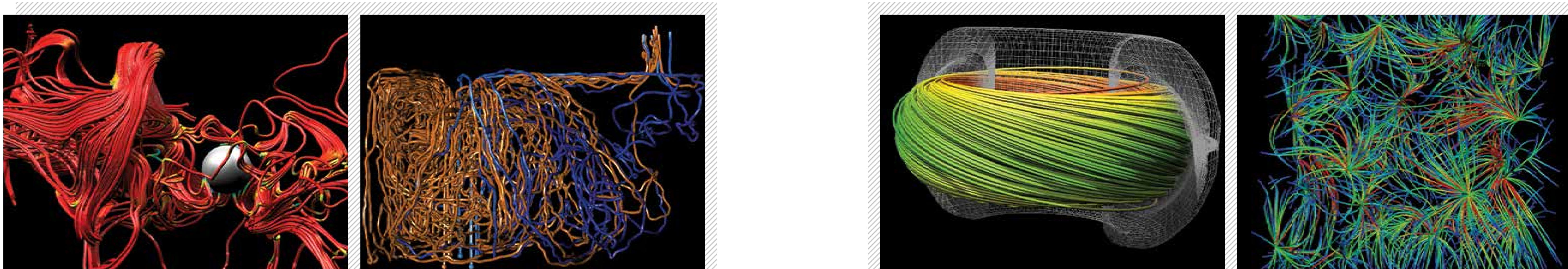

FIGURE 1: Four datasets were used for the parameter sweep in this work. Typical flow visualizations from each are, left to right: astrophysics dataset, thermal hydraulics dataset, fusion dataset, and synthetic test dataset.