**Project Proposal**

Xiaotong Liu

Meraj Khan

2014.3.21

**1. Introduction**

1. **Background**

Cosmological simulations are primarily defined by their volume and number of computational elements that discretize the mass in the Universe. Once chosen, each simulation is evolved from very small fluctuations in an otherwise uniform distribution using gravitational N-body integrators in an expanding background Universe. Over the 14 billion years of evolution, these particles cluster into gravitationally bound structures that pull in baryonic matter that form stars, galaxies, and clusters of galaxies. All of the datasets here are derived from the Dark Sky Simulations, which was awarded a DOE INCITE computing allocation at the level of 80M cpu-hours. The largest simulations run from this project cover nearly 12 Gigaparsecs on a side (38 billion light-years across), and uses 1.1 trillion particles to discretize the volume, totalling nearly half a Petabyte of output. This year’s contest will have access to both the very large-scale data releases from the Dark Sky Simulations project, as well as smaller volumes and particle counts to develop the visualization methods and user interface.

1. **Task Overview**

To-be-added

1. **Scope and Expected Results**

To-be-added

**2. Data set**

There are **three primary types** of data that will be utilized in this years contest. The **first** is the **raw particle data** that is described by a position vector, velocity vector, and unique particle identifier. Each snapshot in time, for which there will be approximately 100, is stored in a single file in a format called **SDF** (https://bitbucket.org/JohnSalmon/sdf). This format is composed of a human readable ASCII header followed by raw binary data. Python and C-based interfaces to the data format will be provided. The **second** type of dataset is called a **Halo Catalog**, and it defines a database that groups **sets** of gravitationally bound particles together into **coherent structures**. Along with information about a given halo’s position, shape, and size, are a number of statistics derived from the particle distribution, such as angular momentum, relative concentration of the particles, and many more. These catalogs are stored in both ASCII and binary formats. The final dataset type links the individual halo catalogs that each represents a snapshot in time, thereby creating a **Merger Tree** database. These merger tree datasets **form a sparse graph** that can then be analyzed to use quantities such as halo mass accretion and merger history to inform how galaxies form and evolve through cosmic time. Merger tree databases are also distributed in both ASCII and BINARY formats.

**3. Tasks**

Select one or more from: <http://darksky.slac.stanford.edu/scivis2015/tasks.html>

* T1: a visualization framework for particle data
* T2: a visualization framework for (time-varying,) multi-faceted halo properties
* T3: a data processing framework for incorporating semi-analytical model (as outputs to new dimensions/facets to T2)
* T4: Comparative visualization of Halo substructures

To-be-discussed

**4. Software**

* vtk
* python-yt

**Yt** has built-in readers for the **SDF** data format. The I/O routines that are in yt have also been packaged into https://pypi.python.org/pypi/sdfpy, which can be installed with 'pip install sdfpy'.

**5. Visualization Design**

Particle visualization?

Interaction: zooming, panning, rotating views, brushing

To-be-discussed

**6. Expected Results and Evaluations**

To-be-added

**Project Requirements**

I. Introduction:

-  Describe the background of the cosmology application for the contest. -  Overview the specific tasks listed in the contest that you will be tackling -  Define the scope of your project and the expected results.

III. The tasks:

-  Explain **the selected tasks** in more detail

-  Describe the algorithmic and data analysis components required to perform the tasks

II. The data:

- Describe the contest datasets: the formats, the information contained in the data sets and the relationships between the data sets.

- If you will perform any transformation, describe your methods and explain how the data transformation will help to perform the tasks.

IV. Software:- Describe the **existing software** that you will be using to read, process, and analyze the data. - Describe the software that you will be creating

V. Visualization design

- Describe the visualization that you will be creating for the tasks - Describe how users can interact with the data

VI: Expected results and evaluation

- Describe any actions that you have taken for the project (downloading and cleaning data, etc.)

- Present any preliminary results if any

- Describe the expected results from your project