Developing LAVAFLOW: A Large-Scale Data Analytics Package

Jonathon Nosowitz, Tomasz Plewa
Department of Scientific Computing, Florida State University
jn18z@my.fsu.edu

Introduction

High resolution computer simulations can generate datasets with a size on the order of tens of terabytes. Should more physics components be added to the problem, such as magnetic fields or complex compositions, the data size requirements may substantially increase. Often, analyzing the data requires the use of a computer system of a size similar to that used to produce the actual simulation results. Thus, analysis of simulation results is frequently cumbersome, may require substantial computational resources, and motivates the development of novel data analysis methods in order to identify and understand the role of participating physics in the application.

In order to address this problem, we developed LAVAFLOW, Library for Advanced Variable Analysis. LAVAFLOW is a package comprising several modules for efficient analysis of computer simulation results of complex, multiphysics problems. The modules provide analysis capabilities for both grid-based and particle-based data in multidimensions. The computational mesh can be non-uniform, including block-based adaptive grids, and the data can be provided in a variety of formats including VTK and HDF5.

Package Structure

The code's modular structure provides a high versatility to the user. Modules can be added or removed in a plugin-like fashion and tailored to a specific application. This is aided by the dynamic memory management to enable offline data analysis and processing on desktop computers. The package optimizes the use of data to only the necessary subsets of solution variables to limit the required memory footprint. Figure 1 displays the directory structure of the package. Listed are the directories of the FractalDimension module available in LAVAFLOW. Each directory level provides information about the contents of that directory and makes navigating the package contents easier.

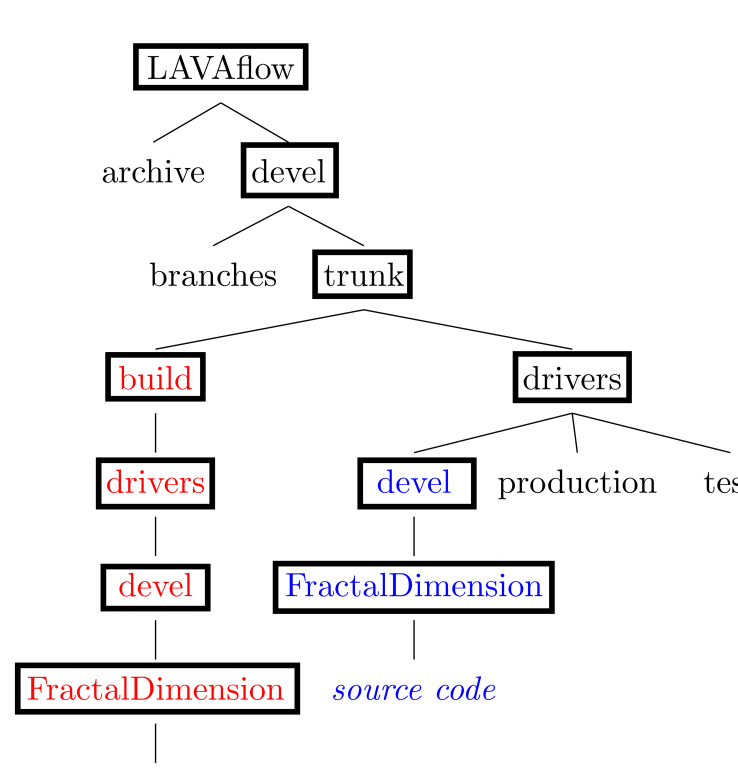


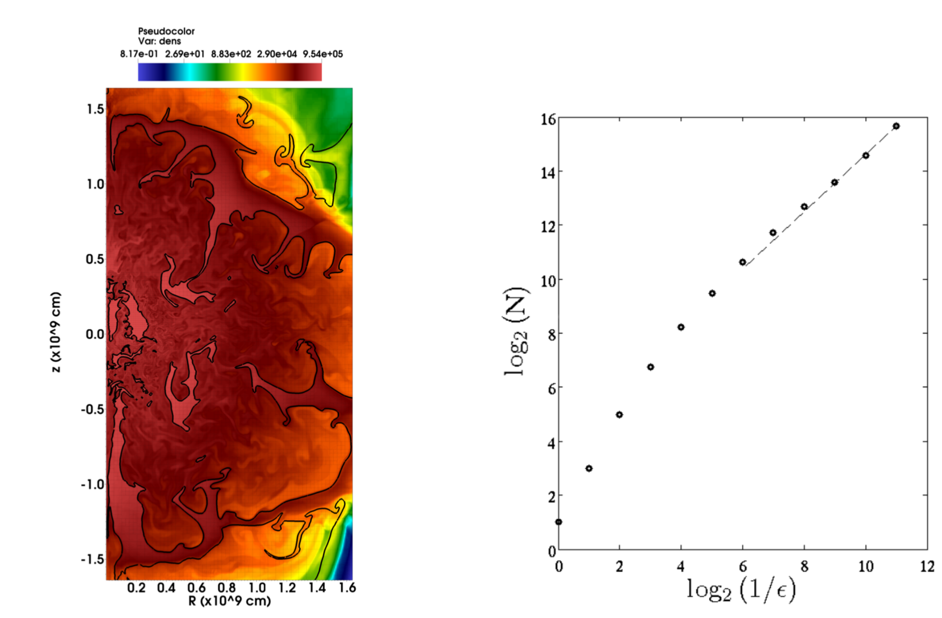
Figure 1: Organization of the LAVAFLOW package related to fractal dimension analysis with location of the source code (shown in blue) and executable building path (shown in red). This structure is generic to other LAVAFLOW analysis modules.

LAVAFLOW's modularity also enables continual development of the package. We are currently investigating the role of diffusive processes in magnetized turbulence. The corresponding analysis modules will be made available in the next release of the package.

Current Capabilities

We use LAVAFLOW for problems in stellar evolution and high-energy density plasma physics. These classes of applications require solving various physics such as magnetohydrodynamics, radiation transport, thermal diffusion, nuclear burning, and the effects of gravity to name a few. Analysis of our application results require obtaining problem-specific statistics such as

- energy flux decomposition for advection and heat transport
- kinetic energy spectra for turbulence
- structure functions for turbulence
- fractal analysis for combustion



(a) Flame Front **(b)** Fractal Dimension

Figure 2: Figure (a) is a model of a Type Ia supernova in cylindrical coordinates where z is the axis of symmetry. The False color shows the density in $\mathrm{g/cm^3}$. The contour of the flame front is shown with the black line. Thin, grey lines show the outline of mesh structure used in the model. Figure (b) shows the regression to find the fractal dimension of the flame contour. The slope of the regression line indicates the fractal dimension about 1.05.

where the analysis modules operate on a native mesh or, if necessary, a uniform mesh.

Parallelism

LAVAFLOW offers a high level of computational efficiency by using modular structure, dynamic memory management, and operating in parallel. The extensive use of parallelism enables the code to streamline performance. The work is distributed among the pool of allocated processes depending on the required amount of data. Table 1 highlights the package's efficiency and how the parallelism decreases the required wall time to perform analysis as the data size increases.

Resolution	Time (s)	Efficiency
128^3	0.5706	1.0
256^3	3.9462	6.9159
512^3	29.0738	50.9530

Table 1: Times required to perform fractal dimension analysis for turbulent combustion simulation results obtained on meshes with resolutions of 128^3, 256^3, 512^3 using 6 cores on a Linux desktop workstation. The times have been normalized by the 128^3 time to determine the efficiency.

Acknowledgements

The initial version of LAVAFLOW was developed by T. Handy. We thank Phil Boehner, Samuel Brenner, Angela Li, and Venkat Pamulapati for their recent contributions to LAVAFLOW.