CS732 Data Visualization - Assignment 1
IEEE Visualization 2008 Design Contest

Shathin R Rao [**MT2020067**]

Abstract

This report describes the tools and methodology used for visualizing different scalar fields from the data set provided as part of the **IEEE Visualization 2008 Design Contest**. The data was obtained from a numerical simulation of the first stars in the galaxy. The report then attempts to make certain inferences from the visualizations generated.

1 Introduction

Scientists have performed three-dimensional radiation hydro-dynamical calculations of ionization front instabilities in which multi-frequency radiative transfer is coupled to eight species primordial chemistry. They have simulated ten scalar fields - **Particle Density**, **Gas Temperature**, and **8 chemical species** (H, $H+$, $H-$, He, $He+$, $He++$, H_2, H_2+) and the **Velocity** (vector) field. The **Tools** section describes the third-party libraries used and the script written to generate the visualizations. The **Data** section briefs about the data that is used for the visualization. The technique attempted for the visualization involve **Contour Mapping (Contour Fill)** and **Quiver (Arrow) plots**. The **Visualization and Inferences** section elaborates on the chosen technique for each field and attempts to make certain inferences from the generated visualization.

2 Tools

This sections describes the libraries chosen for the visualization and the briefs about the scripts written using the libraries to generate the visualizations. **Python** was used as the programming language for generating the visualization.Python was chosen primarily because Python is a very mature language when it comes to the domain Data Analysis and hence can deal with large datasets.

2.1 Libraries

The set of libraries used for generating the visualizations involved -

- numpy → (Library Home) **NumPy** is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

- pandas → (Library Home) **pandas** is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. It provides methods to read data from files and transform them as needed.

- matplotlib → (Library Home) **Matplotlib** is a comprehensive library for creating static, animated, and interactive visualizations in Python.

2.2 Scripts

All the scripts mentioned below have been written in a way to enable fast iteration of the data i.e., with very minimal changes to the code, new visualizations can be generated.

- dataOps.py → Contains methods that reads and processes the data

- formulae.py → Contains the different scalar formulae that can be applied on the fields in the dataset

- scalarFields.py → A configuration file that defines each of the given scalar fields and the derived fields. Each field is a dictionary containing different attributes that define the properties of the field.

- scalarVisualization.py → The python script that reads and visualizes the desired scalar field from the data set. The file contains methods for generating both static and animated contour fill plots.

- vectorFields.py → A configuration file that defines each of the given vector fields and the derived fields. Each field is a dictionary containing different attributes that define the properties of the field.

- vectorVisualization.py → The python script that reads and visualizes the Curl of the Velocity from the vector data set. The file contains method for generating a static quiver plot.

The codebase can be found on Github - Shathin/cs732-data-visualization

3 Data

This section briefly describes the given scalar fields and the derived fields that was made use of for the visualization.

3.1 Scalar Fields

- **Gas Temperature** $(Kelvin) \rightarrow$ A *given* scalar field, defines the temperature of the gas in the region.
- **Total Particle Density** ($\#$ of $particles$ / cm^3) $\rightarrow$ A *given* scalar field, defines the number of particles (of all the chemical species) in a given region.
- **Hydrogen Number Density** ($\#$ of $particles$ / cm^3) $\rightarrow$ A *derived* scalar field, defines the number of Hydrogren particles in a given region.
- **Hydrogen+ Number Density** ($\#$ of $particles$ / cm^3) $\rightarrow$ A *derived* scalar field, defines the number of (positively) ionized Hydrogen particles in a given region.
- **Hydrogen- Number Density** ($\#$ of $particles$ / cm^3) $\rightarrow$ A *derived* scalar field, defines the number of (negatively) ionized Hydrogen particles in a given region.

3.2 Vector Fields

- **Velocity** $(km/s) \rightarrow$ The vector data set contains the i, j and k components of the velocity.
- **Curl of Velocity** $(km/s) \rightarrow$ The curl of the velocity a.k.a vorticity or the rotation, represents the rotation of the velocity at a given in three dimensional space. This is vector field.
- **Magnitude of Curl** $(km/s) \rightarrow$ The magnitude of the Curl of the velocity. This is scalar field.

3.3 Computing Derived Fields

3.3.1 Total Density

Before computing the selected derived fields, the total density is to be computed. The total density (g/cm^3) is a function of the mass abundance of all the chemical species (Xi_x) and the total particle density *tpd*. Following the formula used for computing it -

$$\rho = \frac{tpd * mh}{Xi_H + Xi_{H+} + 0.25 * (Xi_{He} + Xi_{He+} + Xi_{He++}) + Xi_{H-} + 0.5 * (Xi_{H_2} + Xi_{H_2+})}$$

Where mh is the Mass of Hydrogen i.e., $mh = 1.38066 * 10^{-24} g$

3.3.2 H, H+ and H- Number Density

Hydrogen, Hydrogen+ and Hydrogen- Number densities (nd_x) are calculated using the following common formula -

$$nd_x = Xi_{Hx} * \frac{\rho}{mh}$$

Where Xi_{Hx} is Xi_H, Xi_{H+} or Xi_{H-} i.e., the mass abundance of H, $H+$ and $H-$ chemical species, ρ is the total density and mh is the Mass of Hydrogen.

3.3.3 Curl of Velocity

If v is the velocity at $(x,\ y,\ z)$ co-ordinates of the grid, then to compute the curl at $(x,\ y,\ z)$ we required the velocity vector at the co-ordinates $(x+1,\ y,\ z)$, $(x,\ y+1,\ z)$ and $(x,\ y,\ z+1)$ as well. Following is the formula used to calculate the Curl of the velocity at $(x,\ y,\ z)$ -

$$curl_i = v_k(x, y+1, z) - v_k(x, y, z) - v_j(x, y, z+1) + v_j(x, y, z)$$

$$curl_j = v_i(x, y, z+1) - v_i(x, y, z) - v_k(x+1, y, z) + v_k(x, y, z)$$

$$curl_k = v_j(x+1, y, z) - v_j(x, y, z) - v_i(x, y+1, z) + v_i(x, y, z)$$

Following is the formula used to calculate the magnitude of the Curl -

$$|curl| = \sqrt{curl_i^2 + curl_j^2 + curl_k^2}$$

4 Visualization and Inferences

This section describes the methodology used for the visualizations and the inferences made from the visualizations.

4.1 Choosing Timesteps

Out of the entire dataset available only the following timesteps were chosen - 1, 2, 3, 6, 9, 14, 19, 29, 49, 69, 99, 129, 159, 189.

The reason for choosing such a timestep that isn't uniformly distributed is because upon trying to visualize the entire dataset i.e., all the timesteps it was observed that most of the action happens quickly during the initial few timesteps and then propagates slowly i.e., not much happens between contiguous timesteps.

4.2 Choosing Data Slice

The Z axis was kept constant at $Z = 124$ as it will provide a value in the center of the data set. Also, Z being the slowest changing axis it becomes convenient to keep it fixed.

4.3 Visualizations

4.3.1 Gas Temperature

- **Technique** - *Contour Fill*: Contour Fill (Mapping) was chosen to group the regions of (approximately close) temperature together.
- **Color Map** - *Hot [Reverse]*: When one thinks of temperature, one generally thinks of the color red and hence the **Hot** sequential color map was chosen.

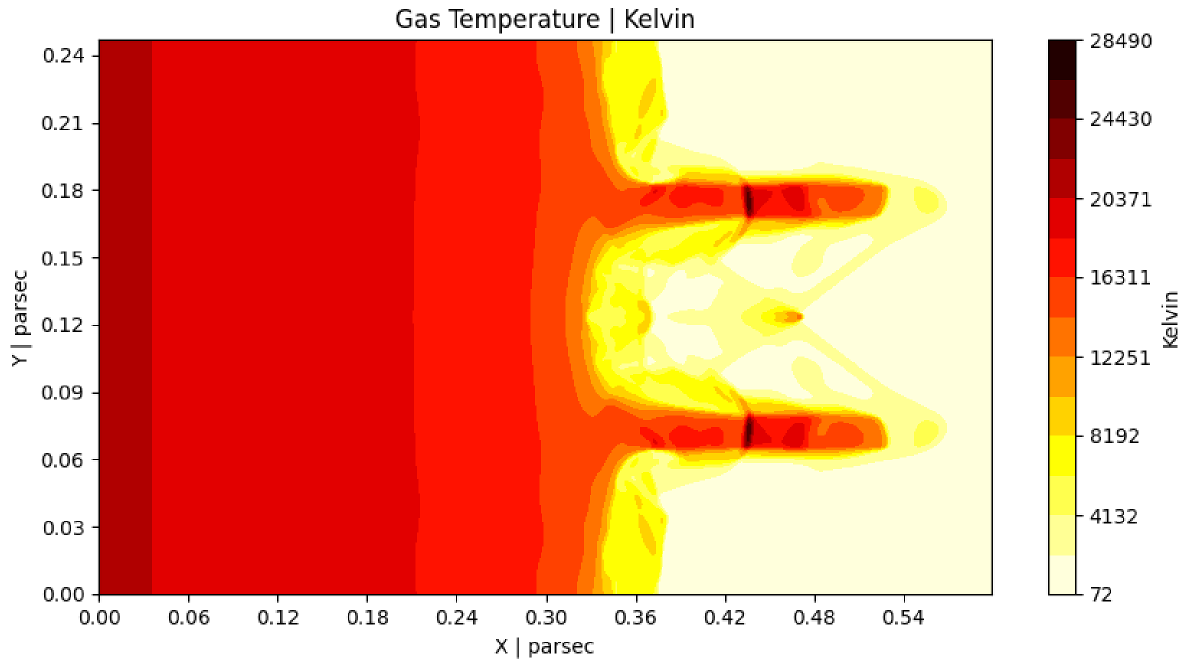

Figure 1: Gas Temperature at timestamp 99 [12.5 kYear]

- **Normalization** -*Min-Max Normalization*: Since the unit of the temperature data provided is in **Kelvin**, no negative values are required to represent temperature. Hence a normal min-max normalization proves sufficient.
- **Colorbar Scale** - *Linear*: Since the values aren't extremely large, a linear colorbar was chosen.
- **Values Utilized** - *Gas Temperature [Kelvin]*: The values utilized from the data set was only the Gas Temperature field. The X and Y co-ordinates were computed based on the information given in the contest webpage.

4.3.2 Total Particle Density

- **Technique** - *Contour Fill*: Contour Fill (Mapping) was chosen to group the regions of (approximately close) particle (number) density together.
- **Color Map** - *Blue-Purple (BuPu) [Reverse]*: This color map was chosen since it was perceptually uniform color map. The reverse of the color map was chosen since region is very narrow and using the non-reversed color map makes it hard to see the regions.
- **Normalization** -*Min-Max Normalization*:Since the unit of the data provided is in number of particles per cm^3, no negative values are required to represent total particle density. Hence a normal min-max normalization proves sufficient.

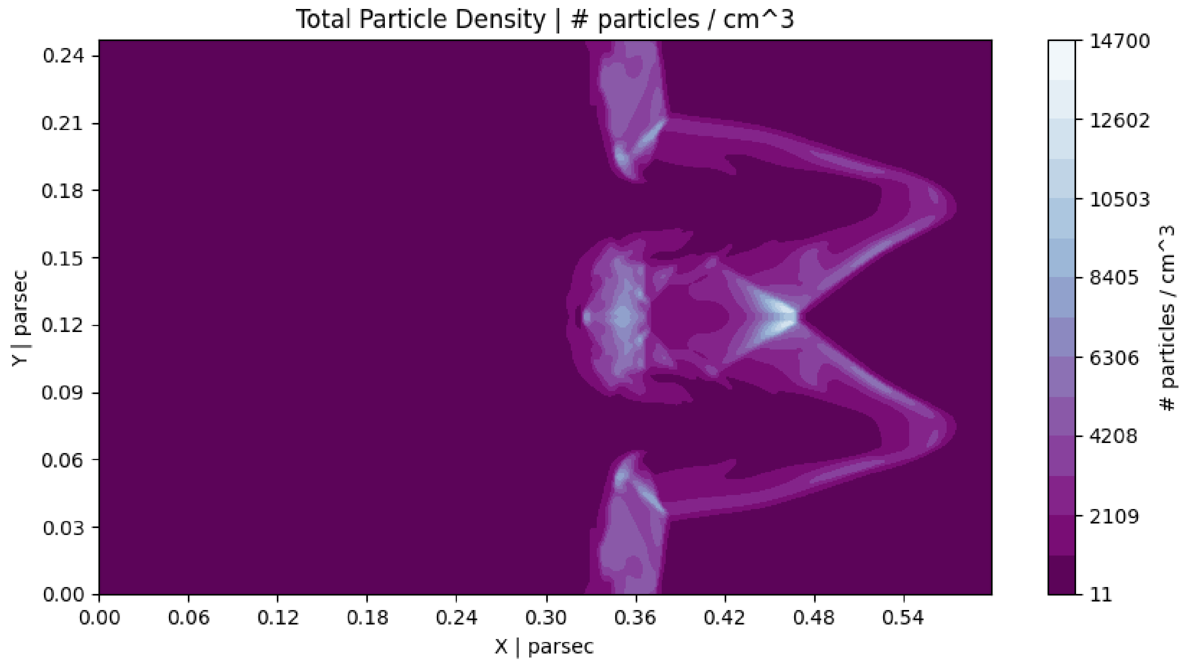

Figure 2: Total Particle Density at timestamp 99 [12.5 kYear]

- **Colorbar Scale** - *Linear*: Since the values aren't extremely large, a linear colorbar was chosen.
- **Values Utilized** - *Total Particle Density [# of particles / cm^3]*: The values utilized from the data set was only the Total Particle Density field. The X and Y co-ordinates were computed based on the information given in the contest webpage.

4.3.3 H Number Density

- **Technique** - *Contour Fill*: Contour Fill (Mapping) was chosen to group the regions of (approximately close) Hydrogen particle (number) density together.
- **Color Map** - *Blues [Reverse]*: This color map was chosen since it was perceptually uniform color map. The reverse of the color map was chosen since region is very narrow and using the non-reversed color map makes it hard to see the regions.
- **Normalization** -*Min-Max Normalization*:Since the unit of the data provided is in number of particles per cm^3, no negative values are required to represent Hydrogen Number density. Hence a normal min-max normalization proves sufficient.
- **Colorbar Scale** - *Linear*: Since the values aren't extremely large, a linear colorbar was chosen.
- **Values Utilized** - *H Number Density [# of particles / cm^3]*: This is a derived field and hence to compute this total particle density field and all the chemical species fields had to be used for

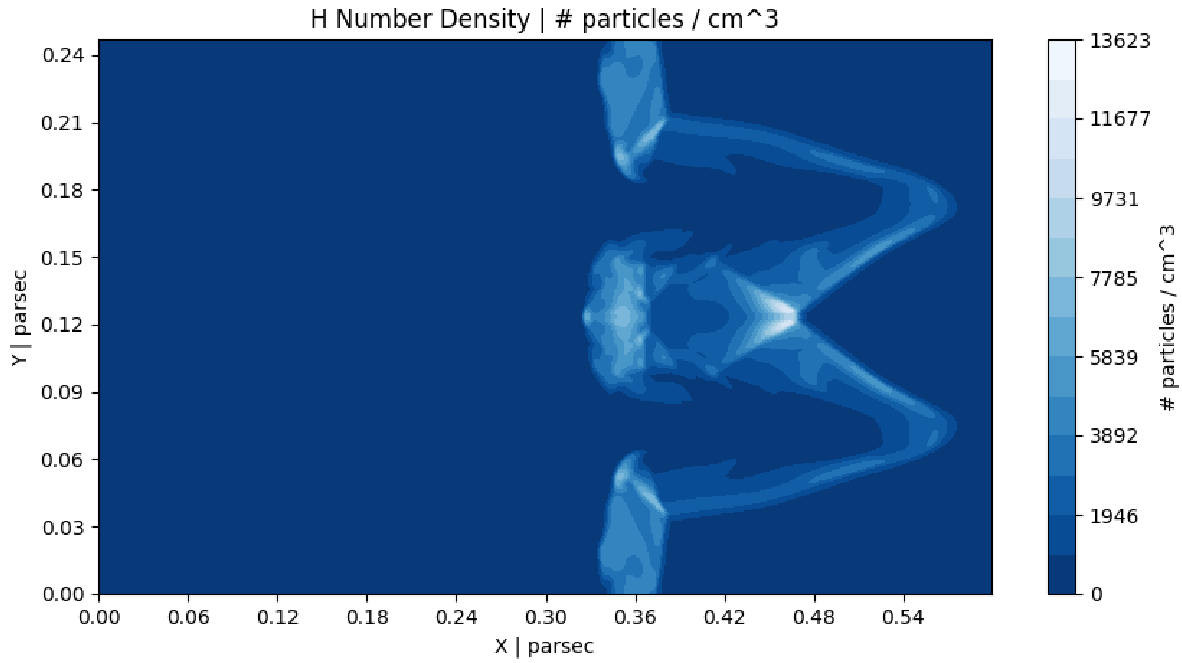

Figure 3: H Number Density at timestamp 99 [12.5 kYear]

the computation. The X and Y co-ordinates were computed based on the information given in the contest webpage.

4.3.4 H+ Number Density

- **Technique** - *Contour Fill*: Contour Fill (Mapping) was chosen to group the regions of (approximately close) Hydrogen+ particle (number) density together.
- **Color Map** - *Reds*: This color map was chosen since it was perceptually uniform color map. Red was chosen because ionized gas is very hot.
- **Normalization** -*Min-Max Normalization*:Since the unit of the data provided is in number of particles per cm^3, no negative values are required to represent Hydrogen+ Number density. Hence a normal min-max normalization proves sufficient.
- **Colorbar Scale** - *Linear*: Since the values aren't extremely large, a linear colorbar was chosen.
- **Values Utilized** - *H+ Number Density [# of particles / cm^3]*: This is a derived field and hence to compute this total particle density field and all the chemical species fields had to be used for the computation. The X and Y co-ordinates were computed based on the information given in the contest webpage.

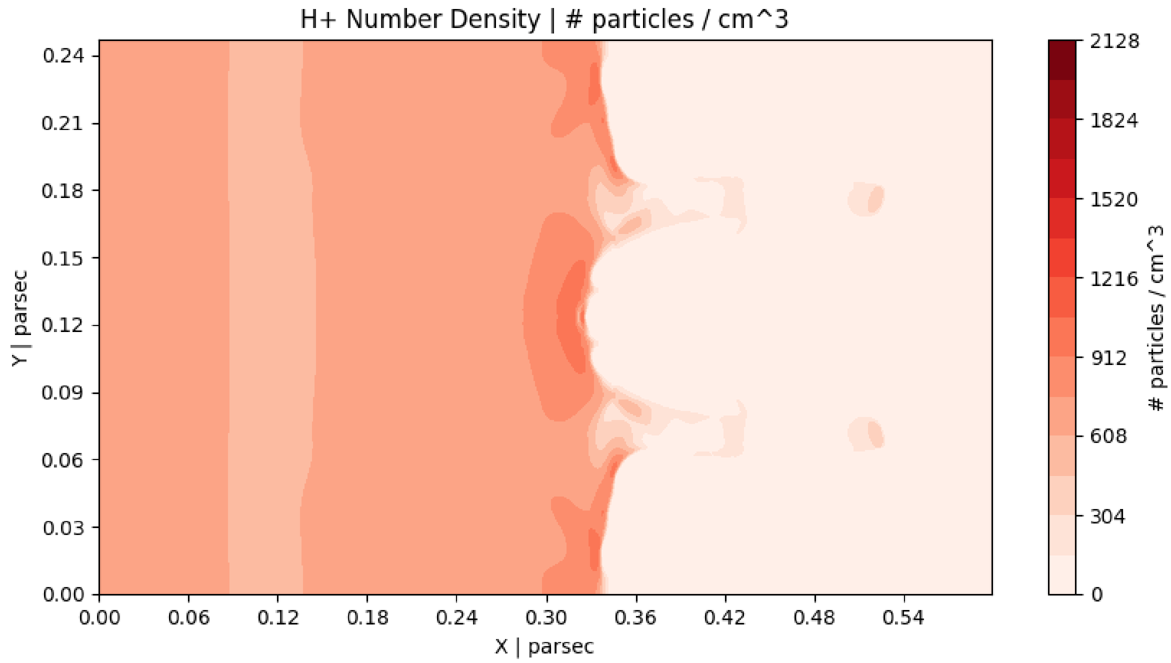

Figure 4: H+ Number Density at timestamp 99 [12.5 kYear]

4.3.5 H- Number Density

- **Technique** - *Contour Fill*: Contour Fill (Mapping) was chosen to group the regions of (approximately close) Hydrogen- particle (number) density together.
- **Color Map** - *Greens [Reverse]*: This color map was chosen since it was perceptually uniform color map. The reverse of the color map was chosen since region is very narrow and using the non-reversed color map makes it hard to see the regions.
- **Normalization** -*Min-Max Normalization*: Since the unit of the data provided is in number of particles per cm^3, no negative values are required to represent Hydrogen+ Number density. Hence a normal min-max normalization proves sufficient.
- **Colorbar Scale** - *Linear*: Since the values aren't extremely large, a linear colorbar was chosen.
- **Values Utilized** - *H+ Number Density [# of particles / cm^3]*: This is a derived field and hence to compute this total particle density field and all the chemical species fields had to be used for the computation. The X and Y co-ordinates were computed based on the information given in the contest webpage.

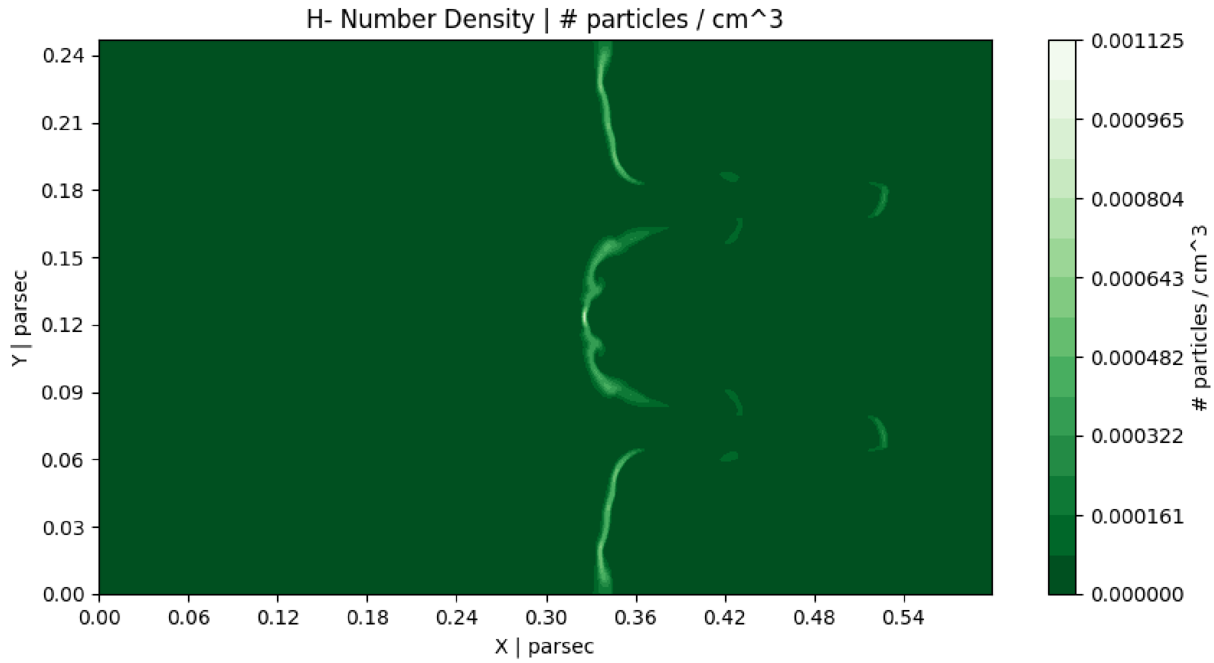

Figure 5: H+ Number Density at timestamp 99 [12.5 kYear]

4.3.6 Curl of Velocity

- **Technique** - *Quiver Plot*: Quiver Plot (a.k.a Arrow Plot) was chose to visualize the direction and magnitude of the Curl of the Velocity.

- **Color Map** - *Viridis*: Viridis is a robust color scheme which helps improve graph readability for readers with common forms of color blindness and/or color vision deficiency.

- **Scale & Normalization** - *Min-Max Normalization*: The units of both the Curl and X & Y co-ordinates were converted to $parsec/s$ and $parsec$, respectively, so that they match. After this, the lengths of the arrows were made to scale as per the X and Y co-ordinates so that lengths of the arrows in different timesteps are accurate. Furthermore, the magnitude, which was used for the color map was normalized between the minimum and maximum of the magnitude in the entire (selected) dataset. Also, vectors of length 0 are not shown on the plot; this was done in order to get a cleaner plot.

- **Values Utilized** - *i & j components of the Curl and Magnitude of the Curl*: The derived vector field Curl of the Velocity and the scalar field Magnitude of the Curl had to be computed. The X and Y co-ordinates were computed based on the information given in the contest webpage.

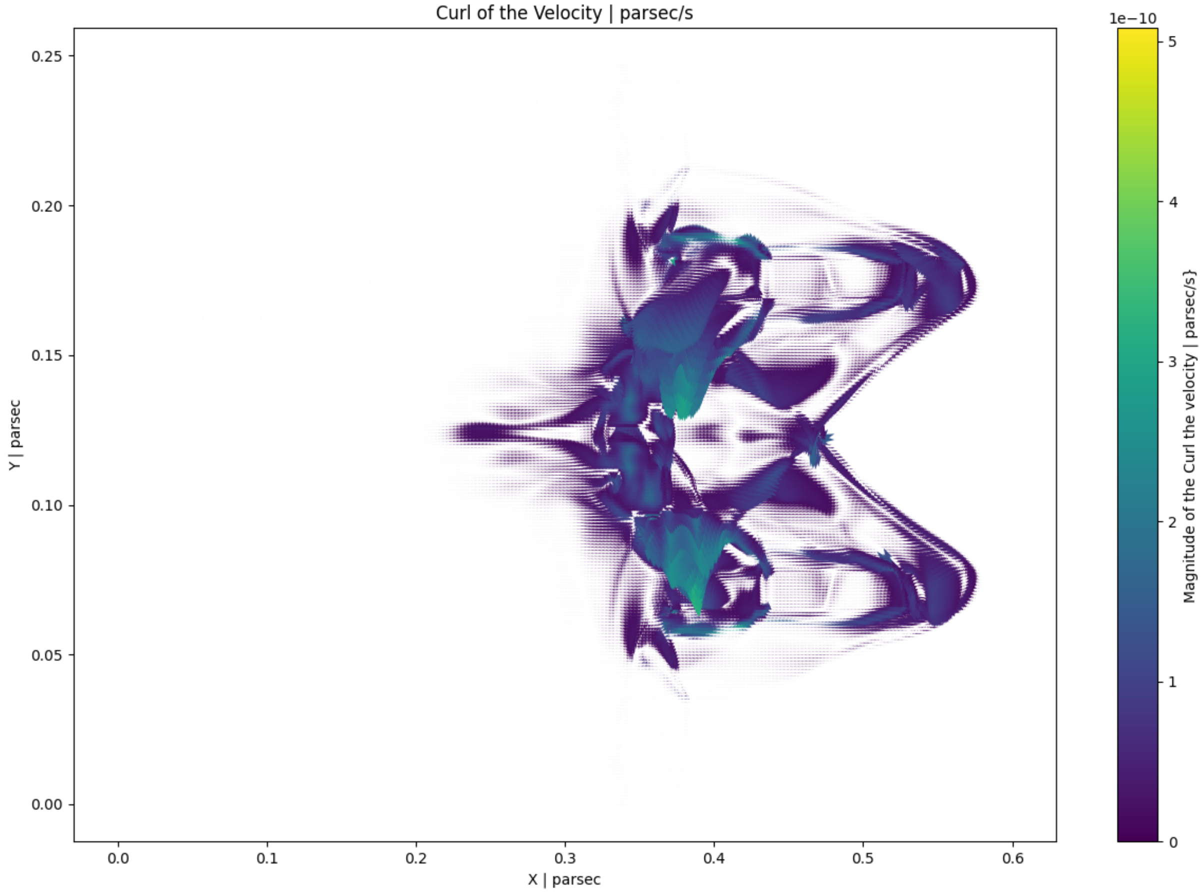

Figure 6: Curl of Velocity at timestamp 99 [12.5 kYear]

4.4 Inferences

- By looking at the Contour Fill Plots for the **Gas Temperature**, we can observe that the ionization front rapidly heats up the temperature of the gases in the region. This is evident by looking at the width of the regions closer to the front of the ionization front. The width of regions towards the left are much wider than the the regions towards the right i.e., closer to the ionization front.

- Another thing that can be observed from **Gas Temperature** plot is that the temperature continues to rise even after the ionization front passes a region and is much higher than the temperature at the ionization front.

- From the Contour Fill plots for the **Total Particle Density**, we can observe that more particles

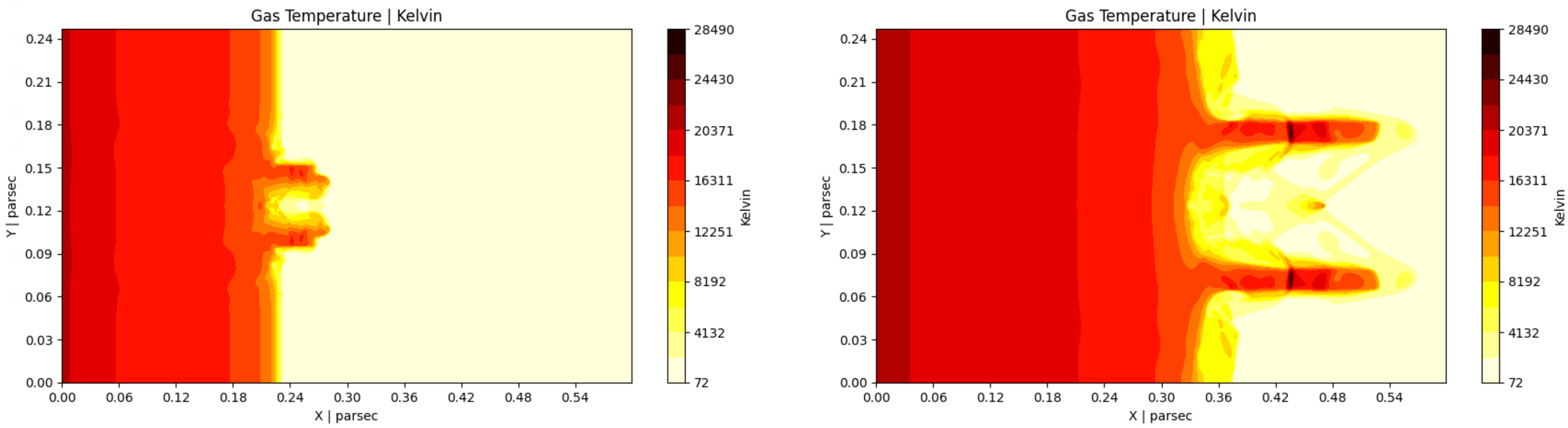

Figure 7: Gas Temperature at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

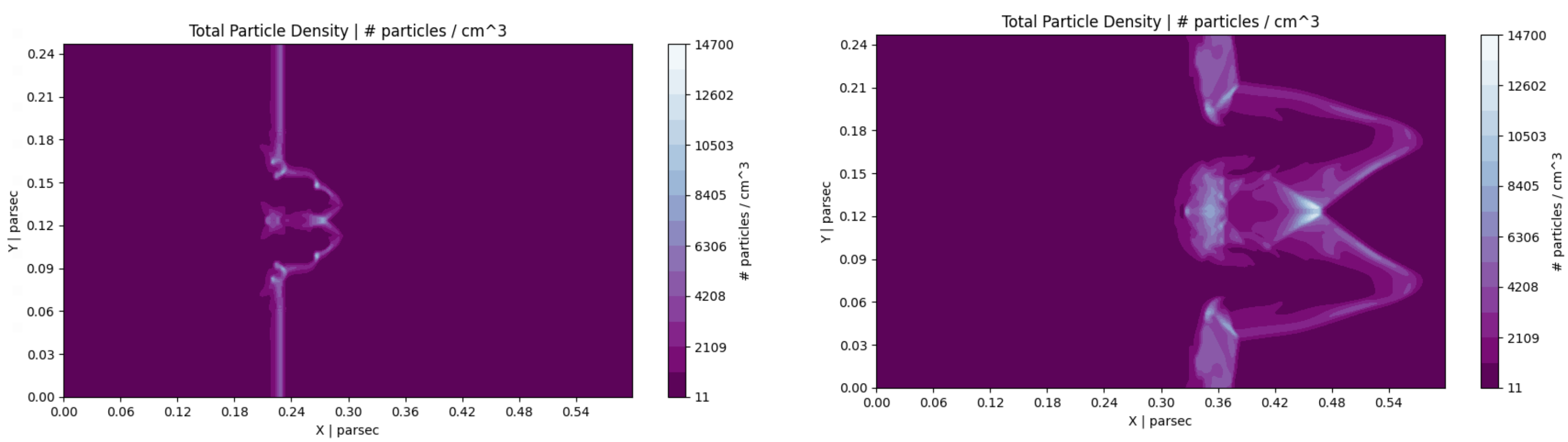

Figure 8: Total Particle Density at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

are concentrated in front of the ionization front leading to the observation that the ionization front has a compressing effect on the gas particles in the region.

- From the Contour Fill plots for the $H+$ **Number Density** it is evident that more particles are concentrated right behind the ionization front leading to the observation that the ionization front leaves behind a trail of hot ionized gas.

- From the Contour Fill plots for the $H-$ **Number Density** it is evident that particles are primarily concentrated closer to the ionization front leading to the observation that $H-$ particles are extremely unstable and don't remain for long.

- Looking at the plots of the H**,** $H+$ **and** $H-$ **Number Densities**, the following can be said -
  - Not all the Hydrogen particles are converted to Hydrogen+ ions and Hydrogen-. This can be said by looking at the maximum number of particles in each graph.

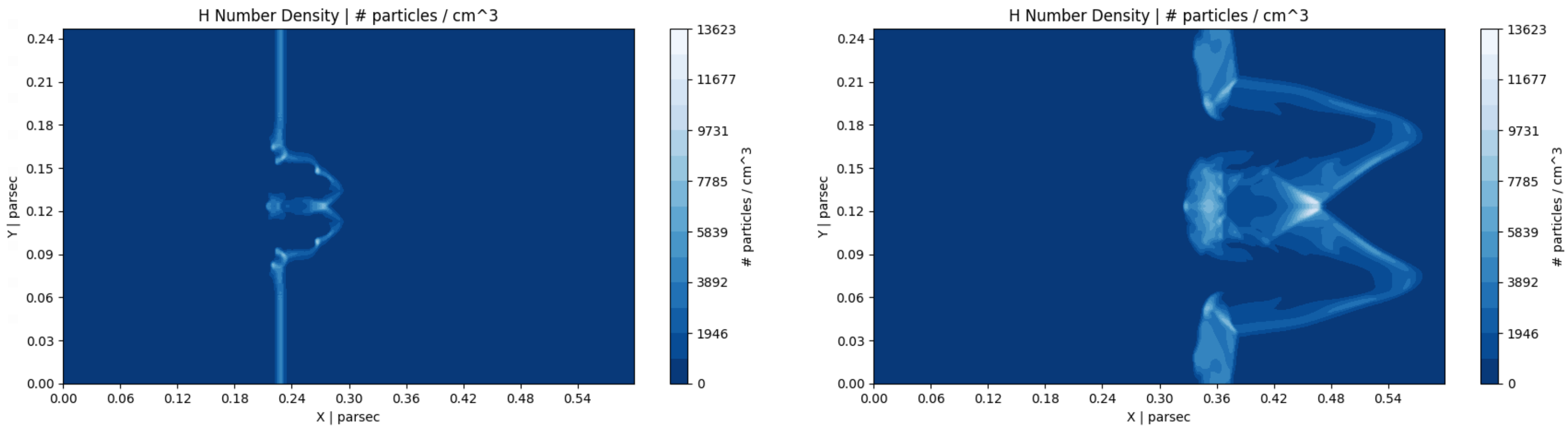

Figure 9: H Number Density at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

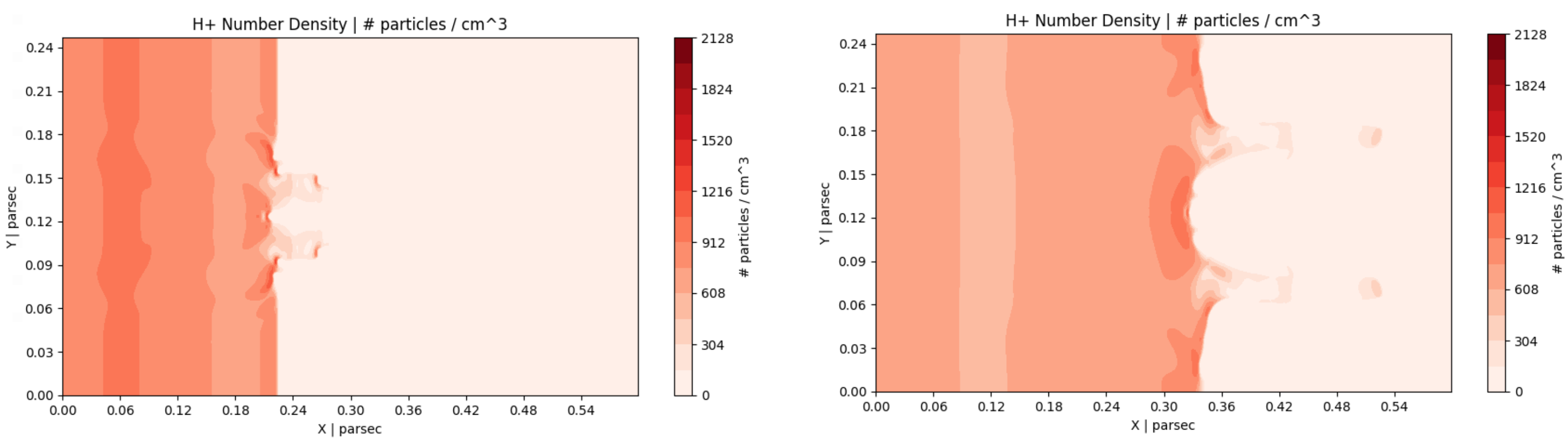

Figure 10: H+ Number Density at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

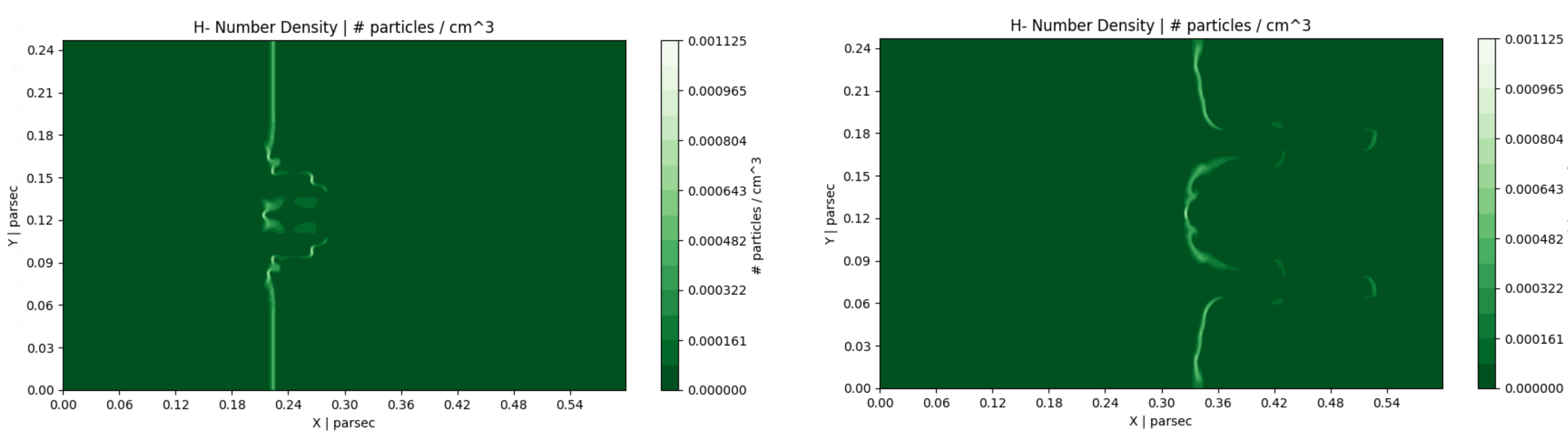

Figure 11: H- Number Density at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

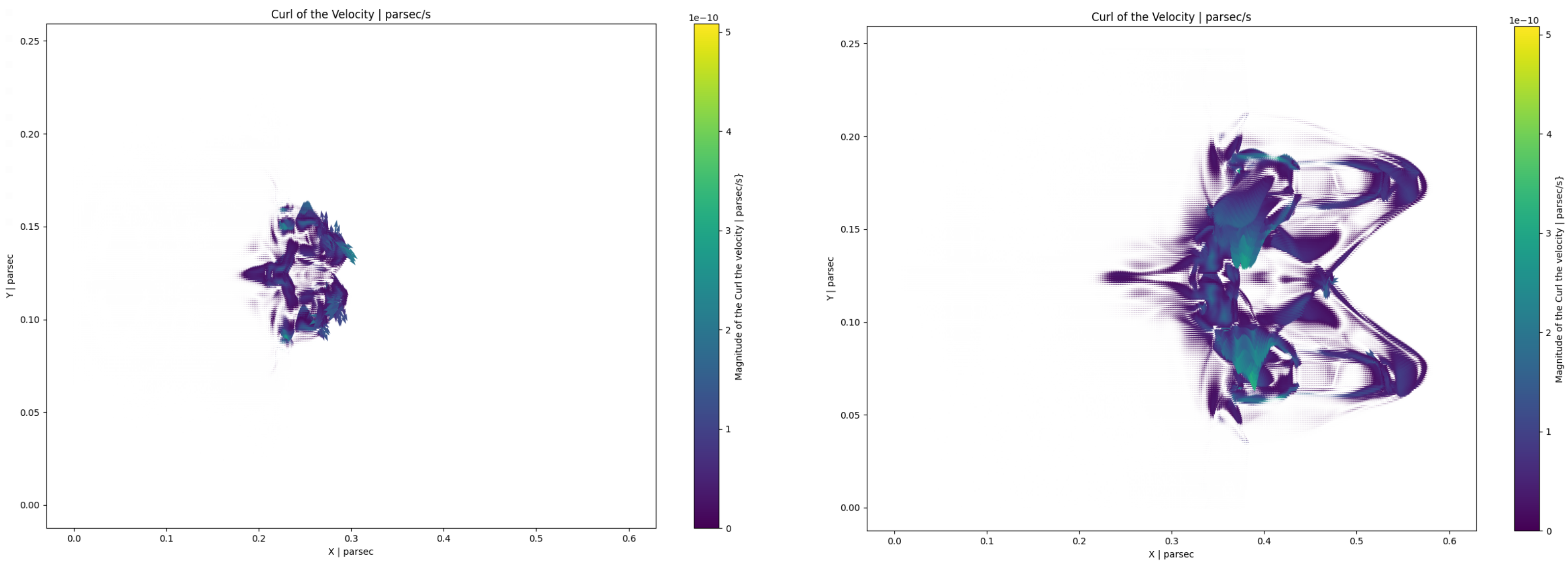

Figure 12: Curl of Velocity at timestep (a) 29 [4.5 kYear] and (b) 99 [12.5 kYear]

  - The conversion rate of Hydrogen to Hydrogen+ is far greater than Hydrogen to Hydrogen-.
  - The Hydrogen- ions only exist for a small period of time and only exist right behind the ionization front. Hence are very unstable compared to H+ ions.

- Looking at the **Gas Temperature**, $H+$ and $H-$ **Number Density** plots we can say the the temperature of $H+$ ions is far greater than the temperature of $H-$ ions.

- The **Curl of Velocity** plot gives us the idea about why the shape of the ionization front changes the way it does i.e., it grows from a single point into a large bow shaped front.

4.5 Experimentation

This section show some experimental plots that were generated by combining the Contour Fill plots of the different scalar fields with the Quiver plot of the Curl of Velocity field

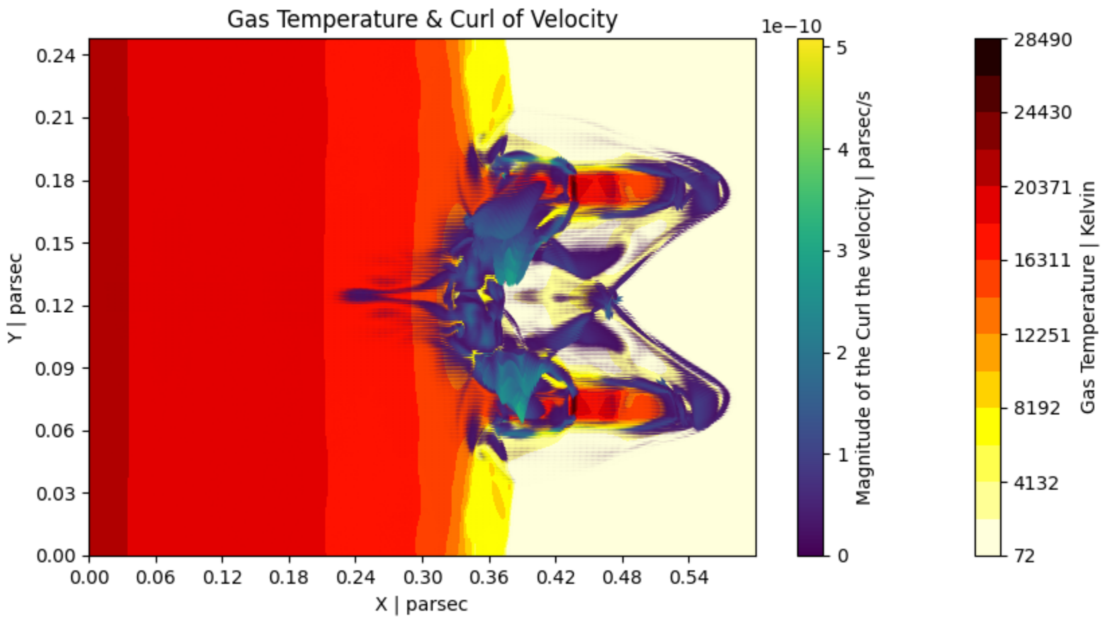

Figure 13: Gas Temperature and Curl of Velocity at timestep 99 [12.5 kYear]

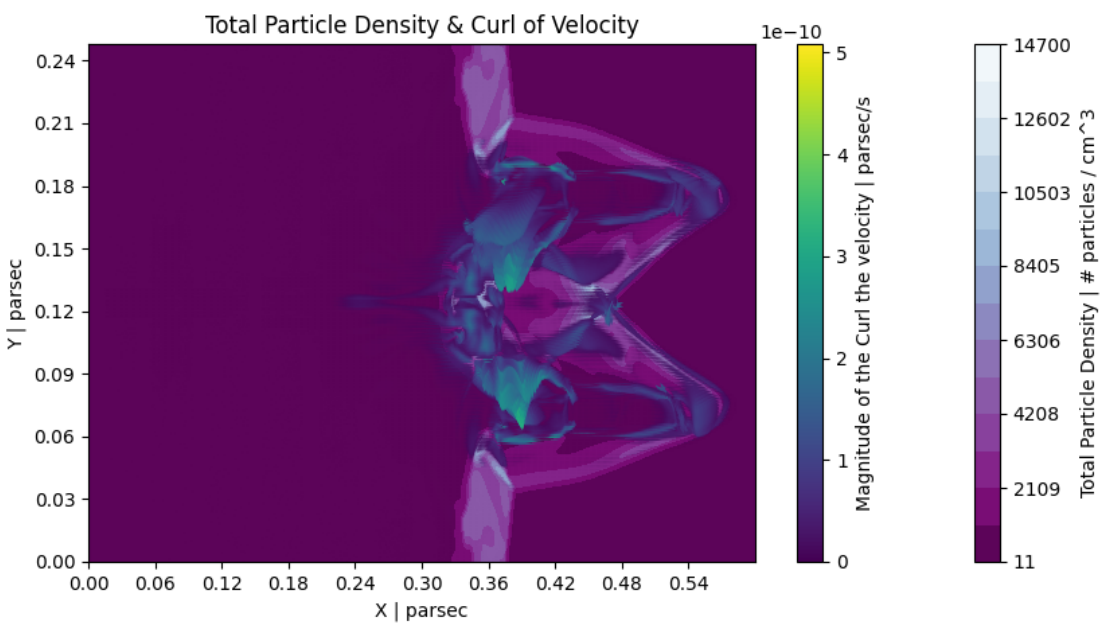

Figure 14: Total Particle Density and Curl of Velocity at timestep 99 [12.5 kYear]

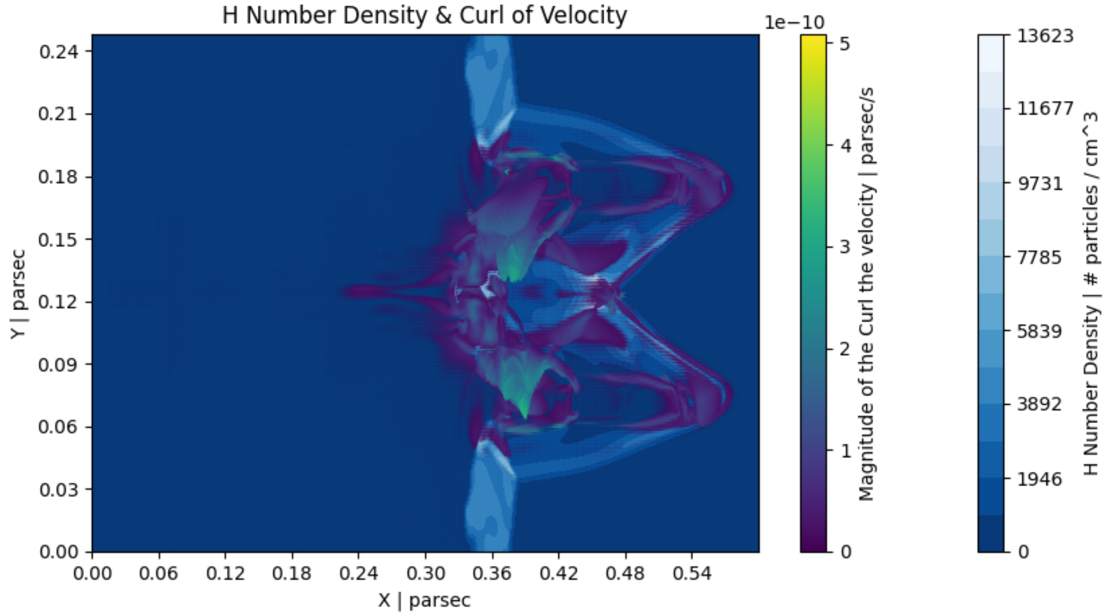

Figure 15: H Number Density and Curl of Velocity at timestep 99 [12.5 kYear]

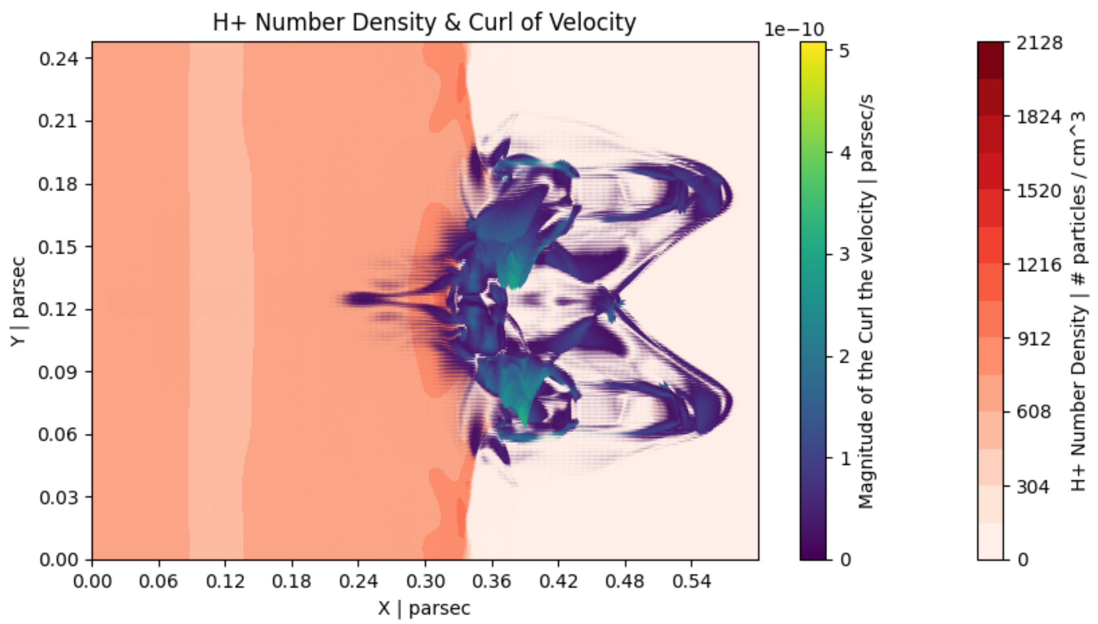

Figure 16: H+ Number Density and Curl of Velocity at timestep 99 [12.5 kYear]

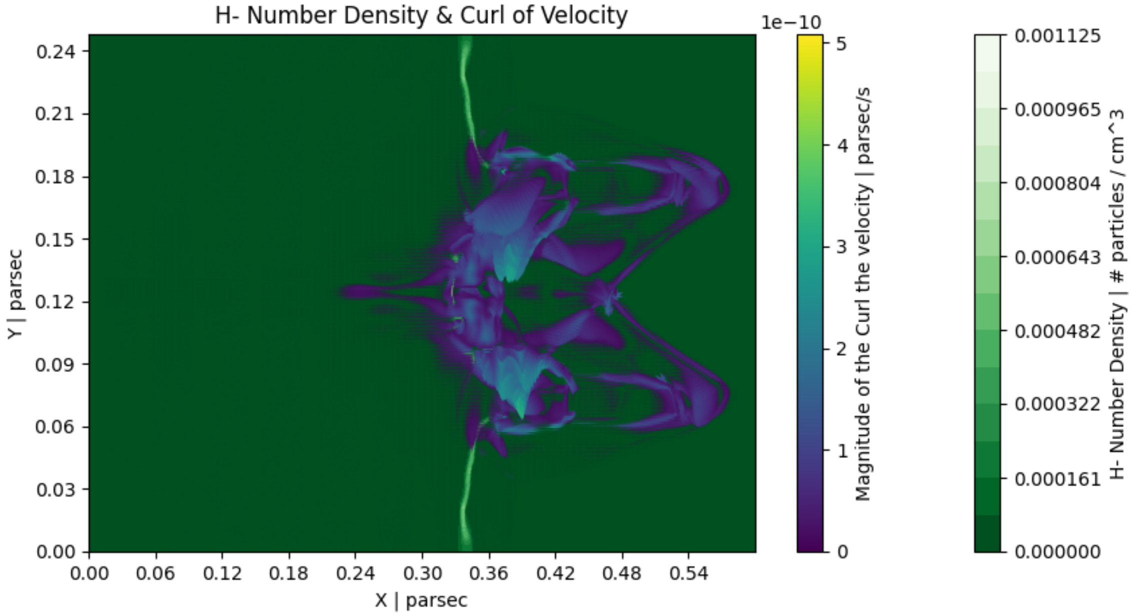

Figure 17: H- Number Density and Curl of Velocity at timestep 99 [12.5 kYear]

5 Conclusion

In conclusion, the report has successfully elaborated on the visualizations performed and the tools & methodology used to perform the required tasks. Such a visualization helps slightly with understanding how the vector field affects the scalar fields