

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Interactive Analysis Tools for Visualizing the Universe

Permalink

<https://escholarship.org/uc/item/31k720gm>

Author

Abramov, David

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**INTERACTIVE ANALYSIS TOOLS FOR VISUALIZING THE
UNIVERSE**

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTATIONAL MEDIA

by

David Abramov

June 2023

The Dissertation of David Abramov
is approved:

Professor Angus Forbes, Chair

Professor Michael Mateas

Professor Joseph N. Burchett

Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by
David Abramov
2023

Table of Contents

List of Figures	vi
List of Tables	xii
Abstract	xiv
Acknowledgments	xvi
1 Introduction	1
1.1 Projects Overview	1
1.2 Brief history of astrovis through the ages	5
1.2.1 Stone age astrovis	5
1.2.2 Telescopes: onward and upward	11
1.2.3 Astronomy visualization in the 20th century	17
1.2.4 Digitization: from observations to simulations	24
2 IGM-Vis: Analyzing Intergalactic and Circumgalactic Medium Absorption Using Quasar Sightlines in a Cosmic Web Context	29
2.1 Abstract	29
2.2 Introduction	31
2.3 Background & Related Works	33
2.4 IGM/CGM Task Analysis	35
2.4.1 Data Collection Tasks	39
2.4.2 Identification Tasks	40
2.4.3 Analysis Tasks	42
2.4.4 Presentation Tasks	43
2.5 The <i>IGM-Vis</i> Application	44
2.5.1 Universe Panel	45
2.5.2 Galaxy Panel	46
2.5.3 Spectrum Panel	48
2.5.4 Equivalent Width Plot Panel	49
2.5.5 Implementation Details	49
2.6 Evaluation	50
2.6.1 Use Case 1: Investigating Coherence of Multisightline Absorption Signals	52

2.6.2	Use Case 2: Investigating Absorption Patterns among Galaxy Filaments using Equivalent Width Plots	53
2.7	Conclusion & Future Work	54
2.8	Acknowledgements	55
3	CosmoVis: An Interactive Visual Analysis Tool for Exploring Hydrodynamic Cosmological Simulations	56
3.1	Abstract	56
3.2	Introduction	57
3.3	Cosmological Background	59
3.4	Related Work	62
3.5	Analysis Tasks	64
3.5.1	Task 1: Identify structural patterns in simulation data	69
3.5.2	Task 2: Simulate quasar absorption line spectroscopy	71
3.5.3	Task 3: Connect IGM/CGM observables to gas physical conditions	72
3.5.4	Task 4: Compare simulation results	73
3.6	Application Design of <i>CosmoVis</i>	75
3.6.1	Data Preprocessing	77
3.6.2	Interactive Rendering	78
3.6.3	Skewer Interaction	79
3.6.4	User Interface	80
3.6.5	Client-Server Architecture	83
3.7	Scientific Use Cases	84
3.7.1	Case 1: Identifying and Examining Sheet Regions	85
3.7.2	Case 2: Simulating a QSO absorption line study of the circumgalactic medium	88
3.7.3	Case 3: Metal enrichment of the intergalactic medium	89
3.8	Evaluation	90
3.9	Conclusion	93
3.10	Supplemental Material	94
4	Visualizing the Temporal Evolution of Cosmological Simulations	103
4.1	Abstract	103
4.2	Introduction	104
4.3	Background and Related Work	109
4.4	Analysis Tasks	112
4.4.1	T1: Identify key phases of structure formation	113
4.4.2	T2: Track galactic life cycles	114
4.4.3	T3: Compare galaxy evolution across environments	115
4.4.4	T4: Examine galaxy morphology transformation	117
4.5	Application Design	118
4.5.1	Temporal Playback	119
4.5.2	Galaxy Query	120
4.5.3	Halo Close-up	120

4.5.4	Side-by-side Comparisons	122
4.6	Implementation	123
4.6.1	Data Processing	123
4.6.2	Visualization Interface	124
4.6.3	Deployment and Hosting	124
4.7	Science Use Cases	125
4.7.1	Use Case 1: Identify the time of structure emergence . .	125
4.7.2	Use Case 2: Compare low redshift halo with progenitors	129
4.7.3	Use Case 3: Compare halo conditions to its local envi- ronment	131
4.8	Conclusion	132
5	Conclusion	135
6	Appendix	139
6.1	Glossary	139
	Bibliography	143

List of Figures

2.1	A screen capture of the <i>IGM-Vis</i> visualization application, which facilitates a range of analysis tasks using quasar sightline data in order to better understand intergalactic medium and circumgalactic medium absorption features. The Universe Panel (upper left) shows a 3D map of galaxies in the Coma Supercluster, along with “skewers” representing absorption signals in spectra of background quasar sightlines. The Galaxy Panel (lower left) provides descriptive metrics for selected galaxies. The Spectrum Panel (right) displays the spectra and marks nearby galaxies, facilitating comparative analysis between galaxies and absorption, and provides interactive controls with to select which regions of the 3D map are visible, to choose the zoom level of the spectra, and to update the profile plot to its left in the Equivalent Width Plot Panel.	30
2.2	Zooming into a specific region of the Coma Supercluster dataset in the Universe Panel. To reduce visual clutter, a user can toggle on or off different elements or filter the number of galaxies displayed. The top image displays an overview of a large number of galaxies; the middle panel zooms into a region of interest, with skewers and labels toggled on; the bottom image filters out galaxies beyond a user-specified distance threshold from the skewers.	37
2.3	The same “skewer,” sightline SBS1116+523, represented in the Universe Panel (top) and in the Spectrum Panel (bottom). On the top, the brighter coloring along the skewer indicates absorption, which a user can explore in more detail in the associated spectral plot.	42

2.4	A screen capture of the <i>IGM-Vis</i> interface while investigating Use Case 1. The Universe Panel shows 4 skewers with absorption (brighter shading) at similar redshifts, and the Spectrum Panel shows each sightline’s spectrum with the absorption features of interest labeled with yellow arrows. The slider below these spectra (not shown here) has been set to mark the redshifts of galaxies within ~ 500 kpc of the skewers (white vertical lines), only the SDSSJ111443.70+525834.0 sightline has a galaxy present near the redshift of interest (green vertical line).	47
2.5	The far left panel shows an example of how an equivalent width measurement is captured: the user selects (1) the left boundary, (2) the right edge, and (3) a center reference point. The two equivalent width profiles shown in the middle and right panels correspond to Use Case 2. The middle plot shows CGM measurements within filaments near the Coma Cluster, resulting in mostly nondetections, whereas the rightmost plot shows measurements taken in regions > 50 Mpc away from the Coma Cluster, resulting in a greater incidence of detections.	50
3.1	Multiple volume renderings within the <i>CosmoVis</i> software illustrating the various types of gas attributes that can be retrieved from the cosmological simulations. Here, a user interactively places a “virtual skewer” within the <i>EAGLE</i> 12 Mpc simulation, as shown in the leftmost panel (temperature). These skewers can be used to generate absorption line spectra, ion column densities, and other properties. The remaining three panels display gas density, temperature, and metallicity fields within the cross-section shown in the first panel, with the same skewer running through each panel. Plots displaying thermodynamical properties of the gas intercepted by the skewer are shown in the lower-right corner of the second, third, and fourth panels. . . .	57
3.2	Gas temperature volume renderings in simulations of various sizes. Snapshots from left to right: 12 Mpc <i>EAGLE</i> $z=0.0$; 25 Mpc <i>EAGLE</i> $z=0.0$; 100 Mpc <i>EAGLE</i> $z=0.0$; 100 Mpc <i>TNG100-1</i> $z=2.3$. While smaller volumes (left) can be easier to work with, larger simulations (right) provide a much greater sample size of various galaxy morphologies and environmental conditions.	59

3.3	Flowchart comparison between a traditional pipeline (left) versus the simplified <i>CosmoVis</i> pipeline (right) when working with these cosmological simulation datasets. In the traditional pipeline, in order to begin comparing a simulation with observations, a researcher would have to acquire the data, write or install code to read the data, identify interesting regions via database querying, extract physical quantities and perform statistical analysis, each independently. <i>CosmoVis</i> consolidates all of these separate components into one unified interface. Note that the Python packages <i>yt</i> [282] (used for data loading and processing) and <i>Trident</i> [147] (used for computing column densities and synthetic absorption spectrographs that are aligned with the virtual skewers; also requires <i>yt</i>) are useful in both the traditional pipeline as well as for enabling interactive scientific analysis in <i>CosmoVis</i> . For more information on how <i>CosmoVis</i> utilizes these packages, refer to Related Work (section 3.4), Analysis Task 2 (subsection 3.5.2), Data Preprocessing (subsection 3.6.1), Client-Server Architecture (subsection 3.6.5), and Figure 3.7.	65
3.4	This figure shows a series of interactive volume renders of gas fields created with the <i>CosmoVis</i> visualization tool. The data is extracted from the TNG100-1 $z=2.32$ cosmological simulation snapshot. The top row depicts the entire 100 Mpc volume from an angle, and the bottom row shows a head-on perspective of a 15 Mpc thick slice of the volume.	66
3.5	Zoom in on a region containing two distinct circumgalactic environments in the 12 Mpc <i>EAGLE</i> simulation, with color representing temperature. Star particles are represented here as faint yellow disks at the center of these two adjacent systems. There appear to be visible galactic outflows of hot gas above and below the nuclei, indicated by the yellow-orange coloration.	67
3.6	<i>Temperature</i> profile along a virtual skewer. Using the default color map, blue indicates low temperatures (10^4 K) and red indicates high temperatures ($> 10^6$ K). These temperature values along the skewer are also displayed in the line plot in the <i>Observe</i> panel on the right, where the user can also adjust the skewer position and define a custom color map. An additional skewer width option is available to adjust the radius of the cylindrical object.	75

3.7	Architecture of the <i>CosmoVis</i> software. The <i>CosmoVis</i> client interface is accessible via modern web browsers, where the visualization is powered by <i>D3.js</i> for loading and displaying 2D graphs and <i>Three.js</i> for 3D volume rendering. The interface allows for interactive exploration of the simulation dataset and on-demand analysis via the placement of the virtual skewers. Requests are sent via <i>Socket.IO</i> between the client front-end and the server back-end, which can be hosted on an EC2 instance on Amazon Web Services or on the Nautilus distributed computing platform. Metadata describing simulation parameters, skewer placement, and the desired data product are sent to the server. The back-end runs the <i>Flask</i> webserver and hosts a series of custom Python scripts. When the client requests a simulation snapshot (e.g., <i>EAGLE</i> or <i>TNG</i>), it is retrieved using the <i>yt</i> package and rendered using a series of custom fragment shaders. When requested by the user (through the placement of skewers), ion column densities, physical properties, and synthetic spectra are generated by <i>Trident</i> . These data is then sent back to the client browser, where plots are dynamically created using <i>D3.js</i> and rendered atop the virtual skewers using the <i>Three.js</i> library.	76
3.8	Four user interface panels can be accessed on demand to load in and interact with simulation datasets, including retrieving spectra and CGM/IGM gas physical conditions: Data Selection, Layers, Observe, and Spectra.	81
3.9	Example of a cosmic sheet found in the 100 Mpc <i>EAGLE</i> simulation shown from three different angles rotated around the same axis, arranged left to right. Different gas fields are presented in each row, from top to bottom: density, entropy, metallicity, and temperature. See Use Case 1 in Sec. 3.7.1 for more information about these cosmic structures.	84
3.10	The CGM environment analyzed in Use Case 2 (Sec. 3.7.2). Here we have placed ten skewers through an apparent biconical galactic superwind to analyze its thermodynamic structure. . . .	87
3.11	The large-scale metallicity distribution in the <i>EAGLE</i> 25 Mpc volume. This is a ~ 5 Mpc slice of the cosmic web, revealing that much of the intergalactic medium in cosmic web filaments is highly enriched with heavy elements (low to high metallicity colored blue to red). Also seen in here are a number of regions in the immediate vicinities of galaxies likely evacuated of their metal-rich gas due to galactic winds [206].	91

- 4.1 Fields are shown from the 12 Mpc EAGLE simulation from top to bottom: density, metallicity, entropy, and temperature. From left to right: snapshots arranged from high redshift ($z = 20$) to low redshift ($z = 0$). Four enlarged views ($z = 0, 1.487, 3.984, \text{ and } 8.075$) are shown above each film strip, with their thumbnail highlighted in the sequence. Each visualization uses the same blue (low value) to yellow (middle value) to red (high value) color scheme. However, the numbers the transfer function aligns to depend on the specific dataset and type of field. All thumbnails in the film strip have synchronized color scales. . . . 105
- 4.2 The galaxy query tool allows users to filter, highlight and zoom in upon galaxies and halos in the simulation. Here, the metallicity field is shown in the main visualization and a portion of the film strip along the bottom. Hovering over a galaxy ID in the list in the open galaxy panel places a yellow sphere in the volume, indicating its size and position. Clicking on the galaxy ID slices the volume centered upon that galaxy. Clicking on a halo ID loads a higher fidelity zoomed-in volume, as shown in the next figure, Figure 4.3. 121
- 4.3 Three fields are shown from a galaxy halo in the 12 Mpc EAGLE simulation from top to bottom: metallicity, density, and temperature. For each field, three zoom-in views of the galaxy are shown from different points in time ($z = 3.017, 1.487, 0.000$). Below the large views are two rows of film strips. On top is the zoom-in, and below is the full volume with the position and size of the halo indicated by a green sphere. The blown-up views are highlighted in the zoom-in film strip. The next figure, Figure 4.4, shows a skewer piercing through a cross-section of this halo. 126
- 4.4 Temperature is shown for the same galaxy halo cross-section across three-time points ($z = 0.000, 1.487, 3.017$) arranged top to bottom. The cross-section cuts through the core of the galaxy to highlight interior features. In each frame, a skewer placed horizontally pierces through the center of the view. Skewers contain information about physical parameters (i.e., temperature, density, metallicity, and entropy) and column densities of element and metal ions in the medium. While not shown here, metals were detected using the skewers in the $z = 0.000$ snapshot but not in the earlier snapshots ($z = 1.487$ and 3.017). In Figure 4.4, 127

4.5 Metallicity is shown for the galaxy halo cross-section at $z = 3.017$. The cross-section in the top frame cuts through the galaxy halo's core to highlight interior features, continuing the investigation from Figure 4.4. The purple color along the skewer indicates non-detection, whereas white highlights indicate the presence of that feature. While no metallicity was detected in the zoomed-in region around the halo, we can zoom back out to reveal the neighborhood around the halo, as shown on the bottom. A skewer has been placed perpendicular to the one in the upper frame, and a few fields (metallicity, Carbon I, Magnesium I, Nitrogen II, and Silicon VII) are shown on the skewer. This skewer confirms a deficit of metals within the galaxy halo but shows metals in the IGM outside the halo, as indicated by the white lines. 128

List of Tables

1.1	Ancient astronomy visualization and record-keeping artifacts (< 1600s).	6
1.2	Timeline of major astronomy discoveries after the invention of the telescope [99].	12
1.3	Timeline of major astronomy discoveries after the invention of the telescope [99].	12
1.4	20th-century astronomy visualization examples.	19
2.1	A taxonomy of tasks relevant to the IGM/CGM community. . .	36
2.2	Performance metrics for <i>small</i> , <i>medium</i> , and <i>large</i> IGM/CGM datasets. The <i>small</i> dataset includes galaxies and sightlines in the Coma Supercluster, while the <i>medium</i> and <i>large</i> datasets include additional galaxies surrounding the Coma Supercluster. In addition to the number of galaxies, skewers, and the redshift range, we include metrics from Google’s PageSpeed Insights (https://developers.google.com/speed/), including the overall Speed Score, the time-to-interactive (TTI) metric, and the input latency (IL) for both desktop and mobile application simulations, along with the average frames per second (FPS) for a) a 15-inch 2012 model MacBook Pro laptop with a 2.6 GHz Intel Core i7-3720qm CPU and an NVIDIA GeForce GT 650M GPU running macOS 10.14, and b) a custom-built PC with a 4.00 GHz Intel Core i7-8086k CPU and an NVIDIA Titan XP Pascal 12 GB GPU running Windows 10.	48
2.3	Title and research areas of the experts who provided feedback, along with the datasets and software tools they commonly use.	51
3.1	Example particle types and fields made available in cosmological simulation snapshots that can be retrieved and plotted using <i>CosmoVis</i> . Simulation datasets typically organize their data in terms of different particle types (gas, dark matter, stars, and black holes), each expressing a variety of physical quantity fields.	96

3.2	Virtual skewer data products available in <i>CosmoVis</i> . <i>Physical properties</i> and <i>ion column densities</i> linearly align along skewers, and can be dynamically plotted as banding patterns directly onto the skewer, as well as a plot in the <i>Observe</i> panel. <i>Synthetic spectra</i> correspond to observational absorption line spectroscopy measurements taken with specialized telescope sensors.	100
3.3	Comparison of first contentful paint (s), time to interactive (s), and framerate (fps) across three different machines and two server configurations as measured using Google Lighthouse. The local server is running on a Windows 10 PC with an Intel Core i7-8086K CPU @ 4.00GHz, 32 GB Memory, and a 12 GB NVIDIA Titan Xp GPU, and was tested on the same network using a MacBook Pro (13-inch, Apple M1 chip, 2020, 16 GB Memory) and a MacBook Air (13-inch HD+, Intel 1.8 GHz Dual-Core Intel Core i5, 8 GB Memory, Intel HD Graphics 6000 1536 MB). The same machines were used to test the performance of <i>CosmoVis</i> deployed in an Ubuntu Docker container on the Nautilus distributed compute cluster. Running <i>CosmoVis</i> locally results in marginally faster initial load times than the Nautilus deployment. All machines support realtime framerates of 60fps for the smallest volume resolution (64^3).	101
3.4	Performance comparison of loading times and framerates for larger volume resolutions (128^3 , 256^3 and 512^3) across three different machines accessing two different server configurations. The Windows machine containing the dedicated GPU is able to run all default resolutions at interactive rates or better, and even the older consumer laptop with integrated graphics is able to support higher resolutions up to 256^3	102
4.1	A chronological list of cosmological epochs regarding the universe's age, redshift, and what galaxies instrumentation can detect.	106
4.2	A list of several physical properties that can be directly observed, estimated from observations, and available in cosmological simulation snapshots. Only a few measurements are directly observable in our universe (and limited to emission sources), whereas simulations contain multitudes of fields directly accessible and ready to visualize.	108

Abstract

Interactive Analysis Tools for Visualizing the Universe

by

David Abramov

This thesis introduces three novel interactive astronomy data visualization tools for exploring and analyzing information from telescope observations and cosmological simulations. The amount of data produced by modern telescopes and simulations is tremendous and can be challenging for experts and non-experts to explore comprehensively without the help of accessible, interactive techniques and tools. Given the size and complexity of these data, it often requires advanced technical expertise to access, query, view, and analyze the information. Furthermore, working with large-scale simulations requires powerful computing resources that may not be accessible on basic consumer hardware. The projects described in this thesis leverage advances in consumer graphics hardware and browser-based rendering to provide lightweight, accessible tools to the astronomy community that are designed to simplify analytic workflows, reduce cognitive load and the need for users to code, and contextualize the data meaningfully. In chapter 1, the introduction, a brief history of astronomy visualization methods from the stone age to now, is preceded by a description of the three interactive browser-based visualization tools introduced in this thesis: *IGM-Vis* (chapter 2) contextualizes the impact of galaxies in observational absorption spectra through the introduction of "skewer sightlines"; *CosmoVis* (chapter 3) is a cosmological simulation volume and particle visualizer that allows users to place "skewer sightlines" in the volume and return absorption spectra and other physical properties; and finally, a temporal animation extension to *CosmoVis* (chapter 4) that allows for exploring the evolution of cosmological simulations. For each project, the scientific motivations and background are detailed for a visualization audience and contextualized with related software tools. We identify astronomy-based analysis tasks that each tool solves, along with details of the design and implementation of the systems as they relate to enabling the tasks. Finally, a set of scientific use cases and user and performance evaluations are provided.

Acknowledgments

I want to acknowledge the following coauthors for their contributions to two previously published works as well as the final chapter, *I GM-Vis: Analyzing Intergalactic and Circumgalactic Medium Absorption Using Quasar Sightlines in a Cosmic Web Context* and *CosmoVis: An Interactive Visual Analysis Tool for Exploring Hydrodynamic Cosmological Simulations*, that are included in this dissertation: Joseph N. Burchett, Oskar Elek, Cameron Hummels, J. Xavier Prochaska, Angus G. Forbes, Jasmine Otto, Cassia Artanegara, and Farhanul Hasan. I want to thank the members of my advancement dissertation committees for their thoughtful comments and feedback: Angus Forbes, Michael Mateas, J. Xavier Prochaska, and Joseph N. Burchett. I am grateful for the teaching opportunities and experiences with Noah Wardrip-Fruin, Eric Lengyel, Brian Hansen, Nathan Altice, Angus Forbes, Ishaan Paranjape, and Asiih Song. I appreciate the guidance and assistance from my academic advisors, Emily Jackson, Crystal Navarro, and Theo-Alyce Gordon. I also want to acknowledge other colleagues and mentors I have had the pleasure of working with for visualizations, teaching, and research over the past few years: Jeffrey D. Weekley, Scott Davidoff, Hillary Mushkin, Santiago V. Lombeyda, Maggie Hendrie, Jason Rabinovitch, Pierre Boutillier, Walter Fontana, Marianne Weems, George Legrady, Sarah Frost, Manu Mathew Thomas, Mahika Dubey, Montana Fowler, Hongwei Zhou, Kyle Gonzalez, Anya Osborne, Aspen Hopkins, Sandra Bae, Juan Camilo González, Julián García, Antonia Bustamante, Hugo Idárraga and Andres Burbano, Mauricio Alberto Cortes Cely, and Catalina Gonzalez Uribe. I would also like to thank other members of the computational media community at UCSC, my parents, my brother, my partner, and my friends and family who have supported me on this journey.

Chapter 1

Introduction

1.1 Projects Overview

The projects introduced in this thesis were designed and developed in collaboration with astrophysicists at UCSC, NMSU, and Caltech. who work with observational data (i.e., analytical measurements captured by specialized telescope sensors) and cosmological simulations (large, complex physical models that mathematically describe the universe’s evolution based on observations). Working with each data type (observations and simulations) presents its own strengths and challenges. Observational results are the bread and butter that inform us about the structure of our universe and our place within it. For thousands of years, astronomers have found ways to keep track of the movement of celestial bodies, such as the sun, moon, and constellations – usually in the form of tables, but also through mechanical and visual methods. Observational measurements are carried out in *sky surveys*, where data is gathered across the sky to provide a systematic and comprehensive mapping and catalog of celestial objects, such as stars, galaxies, and other cosmic structures. These catalogs are stored in tabular databases containing measurements (i.e., absorption spectra, optical images, celestial coordinates), and it is up to other tools to aggregate and present this data meaningfully for experts and a general audience.

The first project introduced in this thesis in chapter 2, *IGM-Vis* [77], is a browser-based interactive software designed to address challenges in visualizing galaxies in relation to absorption line spectra when studying the intergalac-

tic medium (IGM) and circumgalactic medium (CGM). This tool visualizes a set of galaxies from the Sloan Digital Sky Survey (SDSS) as red and blue spheres (representing active and quiescent star formation rates) contextualized with absorption spectra projected along cylindrical "skewer sightlines." These skewers represent the measured absorption spectrum profile, usually depicted as a 2D plot. By placing the skewers in the same volume as the galaxies, the relationship between the measurements and intervening material becomes more apparent. One of the primary limitations in absorption line spectroscopic measurements is having a bright enough background light source to illuminate foreground objects. Usually, this is done by pointing the telescope at quasi-stellar objects (QSOs), or quasars, which are distant galaxies with a supermassive black hole ejecting photons in our direction, thus appearing very bright. The detection of absorption lines in the spectra of distant quasars, which began more than 50 years ago, has been crucial in exploring the formation and evolution of cosmic structures. By studying the lines-of-sight that pass through the immense and intervening expanses of the Universe, astronomers gain invaluable insights into the composition and history of gas complexes within the IGM. These gas complexes are found within the relatively dense regions of the cosmic void, where galaxies, galaxy groups, and galaxy clusters are also located [46, 47, 246]. Any intervening material, such as metals, will absorb and scatter the light at specific wavelengths, which can be detected in the absorption line spectra. *IGM-Vis* decreases the cognitive load of comparing these spectra and the neighborhoods of galaxies which they pierce. Given the rarity of QSOs and the lack of emittance by the cool gas in the universe, it can be difficult to construct a complete picture of the large-scale structure of the Cosmic Web.

This is where the role of cosmological simulations comes into play. In the past two decades, hydrodynamic simulations of galaxy formation have become an indispensable tool in modern astrophysics. These simulations have provided astronomers with unparalleled insights into the structure, chemical composition, and processes driving the formation and evolution of galaxies in our Universe. While these complex simulations are carefully adjusted to match observable characteristics of real galaxies, they also generate additional

theoretical predictions based on the initial conditions. Thus, simulations play a crucial role in interpreting observations of stars within galaxies and their surrounding gaseous environments. In a single snapshot of a simulation (i.e., an n-dimensional data structure representing the physical structure and composition of a cosmological volume at single point in time, as referred to as a 'box'), there may be hundreds or thousands of galaxy halos embedded in a one hundred megaparsec box, along with dozens of physical parameters associated with gas, dark matter, star, and black hole particles, such as temperature, metallicity, pressure, elemental composition, energy, entropy, star formation, and much more. Cosmological simulation outputs contain much higher resolution information per particle or voxel than is feasibly achievable with observational techniques, including physical conditions of the gas, which we must infer from limited information in the observational domain. For starters, galaxies in simulations can be viewed from any angle, and any field can be visualized as a volume, illuminating an otherwise dark sky. Still, there is a strong relationship between the underlying physical models and starting conditions for cosmological simulations with empirical observations. Measuring the accuracy of a simulation is a matter of comparing the synthetic results back to observations of the universe. If a statistical sampling of galaxies in the simulation strongly matches observations, it may be a reliable model.

Measuring the accuracy of a cosmological simulation involves comparing its synthetic results with real observations of the universe. Uncertainties can arise from various factors, such as: the initial conditions of the simulation can introduce uncertainties when different values and assumptions are used; the simplified assumptions of subgrid models where "particles" actually summarize a volume of space; uncertainties in the exact role of dark energy and dark matter; and the complexity of measuring and modeling hydrodynamic and thermodynamic physics such as gas cooling, star formation, and feedback. By comparing the statistical properties of galaxies in the simulation to real observations, researchers can assess the reliability of the model. Based on the simulation outputs and analysis, the model or starting conditions can be tweaked and rerun, ideally returning better results.

The strength in the amount of data contained inside cosmological simula-

tions is also a significant challenge working with them. This is the inspiration for the second project described in this thesis, *CosmoVis* [35]. This is an open-source web-based visualization tool for the interactive analysis of massive hydrodynamic cosmological simulation data. It was also designed in close collaboration with astrophysicists to enable researchers and citizen scientists to share and explore these datasets and use them to investigate a range of scientific questions. *CosmoVis* visualizes many key gas, dark matter, and stellar attributes extracted from the source simulations, which typically consist of complex data structures multiple terabytes in size, often requiring extensive data wrangling. In this tool, the "skewer sightline" graphical metaphor from the observational data tool *IGM-Vis* is extended into the realm of simulations as "synthetic skewers," whereby the same data product is produced by the software as would be collected from a sensor onboard the Hubble Space Telescope. These "synthetic skewers" are simulated analogs of absorption line spectroscopy that act as spectral probes piercing the volume of gaseous cosmic medium, which users can place anywhere within the volume. These synthetic spectra can be used to gain insight into the source datasets and to make functional comparisons with observational data.

When originally creating *CosmoVis*, we were focused on being able to investigate a single cosmological simulation snapshot at a single redshift, or in other words, a the state virtual universe inside of the computer at a single point in time. The third project discussed in this thesis is new work that enables temporal analysis in *CosmoVis*. These simulations generally contain dozens or more sequential snapshots in time at various *redshifts*, a cosmological measurement that can be used to describe the universe's age. In observational astronomy, the redshift, or z , that is measured is a velocity due to either the expansion of the universe or motion of the object under the influence of gravity. Within the context of the simulations, redshift is simply a measure of how much time has passed. With observations of galaxies in our universe, we can only look at galaxies from a single angle (our planet) and at a single time point determined by the expansion of the universe and the amount of time it took the light to travel here. With simulations on the other hand, we have access to the entire history of the same galaxy or filament or other feature evolve over

multiple redshifts. While the galaxies and their histories within these simulations are very detailed and information rich, they are far from perfect. The temporal extension to *CosmoVis* introduces animation, galaxy querying, halo zoom-ins, merger tree tracking of the largest progenitor halos, and a small-multiples film strip with linked views for quick navigation. Additionally, the temporal features work with the synthetic skewers introduced earlier, enabling more scientific use cases and analysis tasks.

All together, these tools fit into the broader research area of astronomy visualization, or *astrovis* for short. A brief history of *astrovis*, a rich field that has its start well-before digital computers, is discussed in the next section. Connecting all of the techniques, machinations, and visualization methods discussed in the next section is figuring out our place in the universe by carefully recording observations, and sharing that knowledge with others.

1.2 Brief history of *astrovis* through the ages

Humans have looked toward the sky out of wonder and curiosity throughout the ages, tracking constellations to mark the passing of time and informing people about the changing of seasons and when to plant their crops [189]. People told countless stories through the stars, with many civilizations naming the constellations and keeping track of their movements. The earliest astronomers had to rely on their visual acuity when observing the cosmos. Table 2 below shows a handful of visualization and record-keeping technologies that ancient astronomers used before the widespread utilization of the telescope.

1.2.1 Stone age *astrovis*

Lunar phases carved into eagle bones are the earliest evidence of astronomy visualization, dating to c. 30,000 BCE in France. Other bone carvings are historically significant in ancient China (1600 - 900 BCE) [162]. Paintings (c. 15,000 BCE) in Grotte de Lascaux, the famous Lascaux cave in France, depicts stars and constellations. Stone structures are another type of ancient astronomy visualization method. Wurdi Youang in Victoria, Australia, is one out of hundreds of stone arrangements constructed by aboriginal peoples in the

Visualization Method	Period	Notes
Bone Carving	30,000 BCE	Eagle bone carved with phases of the moon (France). Bones inscribed with astronomy phenomena have also been found in China dating to c. 1600 - 900 BCE [162].
Stone Structures	25,000 BCE	3D monuments
Lascaux Cave Painting, France	15,000 BCE	2D illustration, stars, and constellations
Clay Tablets	2193 BCE	2D illustrations/recording
Armillary Spheres	195 BCE	3D sphere
Astrolabe	150 BCE	2D, Startracker, Mechanical Calculator, Timekeeper, Measuring tool
Antikythera Mechanism	100 BCE	Mechanical computer, Modeling, Education
Ephemeris	100 CE	Tables/catalogs, Data Wrangling
Line graph of planetary motion	900 CE	2D graph
Tycho's Great Globe	1580 CE	3D + Mechanical calculator

Table 1.1: Ancient astronomy visualization and record-keeping artifacts (< 1600s).

region, who occupied the area from c. 25,000 BCE to 1835 CE until colonizers arrived. This structure appears as an egg-shaped arrangement of stones placed about the perimeter when viewed from above. Historians hypothesize to keep track of the sun's position at the solstice and equinox [204]. Another stone structure in Scotland is speculated to be an early lunar calendar [207]. Other notable stone monuments with astronomy significance include Stonehenge and the Intihuatana stone of Machu Picchu [179]. The Intihuatana stone, created by the Incas, is aligned with the sun at the solstice. It is speculated that there were other Intihuatana stones in other Incan cities and that they played an important cultural role. Still, this information was lost after the Spanish colonization of the region [179]. The origins of Stonehenge are even more of a mystery, and whether or not they were created as an early astronomy observatory is a contentious topic in archaeoastronomy [237]. The Chaco Canyon Culture National Park in New Mexico is home to what has been named the "Sun Dagger" petroglyph, whereby the sun peeks through a stone slab formation in a dagger shaped beam onto a spiral drawn upon the stone. Based on the time of year, the solstices and equinoxes can be tracked by following the dagger [175].

Tablets, tables, and other devices

Clay tablet artifacts dated to Sumer in 2193 BCE are one of the earliest visualizations of a cataclysmic cosmic event. These tablets illustrate the trajectory of a large meteor relative to the position of constellations, which presumably decimated the region and led to the collapse of the Sumerian civilization [250]. Ancient astronomers invented methods and tools that have carried on for generations. For example, ephemerides (meaning "diaries" in Greek) are astronomical tables constructed from observations that can predict the trajectories of various celestial objects, such as stars, planets, or comets. These astronomical "diaries" date back to ancient Babylon [191]. While the ancient Babylonians inscribed their observations onto clay tablets, today, astronomy data is recorded and distributed in the digital realm in various file formats. Hipparchus (190-120 BCE) is known for creating the first accurate star chart containing nearly 1000 stars and their location to $1/2^\circ$ of precision, along with

their luminosity along a 6-point scale, similar to the one used by modern astronomers [99]. Claudius Ptolemy (90-168 CE) later references the stars in his famous book *Almagest*, an ephemeris containing tables of astronomical trajectories [2]. Ptolemy's system was incorrect in assuming a geocentric cosmology with all celestial objects moving in perfectly circular orbits. Despite this, the *Almagest* is the first known model that maps star trajectories and thus became the quintessential astronomy book for another thousand years [10]. The legacy of these ancient ephemerides is carried on in modern astronomy in the form of various star and galaxy catalogs recorded from observational sky surveys, albeit now we have much higher precision.

The relatively fine precision of the star position in these early catalogs can likely be attributed to the astrolabe, a disk-shaped star-tracking instrument (from Greek, meaning star-taker). These devices date back to Greece in the 2nd century BCE and were widely used through the 17th century [202]. Astrolabes were used to measure the current time based on the sun's or other stars' altitude [145], for triangulation, predicting the position of celestial objects and constellations at different times of the year, and for religious purposes [93]. The Islamic caliph Al-Manṣūr (712-775 CE) took a special interest in the astrolabe as part of a measure to gain legitimacy as a leader by bringing Greek science, especially astronomy and astrology, into the Arab world. The astrolabe was useful for Islamic people to determine the direction of Mecca [138]. In addition to celestial purposes, the astrolabe was used for surveying purposes, such as measuring the height, depth, and distances of different structures such as a tower or a well [138]. Curiously, some astrolabes have two measurements of time: the equal division into 24 equal hours similar to what we are accustomed to, and a second system that divides hours unequally based upon the season. During the summer in this alternative system, hours during the day are longer compared to hours at night, whereas the reverse is true in the winter months [138]. In addition to astrolabes and celestial diaries, there are a few other astronomy visualization tools used by the ancient Greeks. Celestial globes are much like a standard globes. However, their surface is illustrated with star positions rather than geographic land and water features. While no ancient celestial globes have survived, written evidence indicates that the

Greeks knew how to construct them [103].

A similar device is an armillary sphere, a mechanical device invented to track the sky and can still be useful in educational settings. Historians attribute the invention of the first complete armillary spheres to the Greek scholar Eratosthenes of Cyrene (276-195 BCE), who is additionally known for directing the Library of Alexandria [6]. These three-dimensional spherical mechanisms place the earth at the center, concentrically encompassed by several movable rings. These rings are a visual tool for taking astronomical measurements about multiple spherical reference coordinates: altazimuth, equatorial coordinates, and ecliptic [9]. Some differences in the celestial coordinate systems across cultures were used in the design of armillary spheres [4]. Advanced armillary spheres contained multiple rings that could be used to convert between these reference systems, which surely aided in avoiding manual conversion when dealing with different star catalogs [223]. Armillary spheres also date back to China as early as the 2nd century CE. Chang Heng created a spherical astronomy apparatus (125 CE) constructed with a horizon and meridian rings [4]. In India, Peng Lu (2015) visualized instructions dating back to 629 CE for creating an armillary sphere (called *gola* in Sanskrit) written by Bhāskara I [174]. By the 8th century CE, armillary spheres were also being constructed by Islamic astronomers [159]. These spheres also continue to capture modern minds in VR [313].

An interesting bronze geared contraption also dates back to ancient Greece (c. 150-100 BCE): the Antikythera Mechanism. The device's namesake is from the location of the merchant shipwreck (80-60 BCE) that was found and salvaged (1900-1901 CE) along with other valuable objects of the time, such as statues and glassware [153]. It is an example of a very early mechanical computer or calculator [108]. The fragmented state in which it was found limits our understanding of its full functionality. Using X-ray tomography techniques, researchers believe the mechanism could be used to predict lunar and solar eclipses and planetary positions, at the very least [116]. Some historians suggest that the device could have also been used for educational purposes, as a tangible object for explaining the current state of astronomy knowledge at the time [108].

Timekeeping and star tracking

Early astronomical clocks were primarily driven by water. These clocks have a rich history in ancient China, but some notable ancient Greek examples also exist. Sometime between 300–200 BCE, the Greek physicist Ctesibius created a water-driven time-keeping cylinder. Around 140 BCE, Hipparchus reportedly made ‘the anaphoric clock’, a water power astrolabe [148]. Early astronomical clocks in China date back to the Later Han dynasty (25–220 CE) when Zhang Heng (78–139 CE) made a water-driven celestial globe. Water-driven technology would be used in China for many centuries to power a variety of astronomical clocks. A water-driven bronze armillary sphere was built by Yi Xing (683–727 CE) and Liang Lingzan [264]. In the 10th century, Zhang Sixun constructed a four-meter tall water-powered astronomical instrument that could be used for tracking time and the trajectories of celestial objects [264].

One of this era’s most notable and well-documented water-powered astronomical clocks is The Water-Powered ‘Cosmic Engine,’ designed by Su Song (1020–1101) and Han Gonglian in 1086. This multi-story tall contraption used water to concurrently drive the motions of an armillary sphere, celestial globe, and time-keeping mechanism. In India, China, and the Middle East, people continued to create and advance upon water-driven clocks, such as Ismail al-Jazari’s (1136 - 1206) water-driven Peacock Clock [148] [38] and elaborate astronomical Castle Clock [88]. Mechanical time-keeping was becoming more common, with non-water-powered clock-making emerging by the 12th century. Giovanni Dondi’s Astrarium (1348) was one of the first and most detailed astronomical clocks to emerge [52]. While the original was lost to time, Dondi provided enough detail in his descriptions and illustrations to reconstruct the device roughly 600 years later [52].

Accurate time-keeping would lead to more accurate celestial measurements and help disprove geocentric models of the universe and solar system. One of the earliest quantitative graphs in recorded history is related to astronomy. This graph plots different planets’ inclination along the x-axis. It dates back to the 10th century CE, found in the appendix to commentaries by A. T. Macrobius on Cicero’s *In Somnium Scipionus*. This visualization is remarkable

because it greatly predates other graphing and graph paper examples by 500-600 years [117] [122].

The Danish astronomer Tycho Brahe (1546-1601) very nearly lived to see the dawn of the telescope, as discussed in the next section. Despite this, his observations and instruments remained influential for the scientific revolution. Beginning in 1580, Tycho operated an astronomy observatory called Uraniborg for over twenty years in Denmark. Rather than a telescope, Tycho used exceptionally large quadrants, a curved astronomical instrument with gradations that could accurately measure different angles, which was pointed out of a hole in the wall. The large size (6 ft radius) of the Mural Quadrant that Tycho built diminished resulting observational error [13] [298] [270]. While Tycho did not invent the quadrant, he greatly improved upon its design. This device required multiple people to work in coordination. One person would be operating the quadrant, one person would be watching the time, and another person would be recording the position and time as reported by the other two operators [5].

In addition to the Mural Quadrant, Tycho also constructed the ‘great globe’ that came into commission in late 1580, which had two purposes: to record the positions of the stars as they were observed and compute a celestial coordinate conversion. The great globe was 1.6 meters in radius and was used to record over 1000 stars accurately. These devices and more were illustrated in his book “*Astronomiæ instauratæ mechanica*” (1598) [70]. After succumbing to what was most likely a bladder infection (determined after exhuming the body), his research was carried on by his assistant Johannes Kepler, who defined their three planetary laws of motion [89].

1.2.2 Telescopes: onward and upward

Table 1.3 outlines astronomy discoveries during the early telescope era, starting with Galileo and finishing with the start of photographic sky catalogs in the late 19th century. During this period, the main forms of visualization methods employed by astronomers in this era were sketching, printmaking, mechanical instruments, and later photography.

Table 1.2: Timeline of major astronomy discoveries after the invention of the telescope [99].

Persons, artifacts, and events	Period
Kepler, Galileo, and Newton laid down the foundations for the modern era of astronomical sciences	1500-1700
Halley discovers proper stellar motions (Aldebaran, Arturo, and Sirio)	1718
Bradley discovers the effect of stellar aberration due to terrestrial motion	1725
J. Lalande determines the parallax of the Moon and Mars	1751
Nautical Almanac - ephemeris, annual publication of the main celestial bodies	1767
Herschel discovers more than 800 double stars, Uranus, and “island universes” (galaxies)	1783
Carte du Ciel - Astronomy catalog that measured magnitude up to 11.5 for 20,236 stars	1887-1931
Astrographic Catalog - The first photographic map of the sky with data from 20 observatories across the world. Contains 22,000 photographic plates (4.5×10^6 stars) with a magnitude of up to 14	1895-1950

Table 1.3: Timeline of major astronomy discoveries after the invention of the telescope [99].

Early telescopes

Galileo Galilei ushered in a new era of astronomy in 1608 when he first pointed a telescope to the sky and readily published his findings in *Sidereus nuncius* [7]. Despite the association, Galileo neither invented optics nor the telescope. The history of optics precedes the invention of the telescope by several thousand years and is a rich topic that will not be discussed in detail here [283]. Major advances in optics started several hundred years before the telescope, namely in the convex polishing of non-precious stones to aid poor eyesight. It is difficult to name the true first inventor of the telescope, as there are discrepancies in the historical record and controversy over the reliability of some sources [285]. Regardless of who may have originally invented the telescope, its emergence in seventeenth-century Europe can be attributed to the advancement of the craft of lens-making for eyeglasses. In his writings, Galileo reveals that the first telescope he put together was created using lenses from the shop of a spectacle maker [285]. His first telescope had a magnification of 8X, whereas his later telescope could magnify up to 20X [32].

To the ancients, telescopes were revolutionary for clarifying our place in the universe. With only a slightly higher magnification than the naked eye, these early telescopes were not able to resolve single stars. However, some stars that appear singular to the naked eye are massive collections of stars inside galaxies and nebulae. Since telescopes collect a lot more light than our eyes' pupils are capable of, they can reveal faint objects in the sky. The telescope also enabled getting a closer look at more nearby celestial objects, such as other planets. While ancient astronomers could visually identify and track the trajectories of planets up to the seventh planet (Saturn) without any special optics, early telescopes could resolve moons around other planets in our solar system. Only two years after the invention of the telescope, Galileo was able to identify four moons of Jupiter in January 1610. It took another 200 years (1892) before E.E. Barnard discovered another moon around the planet [184]. Other astronomers in the 17th century (Huygens and Cassini) discovered a handful of moons around Saturn as well [286]. Telescopes ushered in a new era of astronomical observation, with a legacy carried out to this day as larger telescopes are built. These early telescopes were limited in their

resolving capabilities only to view bright and close enough objects. Later, with photographic methods such as time-lapse or long-exposure techniques, more light can be captured with the telescope allowing dimmer objects to become visible. The observable universe at this point was much smaller than what is provided by modern space telescopes such as the Hubble Space Telescope, which can detect even the faintest of galaxies billions of light years away.

Sketching and print media

During this time, sketching became a primary means for sharing telescope observations with others since photo-sensitive media had not been invented yet. Remarkably, during this time, one of the earliest records of a small-multiples type of visualization scheme was applied in an astronomical context by Christopher Scheiner to illustrate the changing of sunspots patterns observed and recorded in 1611 [117]. Small-multiples is a data visualization method that uses an array of many plots arranged on a grid to communicate multivariate information, such as in this case, a time series of sunspot changes [117]. This method is still a popular choice for data visualizations as it can concisely display plenty of information that can be cross-referenced with a quick glance. Tufte elaborates upon other small-multiple astronomy illustrations in his book *Visual Explanations* (1997, p. 106-108) by Galileo and Christiaan Huygens cataloging observations of the rings of Jupiter [277]. Mass reproduction of sketches was possible at this time as the printing press had recently been invented, which likely heavily aided the relatively quick technological advances since the Renaissance. This was before the invention of the photograph, so visual representations were limited to hand-drawn sketches and diagrams. To create copies of these figures, illustrations would be engraved upon copper plates by skilled artisans. The divots and grooves from the engraving could then be filled with ink and effectively stamped, printing the images onto paper [137]. Not all visualizations from this time period were of observations, as some were illustrations and diagrams used to help explain the history and state of astronomy knowledge at the time. These printed illustrations and diagrams likely were easier to manage than physical contraptions such as armillary spheres and globes. Bookmaking allowed knowledge to be

shared more efficiently, especially with the added diagrams.

Orreries

In addition to printed media, people at this time were still crafting new mechanical astronomy visualizations, driven by the advances in clockwork. In the 1700s the orrery was developed. These devices help illustrate the paths of orbiting celestial bodies. For example, a model of the sun would remain static at the center, while the planets and their moons would rotate and orbit the sun based on fixed gear ratios [73]. Orreries are named after the 4th Earl of Orrery, Charles Boyle, despite the first being made by George Graham, and copied by John Rowley in 1713 for the Earl whom it was so graciously named [190]. Orreries were useful to help convince others to discredit the geocentric theory of the universe and adopt a heliocentric view. Therein lies one of the strengths of mechanical visualization methods that are limited in printed media: the visual experience of the dimension of time as a continuous stream rather than a series of discrete points. It may be easier to intuitively grasp a concept by viewing an interactive or animated 3D representation over a static 2D image.

Astrophotography

In the mid-19th century, photography began to play a role in recording astronomy observations. In 1840, J.W. Draper took the first daguerreotype image of the moon [71]. In 1872, American astronomer Henry Draper took the first photograph of the spectrum of a star (Vega) [107]. William Huggins used absorption lines of these spectra to measure the redshifts of stars, which gave astrophysicists the first indication of how fast stars and galaxies are moving through space relative to us [128]. Absorption line spectra remain an essential observational technique for astronomers to interpret the chemical composition of different stars or galaxies and their age, as the wavelength of light from more distant sources stretches out due to the universe's expansion appearing redder. Modern telescopes such as the Hubble Space Telescope can detect galaxies at a maximum redshift $z=11$, when the universe was about 400 million years old [193].

With early telescope observations, it was acceptable for astronomers to

sketch their observations to share what they saw, such that astronomers needed some proficiency in analytical and artistic domains. During the infancy of photography, methods were developed that combined illustration and photographic print to compensate for the limitations of both. By the late 1890s, photo reproduction became more advanced and reliable, driving illustrations in astronomy publications close to obsolescence. Photographic media at the time was seen as a more reliable and trustworthy pictorial representation than an illustration or diagram in this context because it captures the object of interest as it is with standardized mechanical precision [196].

Historically this could be considered as overenthusiasm for the ‘visual truth’ that photography offered, whereas modern data visualization seeks to find a balance between the impulses of displaying just the raw data and an interpretative element. Early photography did not lend itself well to the visual integration of information from multiple sources, such as the small-multiple example from before. Additionally, labeled diagrams might explain certain complex phenomena better than a photograph could on its own. However, individual photographic plates could be comparatively measured and archived, much like a physical version of a digital database. Indeed many discoveries were made possible through photographic techniques. Modern astronomy visualization tools for observational data often integrate data from multiple sources, such as optical images, absorption spectra, and derived metadata, such as virial radius (gravitational extent) or star formation rate.

This issue of visual trustworthiness is still relevant in modern data visualizations, as people can often be skeptical about the reliability or validity of a graphic. People can be suspicious of the validity of visualizations or inherent biases in visual encodings. Visualizations can be easily dismissed or misunderstood if they do not include certain elements, are poorly labeled, or use an inappropriate color scale. In order to make meaningful deductions, it helps to have multiple examples to compare and find statistical insights. With modern data visualizations, it is just as important to take these considerations into account. For example, the visual representation that is chosen can be susceptible to misinterpretation. Some people argue that 2D representations of data are always better than 3D representations, as perspective could make some data

representations misleading. I argue that interactive 3D representations can be more powerful than static 2D data representations in exploring a complex dataset such as a cosmological simulation. In many cases, a 3D data visualization may not be the best approach, but for data with an underlying spatial structure, the depth perception afforded by 3D can help highlight the spatial relationship between different features. Nowadays, scientific data collected by telescopes results in higher dimensional data than, say, simply the magnitude of a light source. Different sensors are calibrated to collect information all along the electromagnetic spectrum, which can't always be simply encoded as a color to be made interpretable. This can lead to multimodal visualizations that might present multiple plots laid out in a dashboard. The choices that go into creating modern astronomy data visualizations follow a lineage of early astronomers making artistic or design decisions while creating their sketches, as there are conventional methods for illustration just as there are for making data visualizations.

Photographic techniques by the end of the 19th century tackled some of the issues with the interpretative element of illustrations and started a new era of observation catalogs composed of images on photographic plates with *Carte de Ciel* and the *Astrographic Catalog* [99]. Astronomy was also making its way into mainstream popularity with the emergence of popular science writers. In France, Camille Flammarion was a science author that provided illustrations in his books such as *Astronomie populaire: description générale du ciel* (1890) [112]. We will see in the next section later there is a shift in data visualization towards mathematical statistics in a time dubbed the “visualization dark ages” in the early 20th century. Although there was an emphasis on statistical plots in the early 1900s, some advances made it easier to share astronomy knowledge with the greater public. Additionally, by the 1990s modern computer data visualizations have moved back into the realm of using more graphical and illustrative examples.

1.2.3 Astronomy visualization in the 20th century

The early 20th century marks the start of the next era of astronomy visualization in this timeline. Table 1.4 includes a list of 20th-century astronomy

visualization examples. In the early 1900s, scientists and astronomers made great strides in advancing our understanding of our physical universe. Observational methods at the time used photographic plates to record absorption line spectra, which provided astronomers with observational evidence supporting the universe’s expansion [102]. During the first decade of the 20th century, Albert Einstein published his theory of special relativity, which introduced the concept of “spacetime”, fundamentally disproving earlier models of the universe as having some kind of ‘aether’ structure [109]. Minkowski’s space-time diagram was a 2D plot that first unified the two concepts of space and time, by which the rate at which time passes depends entirely on an observer’s frame of reference. Minkowski’s diagram is like a typical time series with time on the y-axis and distance on the x-axis, except in this case, an additional diagonal temporal axis is added that represents the speed of light, which is theoretically impossible to pass [293] [123]. This diagram is useful for presenting object trajectories in the spacetime continuum.

Early 20th-century astrophysicists were generating complex physical models of the universe that they were testing and validating with observational data without the aid of digital computers. Throughout this section, we will see a shift from analog and mechanical methods of visualization, such as photographic plates/slides, and planetarium projectors, towards more modern methods of cinema and computer graphics. Immersive experiences, public engagement, and the visualization dark ages Average citizens at this time kept up with astronomical events such as eclipses through newspapers, and were energized by popular science writings at the time, such as Flammarion’s French popular science book, *Astronomie populaire: description générale du ciel* [112].

A photograph titled *Pendant l’éclipse* captured by surrealist artist Eugène Atget (1912) of Parisians watching an eclipse, who were prepared with shaded spectacles to observe the event as the newspapers alerted the public with optimal viewing times that were predicted in advance [85]. Despite the great advances in physics and astronomy in the early 20th century, it was not immediately straightforward for people at the time to create a mental image of these new concepts such as the expansion of the universe or the spacetime continuum. Watson Davis in 1931 wrote, “To journey mentally out into the expanses

Visualization	Time	Notes
Minkowski space-time diagram	1908	2D graph, expresses spacetime continuum with additional time axis for the speed of light
Orbitoscope	1912	First projection planetarium
Atwood Sphere	1913	A metal celestial globe large enough for a small group of people to sit inside and learn about the movement of the stars
Hertzsprung-Russell Diagram	1914	An early scatterplot that combined star data from multiple sources into one figure, which helped astronomers learn about star evolution.
Zeiss Mark I planetarium	1923	A widely popular projection planetarium that accurately depicts the motions of celestial objects in an immersive dome theater.
Cinema / Newsreels	1920	Footage exists from the time of early cinema about astronomy
Astronomical Image Processing System (AIPS)	1980	Software for controlling and processing data from radio telescope arrays
Flexible Image Transport System (FITS)	1981	A astronomy data standard that is still commonly used

Table 1.4: 20th-century astronomy visualization examples.

of the universe it is necessary for those of us who are three-dimensionally minded, owing to our existence here on earth, to abandon hope of too definite visualization of what the universe really looks like from the outside looking in” [102].

Einstein showed the world that space, time, and light behave differently than classical mechanics would lead you to believe. Prior understanding of the cosmos put the Earth at the center, and later the Sun, both of which are easier to conceptualize. With advanced physical concepts, sometimes it is not enough to visualize something in only static two-dimensional representations. Even three-dimensional visualization can be lacking in the fourth dimension, time. Today, special and general relativistic rendering techniques can be programmed into graphics shaders to visualize traveling near the speed of light [295]. Literature in astronomy visualization during the mid-twentieth century is a bit sparse. Perhaps this was a transition period as World War II ravaged the world, shifting global scientific efforts toward ending the war. Others have noted the time between the early 20th century as the “modern dark ages” of visualization, where this darkness was not restricted to just the field of astronomy [117]. It has also been observed that visualizations in general during this period were more focused on quantitative statistics and less on other visual representations [117]. A transition period between analog and digital astronomy started taking place in the mid-20th century. Observatories in the late 19th and early 20th centuries produced troves of photographic plate images. While many very important discoveries were made with these analog data, early photography was not as streamlined and efficient as digital photography is today. Institutions such as Harvard hired people (oftentimes women because their labor was cheaper and they were not allowed to run research groups) as ‘human computers’ to complete these physical data-driven tasks of maintaining photo catalogs [302] [200]. Scatterplots, one of the most commonly used forms of data visualization, did not emerge until the late 19th century and early 20th century but became a very influential technique. One of the earliest examples of a scatterplot in astronomy is the Hertzsprung-Russell Diagram [259] [117], a wide success in its time. This diagram is a logarithmic plot of star luminosity and temperature for stars [117] and is a collection of

data made by multiple observers [259]. This visualization was influential for astronomers in understanding different stages of star formation over time statistically and is still a useful dataset to explore pedagogically in data science for exploring different smoothing techniques [259].

In addition to astrophotography catalogs and new spacetime and statistical plotting techniques in the early 1900s, people were still building new physical apparatus for visualizing the sky, much like the ancient astronomers. The Atwood Celestial Sphere, a rotating metallic spherical dome with holes denoting star locations that can hold a small group of people, debuted at the Chicago Academy of Sciences (CAS) in 1913 [228]. This device is still in use at the Adler Planetarium in Chicago, but for some time, it was in use for training by the U.S. Navy. The Atwood Sphere can be useful in an educational context for learning about constellations and how they change over time in different seasons. One of the drawbacks of this device is that planetary motion can't be fully accounted for as the whole sphere moves at once. However, planetary positions could be optionally shown by uncovering their openings [228]. The Atwood Sphere became less popular once the optical company Zeiss released its more sophisticated projection-based planetarium about a decade later [228] [92]. The orbitoscope, invented in 1912 by Prof. E. Hindermann, was a spring-based contraption with two planets revolving around a sun, which could aid in teaching planetary retrograde cycles. The orbitoscope is considered the first projection-based device for planetary motions, although it was very rudimentary compared to what was soon to follow [92]. In the early 20th century, the director of the Deutsches Museum, Oskar Von Miller, wanted an interactive exhibit to fill an entire room. Initially, the plan was to have a large-scale physical mechanical model, but then a technician suggested optical projections might be more promising. Oskar Von Miller commissioned the Carl Zeiss optics company, renowned for their high-quality glass, to build an optical mechanical contraption for projecting celestial objects in a dark room [85] [183]. Several years of research and development went into creating the initial planetarium instrument, with film slides of 4500 stars that would move along their trajectories and other planets. The creators reportedly relied on the work of orrery-maker Christian Huygens from the 17th century for the

mathematical fractions required for the proper motion [92]. The Mark I planetarium, installed in Munich in 1923, was an optical mechanical planetarium that would be refined and inspire other cities to build planetariums with large viewing screens. The installation of these first projection planetaria coincides with the introduction of cinema, and as early as 1926, there is footage of the Zeiss Mark II planetarium projector [20].

Early cinema

By the mid-1930s, sound-synchronized color cinema became available and seemed a promising medium for visual science education and demonstration. An article titled “Films and Science” (1936) outlines how cinema could be used to advance science education. As one of the authors wrote, “Cinema can bridge the gulf which now separates people who have a good visual imagination from those who have not” [141]. When astronomers first started using telescopes, they had to rely on sketching. With the photograph enabling astrophotography, that artistic sketching element was phased out. However, a photograph is the projection of a three-dimensional structure onto a two-dimensional plane. Depicting 3D objects in 2D can have optical illusion-type ramifications when the dimensions are reduced, such as presenting a device such as an armillary sphere in an illustration versus having the device in front of oneself. In the early 20th century, theoretical models were robust. Still, the visualizations afforded by the printing press and 2D representations did not effectively teach these new concepts to a broader audience. One of the authors proposed that cinema could bridge the gap between oversimplified 2D representations and physical apparatus (such as an armillary sphere or orrery) that could be fragile, cumbersome to handle, and would have to be stored and transported for teaching purposes. On the other hand, cinema could open up the possibility of animating different phenomena such as “simple harmonic motion, the precession of the equinoxes, the relation of celestial and terrestrial coordinates of a star, wave motions, the trajectory of a body projected in space” [141]. Video could be magnified with a projector to share the information with a large audience. One of the authors also mentioned that stereoscopic cinematography could enhance the video and make it appear more three-dimensional [141].

These educators in the 1930s seemed to have a vision for astronomy visualization that has finally come to fruition in the present. A foreshadowing of virtual reality technology seems to have been made with the mention of stereoscopic cinematography and its potential role in sharing knowledge [141]. The earliest astronomy films I found were silent shorts from the early 1920s documenting observatories [14], people using telescopes [19] and a few of the solar eclipses in 1921 [16] [15] and 1925 [17] [18].

These early silent astronomy films typically opened with a text prompt describing the context of the scene to be shown and would be a few minutes long at most. These films predate the invention of the television by only a few years, with broadcast TV becoming commonplace in affluent households by the late 1930s [214]. Even into the early 1930s, synchronized sound seemed rare in astronomy films despite being commercially available. A couple of silent films about the moon are dated to 1932, “The Mystery Of The Moon” [21] and “A Trip to the Moon” [23]. A video containing audio narration from 1932 documents the Zeiss planetarium projector used at the Adler Planetarium in Chicago [22], and another video shows this device at the Griffith Park Observatory in 1938 [27]. Another silent film about telescopes from 1938 was *Searching the Heavens* 1933 [24], along with a narrated educational film about Galileo’s telescope [25].

By the mid-1930s, animated films in astronomy cinema emerged. A narrated, animated film from 1934 describes an old dismissed view of planetary formation called ‘tidal theory’ [26] [305]. A French animated video showcases the planets [3]. Another video from the 1930s used a combination of text, animation, and video to talk about the calendar and the length of the year [1]. Astrophotography is noted as a major achievement in a 1939 film celebrating one hundred years of photography [209]. A French video from the 1940s uses footage of different astronomy devices, such as an astrolabe and telescopes, as well as an eclipse [31]. A couple of videos in Dutch from 1943 [28] and 1944 [29] show telescopes and astronomical clocks. An educational Russian film from 1945 called *Great Universe* is a ten-minute film with engaging visuals, with stars overlaid with illustrations of constellations, animations of planetary motion, along with a zoom-out animation of a spiral galaxy with a

smooth rotation [30]. Video still has its place in astronomy for education and outreach, such as at planetariums and television documentaries. Cinema was important in the mid-twentieth century war efforts as a method of training people in scientific and technical domains in Great Britain [248]. Additionally, cinema was used as early as the 1930s for other educational nature films [11].

1.2.4 Digitization: from observations to simulations

In the early 1940s, electrically programmable computers began emerging. Early electronic computers were not yet advanced enough to produce engaging astronomy visuals as effective as some of the physical apparatus described in this literature review so far. This transition between analog and digital media could have contributed to the “dark ages” of visualization. It would take quite a few decades for digitally generated and projected video to replace film-based projection planetariums. Nevertheless, in the mid-20th century, astronomers were propelled into the digital era, first with computer-based hydrodynamic simulations followed soon after by digital photography.

In 1948, John von Neumann proposed a method for modeling hydrodynamic systems that would use numerical computer simulations to converge upon a solution, which would then be studied to build intuition, theorize, and refine the physical models that describe the universe [97]. Early computers were not well equipped to visualize these types of data, but their numerical outputs could still be studied. Simulations such as these provide a way for theorists to code and test their hypotheses, and this framework for studying hydrodynamical systems is still an important area of research as open questions remain in physical models of star formation [44]. By the 1950s, the first stellar evolution codes written by Martin Schwarzschild would take up half of the computing time on John von Neumann’s MANIAC computer [97] [42]. In an effort to make computing mainstream, Neumann wrote stellar evolution codes as an example of the utility of high-performance computing for solving complex nonlinear equations that do not have easily hand-derived solutions [42]. Neumann’s work was pioneering in terms of using electronic computers to model physical systems and set us in the direction where today’s computing power and cosmological models can produce virtual galactic clusters with realistic

masses when compared to observational methods. Some early astronomy simulation was reported with visually simplistic statistical plots describing the results [33] [166] [43]. Another simulation documented in a paper “Galactic Bridges and Tails” (1972) used over a dozen small-multiple point diagrams to illustrate simulation results of a galaxy made up of 120 test particles (modern simulations contain billions of particles) perturbed under different conditions [272]. Another early simulation used a 2D plot with velocity streamlines and isocline grids to visualize the “shock-driven implosion of interstellar gas clouds and star formation” [304]. Visualizing simulation data remains an active area of research as modern cosmology simulations can contain billions of particle tracers and thousands of galaxies.

Supernovae were first identified in the 1930s by astrophysicist Fritz Zwicky [160]. Supernovae are stellar explosions, ejecting material that later becomes the raw materials for future planets and stars. By the 1960s and 1970s astronomers were still measuring supernovae with photographic plates, which are more prone to error than modern instruments [160]. Also in the 1960s, computers were used to produce the first detailed models of supernova explosions [97]. Theoretical astrophysicists use computers to simulate the motions of gas over time during cosmic events such as supernovae, as computers can be useful for approximating solutions to integrals that can be cumbersome to calculate analytically [119]. Some of the challenges with early simulations were that they had to be limited to simple cases such as assuming spherical symmetry, reduced down to one or two dimensions. Modern cosmological simulations operate in three spatial dimensions, capturing snapshots across time. Despite this long history of simulating supernovae, there is no cohesive theoretical model for how they work [84].

In the 1970s, observational astronomy started to shift from analog methods to digital. While computers were already being used for a couple of decades on simulation astronomy tasks, observational recording methods were still heavily reliant on analog photography methods. This started to change with radio telescopes such as the Einstein Observatory X-ray space telescope [8] and the Very Large Array (VLA) that went online in the 1970s and relied on computers to facilitate capturing observations [97]. Additionally, in 1969 the first charge-

coupled devices (CCDs) were invented, and by 1976 the first digital images captured from a 61-inch telescope were made by Jim Janesick (JPL) and Brad Smith (University of Arizona) [12].

Technology behind holography, an optical technique that can use lasers or white light (depending on the method) to record and reconstruct three-dimensional images as holograms, was originally developed in the mid-20th century between the 1940s and 1960s [152]. Holograms seem like they could have been a promising technology for 3D astronomy visualizations, but appears underexplored in the literature at this time. Potential drawbacks for wide acceptance could have been its reliance on photographic slides at a time when everything was becoming more digitized. Recently it seems holograms are becoming a more active area of research. Papers show promising improvement in generating and playing back three-dimensional holographic video at real-time rates (30 fps) in 2020 [41], which is much faster than only ten years earlier in 2010 (0.5 fps) [59]. As recently as 2019, an astronomy research group demonstrated a holographic display that could be used to display astronomical objects [299] [256]. Some of the benefits of holographic displays are that they do not require special glasses to be worn, and can be viewed from multiple angles, which could be helpful in a group setting.

In addition to capturing images and controlling telescopes, computers were used in astronomy during this time for data cleaning and preprocessing. For example, ground-based telescopes are limited by atmospheric conditions, and telescope arrays such as the Very Large Array (VLA) require dimensional reduction techniques to make the data accessible. A software called Astronomical Image Processing System (AIPS) emerged in the 1980s specifically for handling data acquired from the VLA but has been extended to work with other radio telescopes [130]. In 1981 The Flexible Image Transport System (FITS) was standardized as a digital file format for storing astronomy images and non-image data [296]. Both AIPS and FITS continue to be used. The 1980s were rich in astronomical discovery, likely due in part to the continued advances in computing. The survey “The Decade of Discovery in Astronomy and Astrophysics” from 1991, is an important, somewhat binding, document that sets priorities for the entire astronomical community for the next few decades. This

document lists a number of achievements during the 1980s, where "microcomputers came into wide use for the control of data acquisition at telescopes, and theoretical simulations were extended to a wide variety of complex astrophysical phenomena" [97].

During this decade the observational and theoretical astronomy communities both had evidence supporting the "Big Bang" cosmology, galaxy surveys indicated that there is more large-scale structure to the universe than originally thought predicted by physical modeling and that galaxies might contain black holes at their center [97].

This survey asked astronomy experts for recommendations of high-impact projects that would be worth federal investment through the rest of the decade. With regard to computing, the committee recognized the importance of pursuing theoretical (simulation-based) astronomy. Early simulations could only relatively few particles, modeling at most a small group of galaxies [97], whereas modern cosmological simulations such as Illustris-TNG have tens of billions of high dimensional particles and contain millions of galaxies [115]. This timeline is consistent with the 1991 report, in which it suggested that it would take 'tens of person-years to develop' reliable large-scale simulations of different cosmic phenomena [97].

Additionally in this report, the committee emphasized the importance of creating a digital archive of astronomy data, coupled with high-speed networks [97]. The utility of computer networking was revolutionized in early 1990 when the first web browser was released by Tim Berners-Lee [56] [55]. By the early 1990s, both simulations and observational astronomy data required substantial computer storage, where some instruments were producing tens of gigabytes of data per day. While this does not seem like much data with today's consumer-grade storage capacity, at the time it signaled a clear need for an intervention [97]. High-speed networks would make it possible for researchers to easily host and share instrument data remotely and across institutions, ushering in new data visualizations that utilize these public data archives.

By the early 90s, computer graphics had spread to tackling scientific problems, such as astronomy data visualizations. A paper by Robert Haber called "Visualization idioms: A conceptual model for scientific visualization systems"

(1990) discusses how the graphics pipeline can be applied in scientific applications for point and volume renderings [132]. With graphics, data can be converted from data frames and tables directly plotted into visual representations, where colors can be mapped to different variables and be used to assign different color maps to accentuate different features of the data that otherwise might be obscured. Also in the 90s, people were using graphics to visualize effects of special relativity such as time dilation [231] [294].

Discussed in the remaining chapters are three new tools from the past few years that fit into the modern astrovis landscape. Contemporary related works are discussed within each chapter that position the project within the context of the cosmological background and how they compare with other visualization tools. A recent comprehensive survey by Lan et al. [?] categorizes and compares astronomy visualization tools from the past decade. Like many other modern enterprises, astronomers today have incredible amount of data at their disposal and it is a challenge to make sense of it all. While there are general purpose visualization tools such as ParaView and Blender that have been purposed for astronomy data, many other tools have been specifically designed for astrovis. Even within astronomy, visualization solutions for one subfield may not work for another, such as individual star evolution simulations versus an interactive map of the night sky. There are many motivations to creating astronomy visualizations, as this section has shown. Through visualization, researchers can make discoveries, share the results with the larger community in an engaging way either to other researchers or the general public.

Chapter 2

IGM-Vis: Analyzing Intergalactic and Circumgalactic Medium Absorption Using Quasar Sightlines in a Cosmic Web Context

2.1 Abstract

We introduce *IGM-Vis*, a novel astrophysics visualization and data analysis application for investigating galaxies and the gas that surrounds them in context with their larger scale environment, the Cosmic Web. Environment is an important factor in the evolution of galaxies from actively forming stars to quiescent states with little, if any, discernible star formation activity. The gaseous halos of galaxies (the circumgalactic medium, or CGM) play a critical role in their evolution, because the gas necessary to fuel star formation and any gas expelled from widely observed galactic winds must encounter this interface region between galaxies and the intergalactic medium (IGM). We present a taxonomy of tasks typically employed in IGM/CGM studies informed by a survey of astrophysicists at various career levels, and demonstrate how these tasks are facilitated via the use of our visualization software. Finally, we eval-

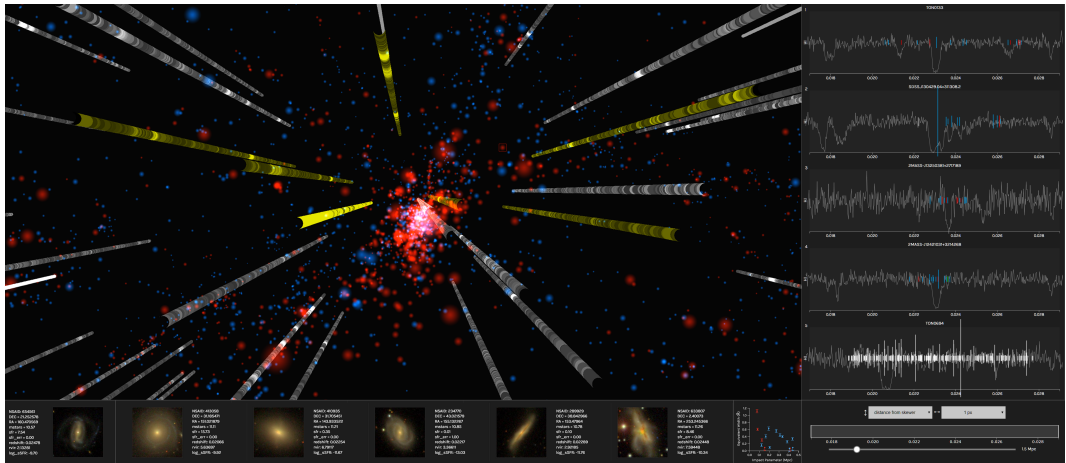


Figure 2.1: A screen capture of the *IGM-Vis* visualization application, which facilitates a range of analysis tasks using quasar sightline data in order to better understand intergalactic medium and circumgalactic medium absorption features. The Universe Panel (upper left) shows a 3D map of galaxies in the Coma Supercluster, along with “skewers” representing absorption signals in spectra of background quasar sightlines. The Galaxy Panel (lower left) provides descriptive metrics for selected galaxies. The Spectrum Panel (right) displays the spectra and marks nearby galaxies, facilitating comparative analysis between galaxies and absorption, and provides interactive controls with to select which regions of the 3D map are visible, to choose the zoom level of the spectra, and to update the profile plot to its left in the Equivalent Width Plot Panel.

uate the effectiveness of *IGM-Vis* through two in-depth use cases that depict real-world analysis sessions that use IGM/CGM data.

2.2 Introduction

Since the first astrophysical discoveries of absorption lines in the spectra of distant quasars over 50 years ago [246], the lines-of-sight to these objects that skewer the vast, intervening expanses of the Universe have served as indispensable probes of cosmic structure formation and evolution [46,47]. These absorption signatures reveal complexes of gas intersected by the sightline; these gas complexes, which compose the intergalactic medium (IGM), inhabit the relatively dense pockets of the cosmic void that also contain galaxies, galaxy groups, and galaxy clusters. On their own, IGM spectra enable scientists to study the evolution of ‘metal’ (in the parlance of astrophysics, any element heavier than Helium) enrichment in the Universe, the conditions immediately after the Big Bang, and the thermodynamic history of the Universe. However, connecting the gas complexes detected in quasi-stellar object (QSO, a class that includes quasars) spectra to particular structures requires substantial ancillary data, which in turn require their own investment of valuable telescope resources, in order to place the IGM spectra in context. In particular, galaxy surveys around the QSO sightlines enable researchers to associate particular absorption features with galactic environments, such as galaxy halos (the circumgalactic medium, or CGM) or the intermediary gas within galaxy clusters.

Galaxies show incredible diversity in their shapes, sizes, and colors. However, this diverse landscape partitions broadly into two categories: those that are actively forming stars and those that are not forming stars at appreciable rates [49]. Young, massive stars are bluer in color than long-lived, lower-mass stars, which appear red in color; thus star-forming (SF) galaxies appear bluer than ‘red and dead’ quiescent galaxies, which have few young stars and primarily contain older populations. Understanding how galaxies evolve from one state to another has become a cornerstone of astrophysics. Galaxies do not evolve as ‘closed boxes’, as evidenced by their abundances of heavy elements [271], the fact that galaxies cannot sustain star formation without

being fed by gas from their surroundings [167], and the gaseous outflows they expel [287]. This dynamic view of galaxy evolution has brought the CGM into sharp focus, as this interface region between galaxies and the IGM likely hosts many of the processes involved [278].

Since the launch of the Hubble Space Telescope (HST) in 1990 and especially the installation of the Cosmic Origins Spectrograph (COS) [129] instrument in 2009, the astrophysical community has amassed large quantities of both the QSO spectra and accompanying galaxy surveys to conduct these analyses, but the analysis methods employed to date have largely been limited to imposing set criteria during sample selection and focusing the ensuing analysis on searching for correlations within those limited parameters. Punzo et al. [225] advocate for the use of visualization tools to assist astrophysicists with a richer set of analysis tasks, taking note of the common pitfalls in many of the current applications used by the astrophysics community, which include the complexity of user interfaces and the lack of interactive analysis features.

We present *IGM-Vis*, a novel visual analytics software application that facilitates more sophisticated IGM and CGM analysis tasks than are available in existing analysis tools and that provides a series of integrated 2D and 3D views of galaxies, quasar sightlines, and analysis plots. The development of *IGM-Vis* was informed by conversations with dozens of astrophysicists who work with these datasets on a daily basis as part of their research, including in-depth interviews with eight IGM/CGM experts who spent a significant amount of time with the application. In addition to the contribution of *IGM-Vis*, we introduce a taxonomy of tasks relevant to the IGM/CGM community, derived from a comprehensive survey of astrophysicists at different career levels, and provide detailed use cases illustrating the scientific workflow of astrophysicists, and that highlight the effective use of *IGM-Vis* for IGM/CGM identification, analysis, and presentation tasks. Astrophysics terminology that may be less familiar to visualization researchers is defined throughout the paper, and summarized in the Glossary (Section 6.1).

2.3 Background & Related Works

A state-of-the-art review of observational and theoretical CGM research is presented by Tumlinson et al. [278], which emphasizes the importance of the CGM within the larger context of galaxy evolution. The CGM can be roughly defined as the gaseous envelope surrounding a galaxy, with a size often expressed as the galaxy's *virial radius*, the approximate maximum distance for which matter is gravitationally bound. Gas flows between the IGM, the CGM, and the interstellar medium, and characteristics of the gas are typically observed by measuring absorption lines in the spectra of light emitting objects behind the gas clouds. Visualizations used in contemporary astrophysics research include spectral plots, which show the data directly and reveal the absorption from material along sightlines, and absorption profile plots such as *equivalent width*, which measures the absorption line strength, versus the projected distances of nearby galaxies. *IGM-Vis* generates interactive versions of these plots on-the-fly for selected quasar spectra, making it easy to quickly associate galaxies with their imprints upon the absorption spectra.

The landmark COS-Halos survey [279] investigates the CGM of forty-four galaxies by selecting both star forming (SF) and quiescent galaxies over a range of mass. Key results of this survey are that the CGM exhibits strong absorption of neutral hydrogen (H I) for both quiescent and SF galaxies [269], and that the CGM contains at least half of all the non-dark matter in galaxy. Other studies also find that galaxies are correlated with the strongest H I absorbers, with the weaker absorbers likely tracing diffuse cosmic filaments and the IGM [95, 224]. *IGM-Vis* provides a novel interface for analyzing both IGM and CGM data, and enables researchers to investigate the relationships among galaxies, cosmic structure, and absorption patterns.

Cosmological simulations based on the cold dark matter paradigm predict that matter in the Universe is organized into a Cosmic Web (also known as large-scale structure), as elongated, interconnected filaments formed from dark matter contain low density IGM gas as well as galaxies and their CGM [230]. Indeed, most of the non-dark matter mass in the Universe likely resides in the IGM [91]. In regions of the Universe nearer to our own Milky Way, large surveys can reveal the Cosmic Web traced by galaxies [125]. A study by Wakker et

al. [291] uses HST/COS to probe one Cosmic Web filament and its imprint of H I absorption lines. *IGM-Vis* facilitates the analysis of multiple filament structures using quasar sightline data.

A range of visualization tools have been created to mitigate the complexity of astrophysics data. Popular web applications, such as the The Sloan Digital Sky Survey’s SkyServer [309] and the World Wide Telescope [127, 234], compile and present an enormous amount of astronomical image data. The European Space Agency’s *ESASky* [48] provides access to data from multiple astronomical archives, and can display the sky at different wavelengths. However, these websites do not provide any tools to analyze the data directly. Similarly, *mViewer* [57] enables a user to merge multiple image layers, using an image mosaic engine to project multiple 2D images into common astronomical layouts. *AstroShelf* [201] also facilitates querying multiple datasets, enabling a scalable navigation of data and data annotations. Sagristà et al. [239] introduce visualization tools to navigate observations made by the Gaia Spacecraft. Luciani et al. [176] introduce an interface to control the transparency of multiple image layers so that relevant data from multiple datasets can be seen at the same time. Work by Boussejra et al. [68] leverages visual programming techniques to filter and analyze multi-spectral datasets. *IGM-Vis* emphasizes the presentation and analysis of spectrum data, and contextualizes these spectra with images for user-selected regions of the Universe on demand.

A number of tools present astrophysical elements as volumes within a 3D view [120, 266]. For example, Pomarède et al. [219] make use of images, videos, and derived isosurface structures within a 3D representation to show galaxy position, velocity and density fields, gravitational potential, and velocity shear tensors. Punzo et al. [225] also note the importance of coupling 3D views with alternative visual representations, and emphasize interactive data filtering in order to investigate relevant elements. Popov et al. [221] explore methods to visualize dynamical structures in cosmological simulation data, showing how 3D plots can be used to compare the resulting outputs from various computational methods. Haroz et al. [136] include a 2D parallel coordinates plot alongside a 3D visualization to emphasize uncertainty inherent to an astronomical dataset or when found through a comparison of datasets. Fujishiro et

al. introduce *TimeTubes* [121], which transforms temporal blazar data into an unusual volumetric structure, using ellipses to encode polarization parameters arranged as a 3D “tube” in order to identify patterns of interest. *IGM-Vis* represents galaxies as an interactive 3D scatterplot in which particular regions of the Universe are pierced by cylindrical representations of sightlines, which can then be more thoroughly examined via linked 2D spectral plots.

Visual analytics tools have been used to explore simulation data that model the evolution of the Universe [135, 222]. Almryde and Forbes [39] introduce an interactive web application to visualize “traces” of dark matter halos as they move in relation to each other over time, and Scherzinger et al. [245] present an innovative application that provides 2D and 3D views to support the analysis of halo substructures and hierarchies. *IGM-Vis* also provides a visual analytics dashboard comprised of integrated panels [100, 113, 178, 240], facilitating a workflow supporting IGM/CGM identification, analysis, and presentation tasks. In addition to these standalone software applications, a range of frameworks and platforms have been developed to support astrophysical visualization. This includes work by Woodring et al. [303], which uses ParaView [37, 45] to analyze cosmological simulation data, and the Aladin Sky Atlas [61, 65], which enables users to add annotation markers to image data catalogs. Tools such as TOPCAT and STIL [265] and Glue [51] are useful for generating and exploring tabular datasets and to explore relationships within and across related datasets. *IGM-Vis* focuses specifically on facilitating analyses of quasar sightlines and their nearby galaxies.

2.4 IGM/CGM Task Analysis

Although astrophysicists utilize a diverse set of very large datasets in their research, including data from ground and space telescopes, data transformed by computational models, and simulation data, to date there are no best practices for effective scientific workflows investigating IGM/CGM data, nor a comprehensive overview of primary analysis tasks relevant to astrophysicists using sightline data. Ludäscher et al. [177] present a formalized system to define scientific workflows, and McPhillips et al. [188] and Etemadpour et al. [111]

Data Tasks	Description
T1: Obtain Sightline Spectra	Query archives; Make telescope observations
T2: Obtain Galaxy Data	Derive measurements from spectroscopy and imaging
Identification Tasks	Description
T3: Identify Foreground Features	Identify galaxies near sightlines; Identify larger structures
T4: Measure Absorption Properties	Find absorption associated with galaxies or structures
T5: Identify Sightline Features	Find relevant features across multiple sightlines
Analysis Tasks	Description
T6: Test Correlations	Quantify relationship between absorption and galaxies
T7: Discover Absorption Patterns	Compare multiple sightlines; Generate hypotheses from analyzing sightlines
Presentation Tasks	Description
T8: Create Derived Datasets	Share data with astrophysics community
T9: Produce Plots	Create plots for presentations; Explore results interactively

Table 2.1: A taxonomy of tasks relevant to the IGM/CGM community.

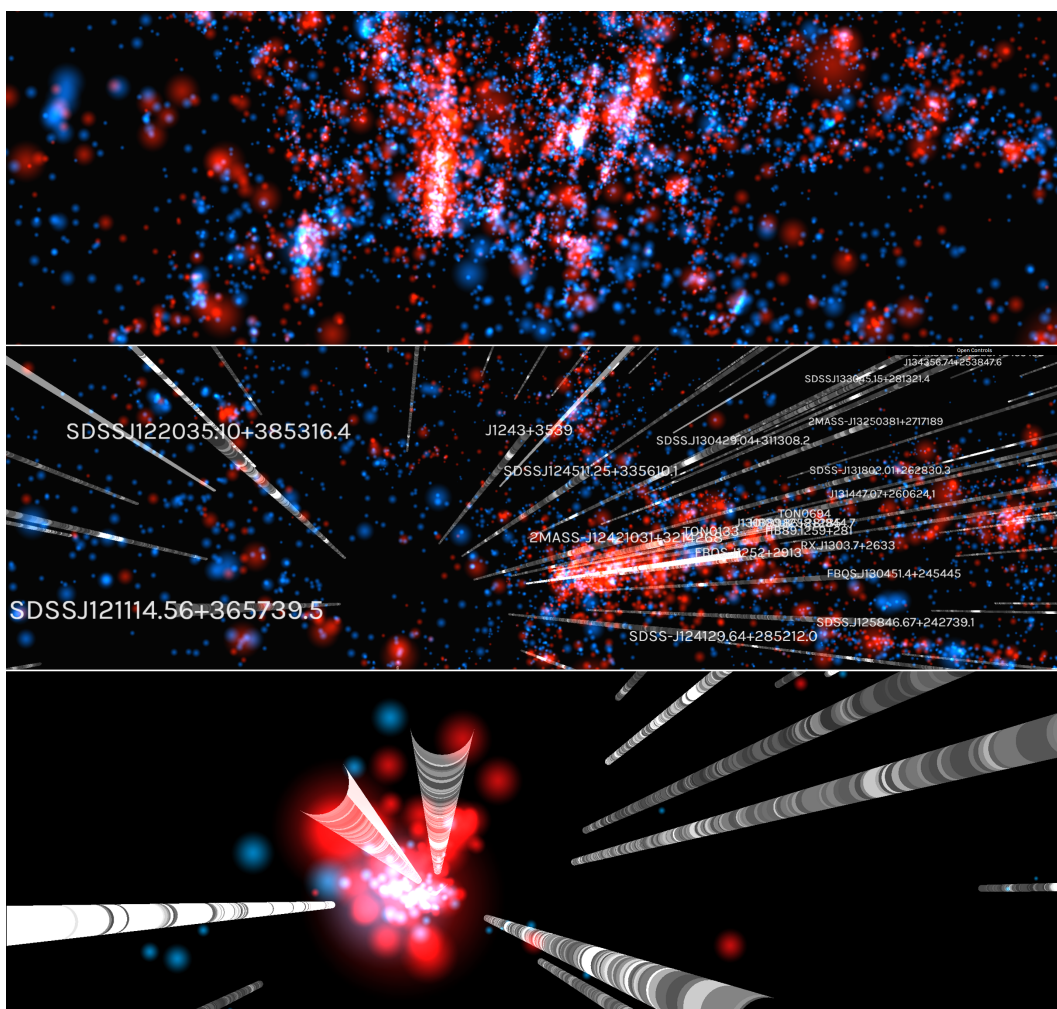


Figure 2.2: Zooming into a specific region of the Coma Supercluster dataset in the Universe Panel. To reduce visual clutter, a user can toggle on or off different elements or filter the number of galaxies displayed. The top image displays an overview of a large number of galaxies; the middle panel zooms into a region of interest, with skewers and labels toggled on; the bottom image filters out galaxies beyond a user-specified distance threshold from the skewers.

present practical approaches to implementing steps within a scientific workflow. Task analyses are a useful way to determine how visualization tools can effectively support cognition [216,273]. Isenberg et al. [150] advocate for evaluating visualization tools situated within the context of their intended use, and Lam et al. [163] introduce scenario-based approaches to evaluating visualization tools, which emphasize understanding environments and work practices. In order to determine the “why” and “how” of abstract visualization tasks [72], we conducted a structured task analysis involving both observations of astrophysics labs and surveys of astrophysics researchers at different career levels. An initial survey was conducted with 40 astrophysicists engaged in analyzing QSO data to determine the visualization needs of the community, and we conducted further in-depth interviews with 8 IGM/CGM experts in order to characterize the specific tasks they perform.

All survey respondents focus primarily on extragalactic astrophysical research (studying objects beyond our own Milky Way), and all are primarily observational astronomers who analyze data from a variety of sources. Each actively collects data from various telescope facilities but also relies heavily on publicly available archives, including the Sloan Digital Sky Survey (SDSS) and the Hubble Space Telescope (HST) archive, which are the two datasets we analyze with *IGM-Vis*. The data products from these sources include not only imaging and spectroscopy that our researchers analyze by, e.g., measuring absorption lines in quasar spectra, but also compiled catalogs of tabular data, such as galaxy positions and brightnesses measured by the SDSS team.

Our respondents indicated challenges due to the lack of shared or distributed data and software, and they also noted that visualization components are not sufficiently integrated within their current software tools. Based on our observations and conversations with researchers, responses from our survey, and a review of recent literature, we defined a list of primary tasks involved in IGM/CGM analysis, which we categorize as *data collection* tasks, *identification* tasks, *analysis* tasks, and *presentation* tasks (summarized in Table 2.1). In addition to guiding the development of *IGM-Vis*, we expect that this taxonomy will be of use for other visualization researchers who plan to create tools for astrophysicists and for students and astrophysicists from other fields who

wish to understand the primary activities of IGM/CGM researchers.

2.4.1 Data Collection Tasks

Studies of the IGM/CGM using absorption line spectroscopy primarily require the spectra of QSOs or galaxies, against which the foreground gas will absorb. These spectra are acquired via new observations or from archival sources, such as the SDSS [309], Hubble Spectroscopic Legacy Archive [212], or Keck Observatory Archive [205]. Because different telescopes and instruments are sensitive to different wavelengths or have other differing characteristics (e.g., spectral resolution), spectra from multiple sources for a single sightline will often be compiled to provide diagnostics from multiple spectral lines. These spectra often require normalization before taking measurements of absorption lines.

The defining characteristic of CGM studies relative to those focusing on the IGM is the knowledge or supposition of galaxies whose halos are probed by the sightline spectra. Relating the CGM absorption characteristics to the properties of host galaxies requires associating the two by their *redshifts*. Redshift is a measure of velocity, calculated from the observed wavelengths of intrinsic spectral features that have been shifted to longer (redder) wavelengths. The redshift provides the best estimate of distance for objects far outside our own Galaxy, as Hubble’s Law [144] and cosmological models provide the link between distance and redshift, which is measured most precisely by spectroscopy. The default dataset we study in *IGM-Vis* includes only foreground galaxies for which spectra exist, and our ensuing workflow description assumes datasets that include spectroscopic information about these galaxies in addition to spectra of the sightlines probing them.

T1: Obtain spectra of objects probing foreground environment This task may involve taking new telescope observations or querying one or more archival sources when different wavelength ranges are covered by different telescopes/instruments.

Redshifts are typically measured from the locations of known spectral features, such as strong emission or absorption lines. A researcher typically will correlate CGM absorption properties with properties of the ‘host’ galaxies,

which must be derived from the galaxy spectra and/or imaging. These include a galaxy’s luminosity, color (ratio of flux between two photometric bands), mass in stars (stellar mass, M_*), dark matter halo mass (M_{halo}), star formation rate (SFR), and metallicity. Several methods exist for measuring these quantities from both imaging and spectroscopy, including directly using spectral line fluxes in calibrated formulae [155, 157] and fitting the spectral energy distributions (SEDs) with stellar, nebular, and dust emission models [292].

T2: Obtain information on foreground objects composing the probed environment This task involves deriving measurements from spectroscopy and/or imaging, unless relevant catalogs of derived measurements are already available.

2.4.2 Identification Tasks

Astrophysical datasets often contain billions of data points, and the datasets are growing: as of 2015, astronomical instruments produce twenty-five zettabytes of data annually [260]. To feasibly define scientific questions and the samples of data to address them, one must select objects and attributes of interest from the troves of information. *IGM-Vis* integrates sky survey data, which provide information and images of individual galaxies, with quasar sightline data, which provide information about the nature of gas in different regions of space. A primary research task is to search for and identify data within a certain region of the Universe that is associated with user-specified features.

A purely IGM study may not require any information about foreground galaxies; the sample to analyze may be selected by absorption features identified with a particular ion, such as H I [170, 301], O VI [169, 275], C IV [82, 258], or a combination of these. CGM studies may follow either what are sometimes known as ‘galaxy-selected’ or ‘absorption-selected’ approaches. In the former, samples comprise particular galaxies with desired properties according to selection criteria, and their sightline probes are analyzed for absorption at similar redshifts. In the latter, one attempts to associate absorption features in background spectra with foreground galaxies projected around the sightline.

Recent CGM studies consider relationships to galaxy environment and focus on objects in dense clusters [83, 307] or groups [81, 218]. These environments

as well as filaments and voids may be investigated from a more galaxy-agnostic perspective to reveal the gas composition, kinematics, and physical conditions in dense and underdense regions of the Universe [267,268,291]. *IGM-Vis* is ideally suited for selecting targets based on environment, as the user can quickly identify filaments and dense structures visually and leverage the data available to generate hypotheses and test them on the fly.

T3: Identify features in foreground environment to investigate with spectral probes This task involves identifying galaxies nearby to sightlines and identifying larger structures comprising galaxies, such as clusters, filaments, or voids.

If adopting a galaxy- or environment-selected strategy, the task then turns to measuring the associated absorption in spectra probing the foreground objects. Often, a researcher adopts some velocity window near the redshift of interest (such as that of the host galaxy) that determines the wavelength range in the spectrum where absorption due to some species transition (such as H I Ly α) will be associated to the foreground object. *IGM-Vis* includes the ability to quickly measure absorption for hypothesis testing and exploratory analysis, and assists the researcher in identifying absorption features that can be more extensively analyzed offline.

T4: Measure absorption properties associated with foreground structures This task involves identifying absorption properties near the redshifts of foreground structures and searching for coherent absorption associated with structures across multiple sightlines.

A researcher may also conduct tomography of foreground structures wherein multiple sightlines pierce the halo of a single foreground galaxy [69], galaxy cluster [308], or filament [291]. Alternatively, structures may be identified through absorption alone, as multiple adjacent sightlines might exhibit coherent absorption at a consistent redshift [87,168]. *IGM-Vis* enables researchers to visually select multiple sightlines within a single filament or galaxy cluster, and encourages the interactive investigation of their absorption properties.

T5: Identify features in multiple spectra to investigate origin This task involves identifying interesting or coherent features across several sightlines.

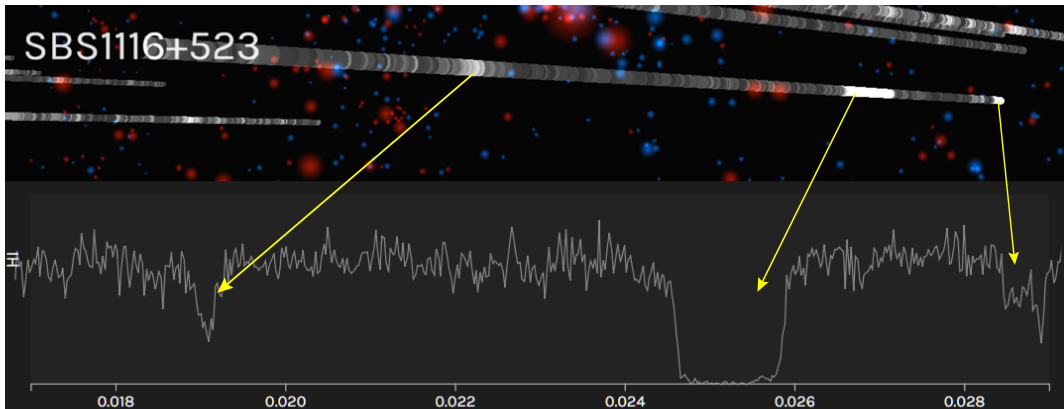


Figure 2.3: The same “skewer,” sightline SBS1116+523, represented in the Universe Panel (top) and in the Spectrum Panel (bottom). On the top, the brighter coloring along the skewer indicates absorption, which a user can explore in more detail in the associated spectral plot.

2.4.3 Analysis Tasks

Once galaxies and sightlines of interest have been identified, researchers analyze them in terms of a variety of metrics, depending on their specific goals or hypotheses. Insight into the formation and evolution of galaxies emerges from observable trends or shared characteristics among the galaxy populations. Perhaps the most important result of a CGM survey is the absorption profile, or absorption strength as a function of impact parameter. If the foreground galaxy redshifts are known, then the *impact parameter* (the projected physical distance between a galaxy and a sightline) can be calculated. The absorption strength is typically quantified by the equivalent width and/or column density, with *equivalent width* being readily measured from any normalized spectrum. The equivalent width is simply defined as the width of a rectangle with area equal to that between the absorption line and normalized flux level ($= 1$), if the height of the rectangle extended from zero flux to 1. Hence, the equivalent width-impact parameter relation is often the first (and often key) result of any CGM study. Much valuable information is encoded in these absorption profiles: How far does the CGM extend? How does the absorption strength change with increasing distance from the galaxy center? How does the profile change as measured in different ions? How do the profiles of galaxies in more isolated environments compare to those in dense structures? The answers to these questions have strong implications for galaxy formation models, and

IGM-Vis enables researchers to reason about these models effectively.

T6: Test correlations between absorption and foreground objects This task involves investigating galaxy and/or environmental properties and absorption metrics, such as by quantifying the relationship between equivalent width and impact parameter.

Any observational dataset is, to some degree, limited by the signal-to-noise ratio (S/N). At any S/N, weak enough absorption features will be ‘buried in the noise’, rendering statistically insignificant detections even if present. Also, archival datasets compiled from heterogeneous sources may comprise observations from programs with vastly different data quality requirements. Thus, in IGM spectra, one sightline might have a vastly greater S/N than another. To address both of these cases, one can employ stacking to boost the S/N in a composite spectrum. In such analysis, collections of sightlines are chosen based on some criterion (e.g., those that probe star forming galaxies within 100 kpc), the spectra are all transformed to some reference frame (e.g., the redshifts of galaxies they probe), and the spectra are coadded by, e.g., the average or median flux value at each pixel [67].

T7: Discover weak absorption by stacking data from multiple sightlines This task involves generating and testing hypotheses about galaxies, clusters, and filaments through a process that merges spectra of subpar sensitivity or from heterogeneous datasets in order to isolate and characterize absorption features not observable in individual spectra.

2.4.4 Presentation Tasks

The standard for sharing large data tables and images in the astrophysics community is the Flexible Image Transport System (FITS) format. For IGM work, data tables commonly include information about the QSOs observed and measurements of absorption systems. CGM studies often include this information as well as the properties of associated galaxies, such as stellar mass and SFR. Researchers can use *IGM-Vis* to export these data for galaxy-sightline pairs of interest.

T8: Build and release IGM/CGM datasets This task involves creating, exporting, and publishing data tables containing derived measurements from

analyzing galaxies and sightlines, so that other researchers can utilize results for future studies.

Current interactive visualization tools for astrophysical data do not accommodate the distinct aspects of CGM/IGM research. CGM/IGM science results are typically presented using static plots. For observational studies, these often include ‘stack plots’ of the spectral data showcasing one or more several transitions within the same absorption system transformed to a uniform velocity frame of reference. The relationships among various absorption and galaxy properties in CGM studies are generally reflected in 2D scatterplots [297] and may include theoretical model predictions when the comparable model data are available and/or relevant. *IGM-Vis* features the use of interactive plots that facilitate data sharing, enabling other researchers to validate analyses or to provide information that supports or challenges a hypothesis.

T9: Produce static or interactive plots This task involves creating static plots for presenting data in scientific articles and presentations or to communicate with the public, and it may involve producing interactive plots that can be used to summarize results, to illustrate novel ideas, to annotate interesting features, and to validate hypotheses.

2.5 The *IGM-Vis* Application

In this section, we provide an overview of *IGM-Vis*, and discuss how our design decisions promote the analysis tasks described in Section 2.4. (Although IGM and CGM research are both enabled by *IGM-Vis*, we chose to title the application *IGM-Vis* as IGM datasets underlie the work in both fields.) By default, *IGM-Vis* provides coverage of the Coma Supercluster and its surroundings to the extent covered by the SDSS. The dataset we employ consists of QSO spectroscopy from the Hubble Spectroscopic Legacy Archive [212] and galaxy information provided by the NASA/Sloan Atlas (NSA) [60, 195], supporting **T1** and **T2**. Additionally, using the NSA galaxy catalog, we calculate a measurement of star formation rate, enabling **T3**. Two key preprocessing steps are conducted prior to integrating the QSO data into *IGM-Vis*: the QSO spectra are trimmed to include only the wavelength range where our two spec-

tral diagnostics (H I and C IV) would be observed over the redshift range of interest, and we normalize the spectra by fitted continua to enable spectral measurements directly within IGM-Vis, supporting **T4**. For the Coma Supercluster data, we used a subset of astrophysical data localizing on galaxies and quasars that fall within a right ascension (RA) range of 115° and 260° , a declination (DEC) range between -4° and 65° , and a redshift (z) range between 0.018 and 0.023. This resulted in 19,268 galaxies and 348 quasar sightlines containing H I and/or C IV absorption data. Though we designed *IGM-Vis* around this initial dataset [75], many other quasar spectral lines, wavelength ranges, and redshift regions of the Universe can be visualized within *IGM-Vis*.

A core design decision in developing *IGM-Vis* is to enable and encourage users to begin their analysis from various starting points and to take different paths during an investigation of the Cosmic Web. To that end, *IGM-Vis* is a modular platform composed of four primary panels, each of which provides a different view of astrophysical data: (1) an interactive 3D visualization of galaxies and QSO sightlines, or “skewers”; (2) image data and metadata from the SDSS for selected galaxies; (3) interactive 2D plots of spectra for selected skewers; and (4) an equivalent width profile plot that is generated dynamically by user interaction. Fig. 3.1 shows an overview of the application. *IGM-Vis* enables comparisons between absorption features in a single QSO sightline and its surrounding galaxies as well as comparisons between multiple QSOs simultaneously. This is useful for identifying absorption patterns of a spectral line that may be related to particular features of neighboring galaxies. One key use is to quickly visually identify cosmic filaments [291] and inspect the influence these structures may have on their gas.

2.5.1 Universe Panel

The main panel provides an interactive 3D plot of the angular position and distance of all galaxies and quasar sightlines in the dataset, supporting the identification tasks **T3** and **T5**. Including a 3D plot was an important design feature requested by the majority of IGM/CGM researchers we interviewed during the development of the project, as it provides a Cosmic Web context to make it easier to reason about relationships between sightlines and galax-

ies within particular redshift ranges, which can be challenging when using 2D scatterplots representations. Galaxies are represented as partly transparent colored circles (Gaussian blurs), where blue represents star-forming galaxies and red represents quiescent galaxies, and where the size of the galaxy indicates its virial radius. Sightlines are represented as cylindrical “skewers” and colored differently along their length to indicate the amount of absorption in the spectrum (by default, neutral hydrogen H I absorption), where dark grey indicates no absorption and white (or yellow when selected) indicates strong absorption. Regions of strong H I absorption appear as bright bands on the skewer cylinders, and a user can quickly discern which galaxies reside near high-absorption regions. The skewers and galaxies are all rendered over a black background, and skewers are outlined in yellow when they are selected by a user. Both galaxy color maps and skewer color maps can be customized by the user via a drop-down options menu. Fig. 2.2 shows different views of the Universe Panel at different zoom levels and with interactive filtering applied.

Each galaxy and skewer is positioned according to their angular coordinates in the celestial sphere: right ascension (RA) and declination (DEC). Navigating the 3D view is controlled using keyboard shortcuts or via the mouse. Text displaying the name of each skewer and the visibility of the skewers themselves can be toggled on or off using either the drop-down menu or a keyboard shortcut.

Several computations are performed on the data in order to be effectively presented in the application. As astrophysical objects are measured in projection on the sky, object redshifts are used in transformations into physical distance. Each galaxy and data point along a skewer has a corresponding redshift, which are converted to physical distances (units of Megaparsecs, or Mpc) via cosmological formulae and plotted in 3D space. We then convert from spherical coordinates by using the RA and DEC angles, lookup the corresponding physical distance for each redshift, and output a 3D position vector.

2.5.2 Galaxy Panel

Directly below the Universe Panel, information about selected galaxies is displayed along the bottom of the application window and is updated each time

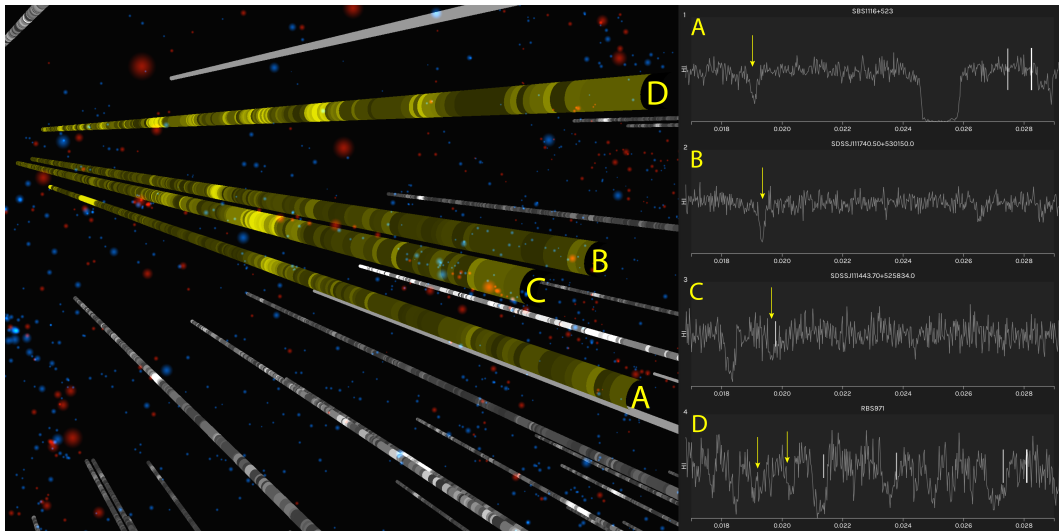


Figure 2.4: A screen capture of the *IGM-Vis* interface while investigating Use Case 1. The Universe Panel shows 4 skewers with absorption (brighter shading) at similar redshifts, and the Spectrum Panel shows each sightline’s spectrum with the absorption features of interest labeled with yellow arrows. The slider below these spectra (not shown here) has been set to mark the redshifts of galaxies within ~ 500 kpc of the skewers (white vertical lines), only the SDSSJ11443.70+525834.0 sightline has a galaxy present near the redshift of interest (green vertical line).

a user hovers over a galaxy in the Universe Panel. Each galaxy contains a list of attributes: its unique identifier (NSAID), DEC, RA, stellar mass (mstars), star formation rate (sfr), star formation rate uncertainty (sfrerr), a log of the specific star formation rate (log_sSFR), redshift, and the virial radius (rvir). When one hovers over a galaxy, this information is displayed along with its corresponding image, retrieved from the SDSS [34]. A user can interactively select and store galaxies of interest, which will then continue to populate the Galaxy Panel even after the user has moved the mouse off of that galaxy. These stored galaxies are also highlighted in the Spectrum Panel, as we discuss below, using either a blue or red tick mark to show the galaxy’s redshift within the spectral plots if its impact parameter is within a user-selected threshold of the currently selected skewers. The bottom of Fig. 3.1 shows a diverse set of galaxies found within the Coma Supercluster dataset. The Galaxy Panel provides context in support of the identification tasks (especially **T3**), and relevant galaxy data can be exported for further analysis (**T8**).

Table 2.2: Performance metrics for *small*, *medium*, and *large* IGM/CGM datasets. The *small* dataset includes galaxies and sightlines in the Coma Supercluster, while the *medium* and *large* datasets include additional galaxies surrounding the Coma Supercluster. In addition to the number of galaxies, skewers, and the redshift range, we include metrics from Google’s PageSpeed Insights (<https://developers.google.com/speed/>), including the overall Speed Score, the time-to-interactive (TTI) metric, and the input latency (IL) for both desktop and mobile application simulations, along with the average frames per second (FPS) for a) a 15-inch 2012 model MacBook Pro laptop with a 2.6 GHz Intel Core i7-3720qm CPU and an NVIDIA GeForce GT 650M GPU running macOS 10.14, and b) a custom-built PC with a 4.00 GHz Intel Core i7-8086k CPU and an NVIDIA Titan XP Pascal 12 GB GPU running Windows 10.

Dataset	Galaxies	Skewers	Redshift Range	Data Size	Speed Score	TTI	IL	FPS MBP	FPS PC
Small	4160	41	[0.0168, 0.0298]	36.4 MB	98/82	1.2s/5.3s	10ms/10ms	~60fps	≥60fps
Medium	19268	348	[0.0168, 0.0298]	217.8 MB	98/80	1.2s/5.6s	10ms/40ms	~36fps	~57fps
Large	37663	348	[0.0128, 0.0348]	257.4 MB	98/78	1.3s/5.8s	10ms/60ms	~26fps	~49fps

2.5.3 Spectrum Panel

The Spectrum Panel is located on the right side of the application, primarily supporting analysis tasks **T4** and **T6**. When a skewer is hovered over in the Universe Panel, it appears in the topmost position of the Spectrum Panel. Fig. 2.3 shows an example of how bright spots on “skewers” in the Universe panel are represented in the Spectrum Panel. Multiple spectral plots can be stored within the panel, making it easy to compare absorption profiles, and each skewer can contain multiple spectral plots (H I and C IV in the Coma Supercluster dataset). The x-axis of each plot is in units of redshift, and the y-axis represents normalized flux.

The range of redshift values displayed can be filtered using an interactive slider, which is mapped to all of the spectral plots for consistent comparison. Colored tick marks are used to represent galaxies within a user-specified impact parameter. The user can interactively toggle the visibility of galaxies beyond this distance within the Universe Panel. The relative height and width of these tick marks can be interactively mapped to different attributes in the galaxy data, such as its impact parameter, virial radius, stellar mass, or star formation rate. Fig. 2.4 illustrates a coordinated analysis using the Universe Panel and Spectrum Panel, described in Use Case 1. The user can also export a file that contains all data within the Spectrum Panel, including the name and spectra for each skewer, along with a list of all nearby galaxies within a

specified impact parameter, supporting tasks **T8** and **T9**.

2.5.4 Equivalent Width Plot Panel

Positioned between the Galaxy Panel on the bottom and the Spectrum Panel on the right is a plot for visualizing the projected distance of a quasar sightline-galaxy pair (impact parameter, x-axis) and the absorption strength (equivalent width, y-axis) of a user selected spectral region. This plot is dynamically generated on demand once an equivalent width measurement is made by selecting boundary points on a spectral line. *IGM-Vis* calculates the equivalent width of the spectral feature and plots the value against the impact parameter of the nearest galaxy to the spectral skewer. The data points are dynamically filtered to within a user-specified impact parameter range and/or a user-specified redshift range. This panel supports identification task **T4** and analysis tasks **T6** and **T7**, and these plots can be exported for inclusion in presentations (supporting **T9**). Fig. 2.5 illustrates how an equivalent width plot is interactively populated from within the Spectrum Panel, along with two example outputs created for Use Case 2.

2.5.5 Implementation Details

IGM-Vis runs on any modern web browser, and makes use of *three.js* for 3D visualization in the Universe Panel and *D3.js* for data management and plotting the interactive 2D graphs and displaying galaxy metadata. Data is preprocessed prior to being uploaded to the application using custom software functions made available with the software that: a) translates the cosmological dataset into Cartesian coordinates used in the Universe Panel, and b) creates a lookup table to enable realtime filtering of galaxies by either redshift or by distance from selected galaxies or sightlines. Fraedrich et al. [115] and Schatz et al. [243] introduce rendering techniques to display very large datasets that may include trillions of points. However, for our initial use cases— which focus on a relatively narrow region of the Universe, as is common for IGM/CGM analysis— we ingest smaller datasets consisting of between 4,160 and 37,663 galaxies and between 41 and 348 skewers. Running on a desktop computer,

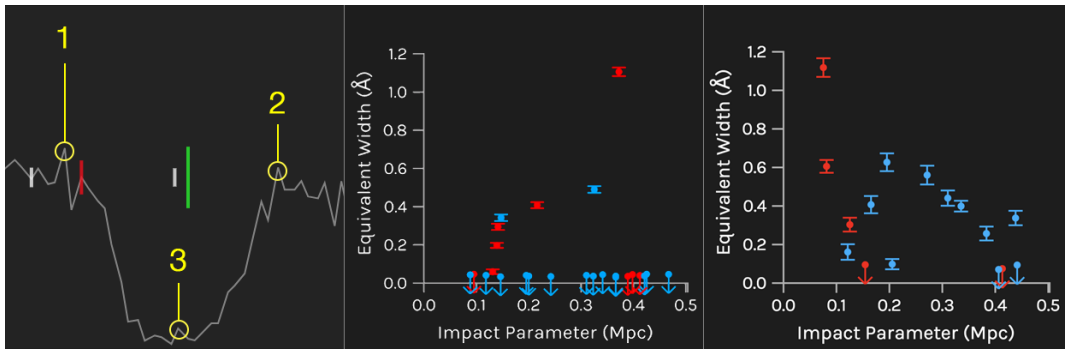


Figure 2.5: The far left panel shows an example of how an equivalent width measurement is captured: the user selects (1) the left boundary, (2) the right edge, and (3) a center reference point. The two equivalent width profiles shown in the middle and right panels correspond to Use Case 2. The middle plot shows CGM measurements within filaments near the Coma Cluster, resulting in mostly nondetections, whereas the rightmost plot shows measurements taken in regions > 50 Mpc away from the Coma Cluster, resulting in a greater incidence of detections.

IGM-Vis renders up to 20,000 galaxies and 400 skewers at close to realtime rates, while a larger dataset encompassing a wider redshift range runs at well above interactive rates. (See Table 2.2 for details.)

2.6 Evaluation

We invited eight experts (four male and four female) at different career stages— a subset of those surveyed regarding workflow and analysis tasks whose responses are summarized in Section 2.4— to spend time with *IGM-Vis* and to provide feedback on their experiences. Four are faculty at research institutions, three are graduate students, and one is a postdoctoral researcher. All employ spectroscopy of galaxies in their research, which includes investigations of fast radio bursts (FRB), galactic winds, and dwarf galaxies, along with IGM and CGM data analyses. Although none of them uses the Coma Supercluster dataset provided for their evaluation in their own work, they were able to navigate the data without any issues, and indicated their familiarity with the spectral analysis and equivalent width profile plots. Table 2.3 lists relevant details about these experts, including their career level, research areas, and which software tools they commonly use for IGM/CGM analysis.

Table 2.3: Title and research areas of the experts who provided feedback, along with the datasets and software tools they commonly use.

Title	Research Areas	Datasets	Software Tools
Professor	IGM, CGM, FRB	Keck HIRES, KCWI, HST, CASBaH	linetools, pyigm
Assistant Professor	CGM, Galaxy spectroscopy	HST COS, HRC, Keck HIRES, Gemini GMOS, HST WFC3	linetools, pyigm, veeper, redrock, specdb, IDL
Assistant Professor	CGM, Galactic winds, Quasar absorption line systems	Galaxy and quasar spectroscopy, redshift catalogs	linetools, specdb, sdss-marvin, pyigm
Assistant Professor	CGM, Galaxy evolution, Galaxy halo-gas connection	Spectroscopy, imaging databases	custom python modules
Postdoctoral Researcher	Extra-galactic astronomy, Galactic outflows, Cosmic reionization	The Hubble Legacy Archive, SDSS, GALEX	custom IDL programs
PhD Student	FRB, Spectra analysis of galaxies	SDSS, Spectroscopy data	python
PhD Student	Dwarf galaxies, Primordial helium abundance	SDSS	CasJobs, PypeIt, emcee, custom python modules
PhD Student	CGM	CGM ² , COS, Gemini GMOS	pyigm, guesses

The reaction from our survey respondents was overwhelmingly positive, with each of the experts noting the novelty of using interactive visualization software for identification and analysis tasks, and described *IGM-Vis* using terms such as “fantastic,” “great,” and “impressive.” One respondent noted that *IGM-Vis* was useful for exploratory analysis, that it provided an intuitive way to “get a feel for the data”, and appreciated that no special installation or downloads were necessary to run the software. Another noted that it gave her “an appreciation for data visualization,” as similar visualization tools do not exist for analyzing IGM/CGM data. The experts reported a diverse set of potential applications for *IGM-Vis*, with multiple users suggesting that identifying interesting configurations of galaxies, sightlines, and absorbers would be aided by *IGM-Vis*. One respondent highlighted the ease of interaction and noted how straightforward it was to compare different galaxies and sightlines. Another wrote that *IGM-Vis* is “a powerful tool to diagnose which galaxies correspond to which intervening absorption systems.” Yet another concluded that *IGM-Vis* “is extremely well-suited to diagnose the physical mechanism that leads to CGM absorption.” Our experts also lauded the potential for scientific outreach given the difficulty of describing astrophysics research methods to the general public. One respondent told us that *IGM-Vis* could be “useful for both experts and non-experts in the field,” and appreciated the ability to share the data and visualization easily in order to facilitate reproducibil-

ity. Another researcher described an insight gleaned from experimenting with *IGM-Vis*, expressing surprise that several skewers showed absorption features while no galaxies existed nearby within $1 r_{vir}$ and, conversely, was surprised that there were a number of galaxies at small impact parameter but did not have detectable absorption, despite the high covering fraction of H I in the CGM [224,261]. Two of the experts explicitly described new hypotheses generated while using *IGM-Vis* and provided details about the process of developing new ideas and performing initial investigations involving star formation rates and filament features.

2.6.1 Use Case 1: Investigating Coherence of Multisightline Absorption Signals

While the large scale structure traced out by galaxies is quite conspicuous in the Universe panel, absorption features are also apparent as bright regions on the skewers. This enables one to quickly explore the absorption features that may be associated with individual galaxies or identify absorption features that appear in multiple sightlines at similar redshifts, both within the context of the large scale environment. Here, we investigate the nature of a coherent multisightline absorption signal discovered using *IGM-Vis*, providing an example of **T5** (investigating origins) and **T6** (testing correlations).

Using multiple sightlines to reveal the spatial structure of absorbing media by sampling several points across the plane of the sky has been employed on a variety of scales. This tomographic approach has been applied to the gas clouds in our own Milky Way [66, 114, 276], the halos of other galaxies [69, 173, 235, 310], and, on the largest scales, the intergalactic medium of the Cosmic Web [87, 168]. Within the *IGM-Vis* volume, we identified at least four sightlines that exhibit H I absorption signals within a narrow redshift range of one another. Fig. 2.4 shows an interface view with these skewers visible in the Universe Panel along with the four spectra of interest shown in the Spectrum Panel. We have annotated the absorption features at $z \sim 0.019$ in these panels. Remarkably, these sightlines are separated by as much as ~ 2.5 Mpc, and only one sightline passes within 500 kpc of a detected galaxy having a similar redshift as the absorption. Using the 3D navigation in the Universe Panel,

we find that a putative filamentary structure of galaxies passes through this group of sightlines at slightly higher redshift ($z \sim 0.021$). Although potentially related to these galaxies, the absorbing complex we have discovered would have velocity separations of > 300 km/s from the approximate central redshift of the group of galaxies near the sightline skewers (crudely estimated from the data presented in the Galaxy Panel of IGM-Vis). If not bound to the same underlying dark matter infrastructure, the gas may be separated by $\gtrsim 5$ Mpc.

$\text{Ly}\alpha$ absorption is nearly ubiquitous within galaxy halos, extending to their virial radii and at least to $3 r_{vir}$ [81, 151, 224, 274]. Weak absorption, extending to larger distances may actually trace the filaments and Cosmic Web structures hosting galaxies themselves, and [224] has estimated the sizes of the weak- $\text{Ly}\alpha$ traced filaments to be ~ 400 kpc. Thus, if the coherent absorption we detect across these four sightlines probes the same gas complex, the size of the system would far exceed this scale and does not coincide with the nearby structure traced by galaxies.

2.6.2 Use Case 2: Investigating Absorption Patterns among Galaxy Filaments using Equivalent Width Plots

The Cosmic Web is composed of vast filaments, sheets, and walls traced by galaxies held together by a skeleton of dark matter. The intersections of these filaments, nodes, are where massive clusters of galaxies form, and these sites of galaxy cluster formation are the densest pockets of the Universe. Here, we explore the hypothesis that the CGM of galaxies within filaments that are in closer proximity to the massive Coma cluster are preferentially stripped of their gas relative to those in apparently less dense filaments further away.

We begin by using the slider below the Spectrum Panel to mark the redshifts of galaxies within ~ 500 kpc of the sightline skewers. We selected skewers that: (a) probed putative filament structures near Coma, and (b) had galaxies with impact parameters $\lesssim 500$ kpc. We then measured the equivalent widths at the redshifts marked in each spectrum panel. As shown in the equivalent width profile panel in Fig. 2.5, we registered mostly nondetections at all impact parameters. This is a bit surprising, because we selected galaxies well outside Coma itself (but within filaments *near* Coma), and a high detection rate of

H I is expected [83]. Next, we selected skewers that also probed filaments but those that are (a) well separated from Coma on the sky, translating to distances of > 50 Mpc and (b) had galaxies with impact parameters $\lesssim 500$ kpc. Similarly, we measured equivalent widths at the redshifts of these galaxies. Intriguingly, we measured 12 H I detections, with only 4 nondetections at all impact parameters.

Our experiment within *IGM-Vis* appears to validate our hypothesis: the higher detection rate of H I and the large Ly α equivalent widths we measure for the non-proximate filaments to Coma are consistent with a picture where the CGM in filaments ‘feeding’ Coma are indeed preferentially stripped. We acknowledge the dangers of possible confirmation bias in selecting the skewers in each category, but a unique hallmark of *IGM-Vis* is the ability to test and retest such hypotheses *in mere minutes*. From here, we will proceed with a more rigorous analysis of the data to establish the statistical significance of this result. As we have identified several skewers with nearby galaxies that do not exhibit detectable absorption, we can easily extract this list of sightlines from *IGM-Vis* (**T8**) and stack the spectra to check whether a signal emerges that is too weak for detection in the native S/N of the data (supporting **T7**).

2.7 Conclusion & Future Work

IGM-Vis was developed through an iterative design process that included multiple rounds of feedback both from astrophysicists and visualization researchers over a 13 month period between February 2018 and March 2019. The use cases presented above demonstrate that *IGM-Vis* is already empowering astrophysical investigation from a wholly different, environmental context-sensitive perspective than those commonly employed by the IGM/CGM community. In particular, *IGM-Vis* enables a range of identification, analysis, and presentation tasks that are not well supported by other visualization tools. Based on the feedback from astronomers with varying interests and at different stages of their careers, we have additionally identified a range of data collection and preprocessing tasks that are not currently supported directly in *IGM-Vis*. Indeed, currently, the the most challenging aspect of our application

is transforming heterogeneous data sources into a form that can be ingested into *IGM-Vis*. However, we have developed open source JavaScript functions that simplify astrophysics data processing; for future work, we plan to incorporate them directly into the visualization. These tools are freely available via our GitHub project repository at <https://github.com/CreativeCodingLab/Intergalactic>, along with detailed instructions on how to use *IGM-Vis* and import custom data, an interactive web demo that was used for Use Case 1 and Use Case 2 and to create all figures in this paper, documented source code, and a video tutorial.

2.8 Acknowledgements

This research is funded by Hubble Space Telescope grant HST-AR #15009. Hardware used to develop the project is provided through the Nvidia GPU Grant program.

Chapter 3

CosmoVis: An Interactive Visual Analysis Tool for Exploring Hydrodynamic Cosmological Simulations

3.1 Abstract

We introduce *CosmoVis*, an open source web-based visualization tool for the interactive analysis of massive hydrodynamic cosmological simulation data. *CosmoVis* was designed in close collaboration with astrophysicists to enable researchers and citizen scientists to share and explore these datasets, and to use them to investigate a range of scientific questions. *CosmoVis* visualizes many key gas, dark matter, and stellar attributes extracted from the source simulations, which typically consist of complex data structures multiple terabytes in size, often requiring extensive data wrangling. *CosmoVis* introduces a range of features to facilitate real-time analysis of these simulations, including the use of “virtual skewers,” a simulated analogue of absorption line spectroscopy that function as spectral probes piercing the volume of gaseous cosmic medium. We explain how such synthetic spectra can be used to gain insight into the source datasets and to make functional comparisons with observational data. Furthermore, we identify the main analysis tasks that *CosmoVis* enables and

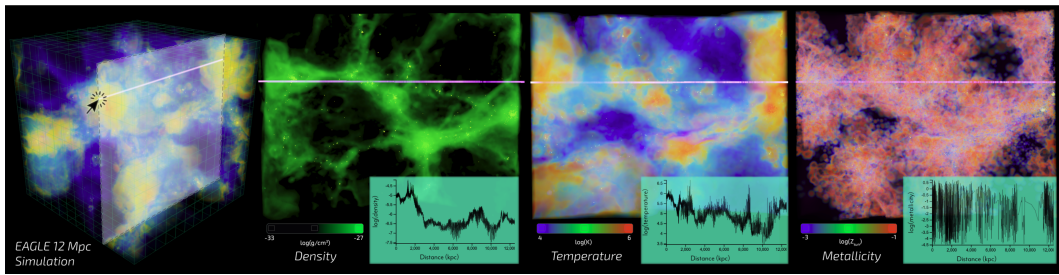


Figure 3.1: Multiple volume renderings within the *CosmoVis* software illustrating the various types of gas attributes that can be retrieved from the cosmological simulations. Here, a user interactively places a “virtual skewer” within the *EAGLE* 12 Mpc simulation, as shown in the leftmost panel (temperature). These skewers can be used to generate absorption line spectra, ion column densities, and other properties. The remaining three panels display gas density, temperature, and metallicity fields within the cross-section shown in the first panel, with the same skewer running through each panel. Plots displaying thermodynamical properties of the gas intercepted by the skewer are shown in the lower-right corner of the second, third, and fourth panels.

present implementation details of the software interface and the client-server architecture. We conclude by providing details of three contemporary scientific use cases that were conducted by domain experts using the software and by documenting expert feedback from astrophysicists at different career levels.

3.2 Introduction

Hydrodynamical simulations of galaxy formation have become an essential tool in modern astrophysics over the past two decades. With their help, astronomers have gained unprecedented understanding of the structural and chemical composition of our Universe, as well as the key processes driving galaxy formation and evolution. While the many variables in these complex simulations are tuned to match salient observable characteristics of the galaxy population in the known Universe, a range of additional features within simulated cosmologies arise from these initial conditions and serve as theoretical predictions. Simulations are therefore critical for interpreting observations, both of the stars within galaxies and the gaseous environments in which they reside.

Although the high resolution and density afforded by modern advances

in cosmological simulations are powerful and informative, the size and complexity of these data are prohibitive for most users to access and understand them. Consequently, the existing software instruments for visualization and analysis of cosmological simulations tend to have steep learning curves and often lack important features necessary for direct, intuitive analysis. In response to this challenge, we have developed *CosmoVis* (<https://github.com/CreativeCodingLab/CosmoVis>), a specialized web-based visualization software for interactive analysis of hydrodynamic cosmological simulation data. Developed in close collaboration with astrophysicists, it targets both expert users as well as interested members of the general public in supporting a range of analysis tasks and data explorations.

CosmoVis accomplishes three core requirements identified by our astrophysics collaborators: (1) Maintaining an extensible bank of large-scale hydrodynamical simulation datasets; (2) Enabling interactive visualization and analysis of both the discrete and the continuous 3D modalities contained in these datasets; and (3) Producing on-the-fly synthetic absorption line spectra and profiles of gas physical conditions, such as density and temperature, for probing the diffuse circumgalactic and intergalactic media within the simulation volumes. *CosmoVis* follows a client-server model in which the visualization interface is accessible through the web browser and on-demand computations are performed on the cloud where the full simulation datasets are stored, significantly reducing the client system requirements.

This paper introduces details about the *CosmoVis* interactive visual analysis tool and makes the following contributions:

- We provide an overview of the cosmological simulations and related tools used by astrophysicists to expand our theoretical understanding of the Universe (section 3.3 and section 3.4).
- We delineate primary analysis tasks relevant for astrophysicists working with simulation datasets (section 3.5);
- We provide a description of the feature set included in *CosmoVis*, along with a discussion of the iterative design process used in its development (section 3.6);
- We describe the design and implementation of “virtual skewers”, a novel

tool that simulates the absorption line spectroscopy used in observational astronomy (subsection 3.6.3);

- We provide an overview of the client-server architecture to interactive exploration and analysis of the simulation datasets via a web browser (subsection 3.6.5);
- We introduce three detailed scientific use cases that document how *CosmoVis* has been used to gain new insight into contemporary research questions in astrophysics (section 3.7).
- We summarize the response to *CosmoVis* from different communities of astrophysicists, and provide more detailed feedback from a pilot study (section 3.8).

Throughout this paper, we define astronomical terms when they are first introduced, and we provide a glossary of relevant concepts in the sidebar on the second page of this article.

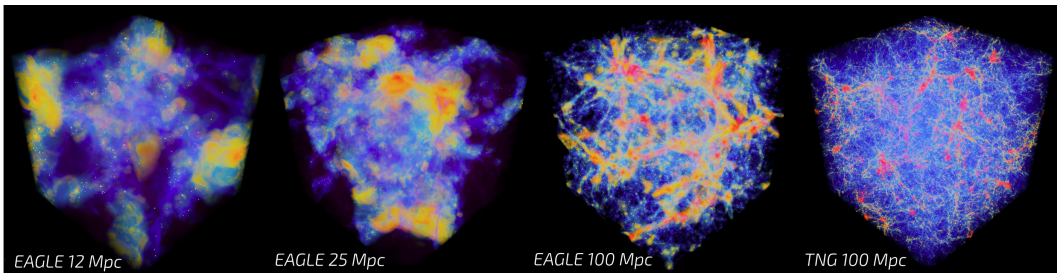


Figure 3.2: Gas temperature volume renderings in simulations of various sizes. Snapshots from left to right: 12 Mpc *EAGLE* $z=0.0$; 25 Mpc *EAGLE* $z=0.0$; 100 Mpc *EAGLE* $z=0.0$; 100 Mpc *TNG100-1* $z=2.3$. While smaller volumes (left) can be easier to work with, larger simulations (right) provide a much greater sample size of various galaxy morphologies and environmental conditions.

3.3 Cosmological Background

The cosmic web represents the largest organizational scheme in the Universe and imprinted in its large-scale structure (LSS) is the cosmological history of the Universe. Embedded within the LSS, ecosystems of galaxies are actively forming and evolving, and in the process, accreting and expelling matter and channeling energy back into the system. Cosmological simulations are essential

tools for expanding our theoretical understanding of the Universe. They predict networks of filaments, sheets, nodes, and voids, and modern simulations with hydrodynamics, and galaxy formation physics also now yield realistic populations of galaxies that inhabit the cosmic web and the circumgalactic and intergalactic gas that permeates it. In the observed Universe, the LSS is readily apparent from the locations of spectroscopically measured galaxies. However, the underlying structure must be inferred from incomplete, partial tracers rather than mapping the LSS directly as it is seen in the simulations. Furthermore, as galaxies do not generally evolve in isolation but in ecosystems within the cosmic web, understanding the galaxy-cosmic web connection is paramount.

Two main types of simulations are employed in studying the evolution of a galaxy in its larger cosmological context. *Large-volume (or big-box) cosmological simulations*, as the name suggests, spread their computational power out over a large computational domain, usually 50-500 megaparsecs (Mpc) in size. These models are able to resolve hundreds of galaxies simultaneously, but at relatively coarse resolution. *Zoom-in simulations* focus on a smaller region, often a single galaxy, and can thus achieve significantly finer resolution in modeling its behavior, while still coarsely sampling the cosmological environment around it. *EAGLE* [187, 244] and *IllustrisTNG* [215] are large-volume simulations, whereas *FIRE* [143], *Tempest* [146], and *FOGGIE* [213] are zoom-in simulations. For the most part, state-of-the-art hydrodynamic simulations include as many “resolution elements” as the supercomputing infrastructure will allow for, about 20 billion resolution elements. Fig. 3.2 shows an example of four ‘big’ boxes ranging from 12–100 Mpc along each edge.

Different simulation suites employ different codes, each with their own distinct implementation of the underlying physics (e.g., hydrodynamics, gravity). Lagrangian codes use a technique called Smoothed Particle Hydrodynamics (SPH) [192] to represent gas parcels as zero-dimensional particles for the purposes of transport, then apply a 3D smoothing kernel to “smear” them into a finite space (e.g., *EAGLE*). The kernel used typically possesses a well-defined edge (e.g., quintic spline) and is applied as a convolution over the entire simulation domain, which preserves a number of physical conservation laws, and

makes it ideal for the fluid problems with a large range in density found in astrophysics. Conversely, Eulerian codes represent the simulation domain as a series of nested Cartesian grid cells [53] and allow gas to travel between resolution elements (e.g., *Tempest*). Finally, hybrid models use tracer particles to flow with the fluid, but define non-Cartesian grid cells at each timestep generated by a Voronoi tessellation [263] based on the particle locations (e.g., *TNG* and *FIRE*). All three of these methods, Lagrangian, Eulerian, and hybrid, are simply different mathematical representations of the same fluid conservation equations, the Euler equations, which govern the inviscid (zero-viscosity) fluid motion found in many astrophysical problems. However, in any real simulation, resolution is finite and thus the fluid representation becomes an approximation. Each of these three methods reveal different types of artifacts based on their differing physical descriptions of fluid motion, not simply due to implementation differences.

Aside from the hydrodynamics, the other primary difference in these simulations is the treatment of energetic feedback from supernovae and active galactic nuclei (AGN, i.e., supermassive blackholes). These two non-linear energy sources can have a profound effect on how the galaxy and its environment evolve with time. The finest spatial resolutions found in these simulations are parsecs to hundreds of parsecs, whereas the scales at which stars and black holes form and evolve are many orders of magnitude below this, thus stars and black holes cannot be modeled self-consistently. The solution is a “sub-grid” model, which provides parameterizations of both how stars and black holes form, age, and interact with their environments through exchange of mass, energy, radiation, and gas composition. Sub-grid models are based on analytic models for how stars and black holes behave as well as results from previously executed external higher-resolution computational simulations. Small discrepancies in the parameterization of these complex non-linear processes can result in significant differences in the outcomes of the simulated galaxy population. By comparing the behavior of galaxies in different simulations using different implementations and sub-grid models, theorists can converge on which galactic behaviors are likely ‘real’, and which are artifacts of a particular numerical implementation.

The largest of these simulations span a length of ~ 100 Mpc along each side and contain physical information about tens of thousands of galaxies. These models take years to develop and are executed on the most powerful supercomputers in the world. The initial conditions of the simulations describe the distribution of matter in the universe shortly after the Big Bang, when gravitational perturbations are still linear. Time is represented in both linear terms of years and gigayears but also as an equivalent cosmological redshift z . This cosmological redshift [144] encodes a Doppler-like effect resulting from the cosmological expansion of the space propagated by light emitted by a distant object at a particular epoch and observed at the present [74].

3.4 Related Work

A recent survey by Lan et al. [164] provides a comprehensive survey of interactive visualization software used by astronomers, categorizing the main functionality of these tools as being related to one or more of the following: data wrangling, data exploration, feature identification, object reconstruction, and education and outreach. Astronomy visualization tools can also be categorized based on whether they work with observational data or simulated cosmologies. In this section, we highlight approaches that focus on or that support the analysis of simulation datasets.

General-purpose visualization tools can be useful in the context of cosmology. The *ParaView* standalone software [37, 45] and Python libraries like *yt* [282] and *Napari* [257] support multiple visualization modalities, such as direct volume and particle rendering. Woodring et al. [303] introduces features for *ParaView* aimed at cosmological simulations, such as halo detection and analysis. The *AstroBlend* library [194] extends the functionality of *yt* to the *Blender* open-source 3D modeling and rendering software. Ultimately, these tools are geared towards offline diagnostics, while *CosmoVis* aims to support online, interactive exploration and analysis of simulation datasets.

Visualization tools with more specialized goals are also available. For example, the *Open Space* “astrographics” system [62] facilitates the interactive display of data from many various sources including simulation datasets, fo-

ocusing on broader scientific communication. *Polyphorm* [110] uses an unconventional nature-inspired approach to interactively reconstruct and visualize large-scale cosmic web density fields from sparse simulated halos or observed galaxy data, but does not handle densely sampled attributes such as gas temperature or metallicity. Hesse-Edenfeld et al. [140] describe an interactive hybrid particle/volume rendering tool with support for multiple data modalities, geared towards temporal analysis of simulated universe evolution. Scherzinger et al. [245] present a similar approach specifically designed for visual analysis of dark matter simulations. *CosmoVis* provides functionality for visualizing and analyzing a wide range of simulation data types across different datasets.

Given the differences between simulation codes and even their individual runs, some researchers elect to develop dedicated tools customized for the respective simulation datasets. For example, the *Illustris TNG* simulation [215] includes its own interactive volume visualizer, and an interactive tool called *FIREFly* [124] supports the rendering of particles from *FIRE* datasets generated from the *GIZMO* code [142]. Work by Schatz et al. [243] shows how multiple GPUs can be configured to interactively visualize millions of particles simultaneously in real-time, using a dataset from the *Dark Sky* cosmological simulation [253], which contains over a trillion particles. The *Hardware/Hybrid Accelerated Cosmology Code* (HACC) framework [133] includes models of baryonic matter as well as active galactic nuclei associated with violent bursts of energy from supermassive black holes, and has been used to illustrate how this energy is imparted to the surrounding gas and affects subsequent structure formation. Several submissions to the 2019 *IEEE Scientific Visualization Contest* explored a 64 Mpc volume produced by the *HACC* simulation. For instance, Nguyen et al. [203] and Fritschi et al. [118] explore these data and observe their temporal evolution through point cloud rendering using *ParaView* and the *Visualization Tool Kit* (VTK) library, respectively. *CosmoVis* encompasses multiple simulations through a unified internal representation and preprocessing workflow.

A key feature of *CosmoVis* is the ability to perform spectral analyses, which are directly influenced by spectroscopic techniques in observational astronomy. Burchett et al.’s *IGM-Vis* [77] introduces the 3D visual representation

of quasar sightlines, measured by the Hubble Space Telescope and positioned within a volume of galaxies, showing the spatial relationship between galaxies and sightlines using empirically observed data. This has inspired the concept of “virtual skewers” (see subsection 3.6.3) as a means to visually represent linear cosmological measurements embedded within a simulated 3D volume. *CosmoVis* extends this analytic approach, facilitating new inquiries within simulations aided by interactive volume rendering of the multi-modal simulation data. This approach is conceptually similar to techniques used in medical visualization tools. For example, *RegistrationShop* [254] uses point and line elements in tomographic scans to annotate features of interest as well as to obtain tissue density profiles surrounding them.

3.5 Analysis Tasks

Our astrophysicist collaborators identified a need for an interactive visualization software that can effectively render large simulation datasets and support a range of simulation analysis tasks. Currently, there is no web-based astrophysical volume rendering application other than *CosmoVis* that allows users to interactively place skewers within a simulation volume rendering and compute multi-modal measurements such as column densities, physical properties, and synthetic absorption line spectra in a unified environment. Moreover, *CosmoVis* provides users with the ability to explore cosmological data interactively and apply on-demand filtering using different thermodynamic properties. Fig. 3.3 shows the complexity of the traditional cosmological simulation analysis pipeline versus *CosmoVis*, which consolidates various repetitive coding tasks into no-code interactions.

A main application of cosmological simulations is to validate theoretical models based on empirical observations, while also accounting for the idiosyncrasies and physical limits of the scientific instruments used to produce them. For the diffuse gas comprising the IGM and CGM, absorption line spectroscopy is the primary observational technique capable of detecting this material. Despite recent advances in computing absorption spectra using cosmological simulation data, deciding the placement of skewer sightlines remains nontrivial

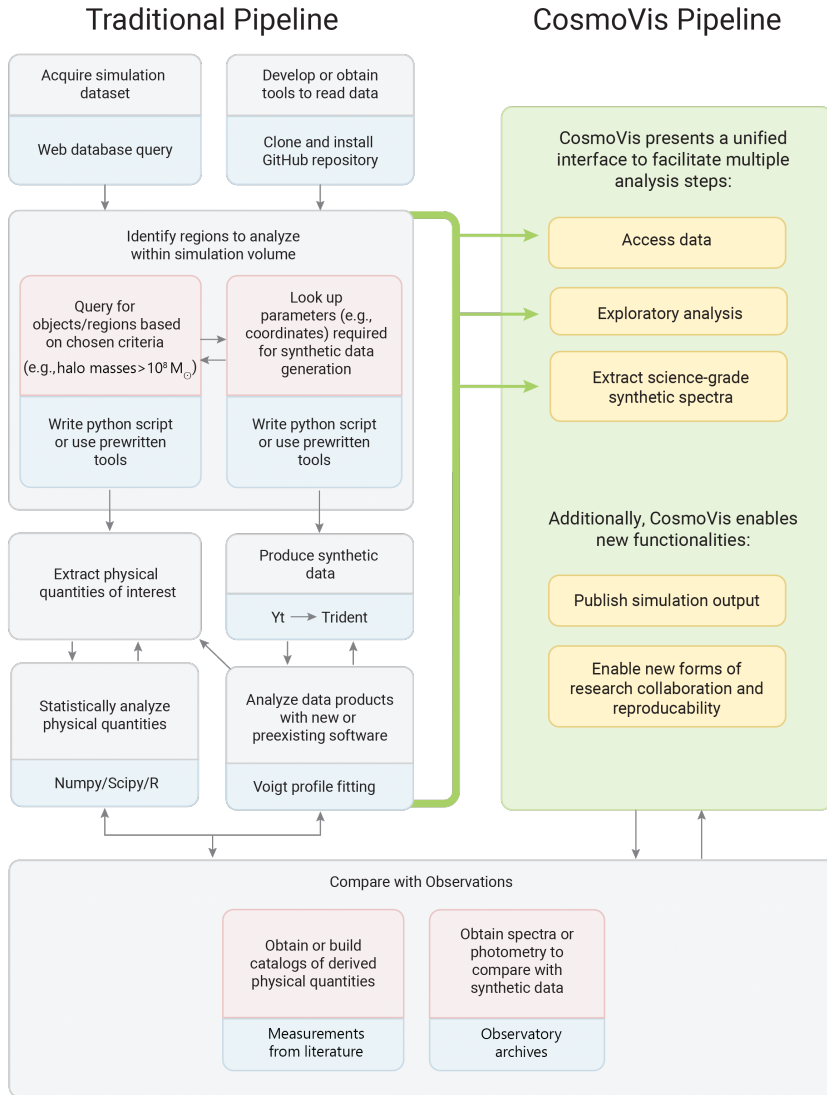


Figure 3.3: Flowchart comparison between a traditional pipeline (left) versus the simplified *CosmoVis* pipeline (right) when working with these cosmological simulation datasets. In the traditional pipeline, in order to begin comparing a simulation with observations, a researcher would have to acquire the data, write or install code to read the data, identify interesting regions via database querying, extract physical quantities and perform statistical analysis, each independently. *CosmoVis* consolidates all of these separate components into one unified interface. Note that the Python packages *yt* [282] (used for data loading and processing) and *Trident* [147] (used for computing column densities and synthetic absorption spectrographs that are aligned with the virtual skewers; also requires *yt*) are useful in both the traditional pipeline as well as for enabling interactive scientific analysis in *CosmoVis*. For more information on how *CosmoVis* utilizes these packages, refer to Related Work (section 3.4), Analysis Task 2 (subsection 3.5.2), Data Preprocessing (subsection 3.6.1), Client-Server Architecture (subsection 3.6.5), and Figure 3.7.

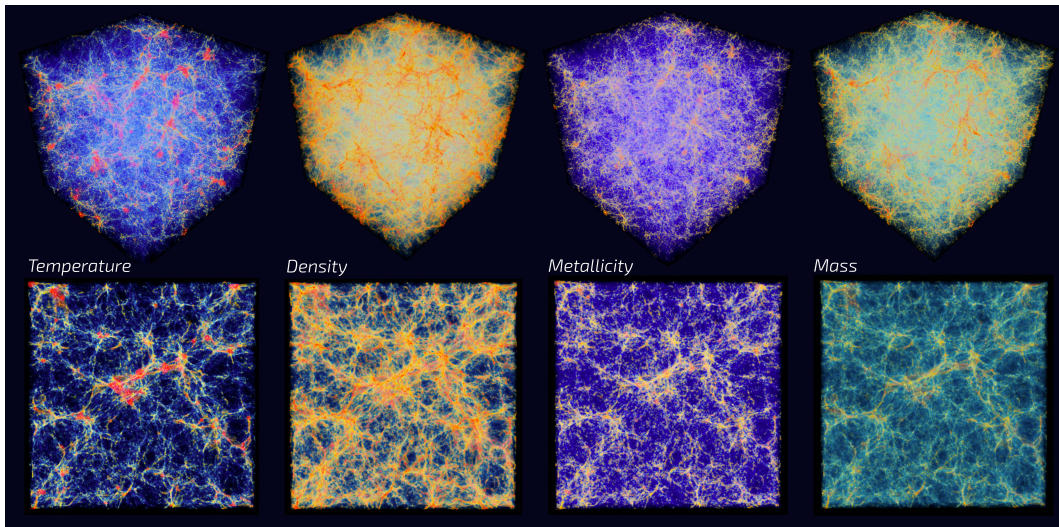


Figure 3.4: This figure shows a series of interactive volume renders of gas fields created with the *CosmoVis* visualization tool. The data is extracted from the TNG100-1 $z=2.32$ cosmological simulation snapshot. The top row depicts the entire 100 Mpc volume from an angle, and the bottom row shows a head-on perspective of a 15 Mpc thick slice of the volume.

as it requires manually coding the ray endpoints without necessarily having a good understanding of the cosmic environmental context in which they are placed.

An important goal of *CosmoVis* is to overcome the steep learning curve of extracting measurements from simulations that may be compared to those from observational techniques, which can be counter-intuitive to observers or theorists unfamiliar with a particular simulation. To maximize the usability and accessibility, we determined that the solution should be open source, provide the means to interact with multiple simulation suites, support novice and expert simulators and observers alike, and enable the generation of synthetic absorption-line spectra by pointing-and-clicking on interactive maps of a simulation in the web browser.

There are a number of studies that astronomers conduct using cosmological simulations, given the wide breadth of information they contain and how they relate to the physical underpinnings of our own Universe. As we describe in more detail below, some of the open questions about our Universe that astronomers are investigating include the following:

- What drives the evolution of galaxies?

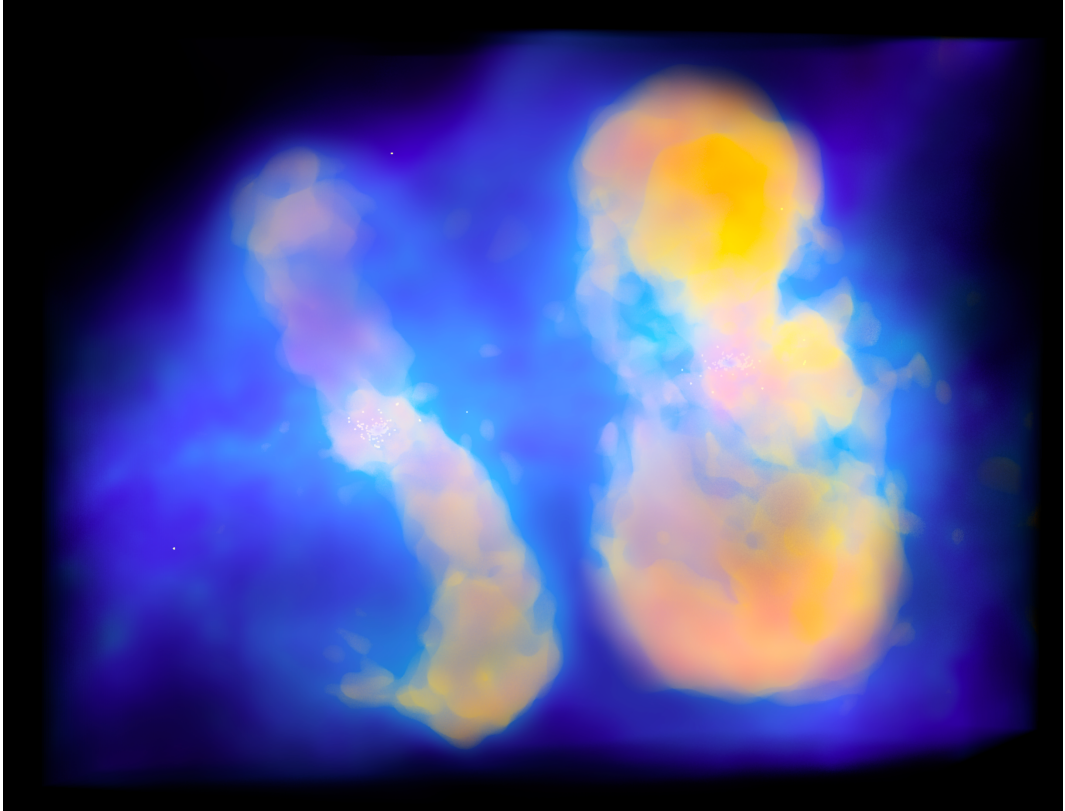


Figure 3.5: Zoom in on a region containing two distinct circumgalactic environments in the 12 Mpc *EAGLE* simulation, with color representing temperature. Star particles are represented here as faint yellow disks at the center of these two adjacent systems. There appear to be visible galactic outflows of hot gas above and below the nuclei, indicated by the yellow-orange coloration.

- What causes massive galaxies to lose their cool gas supply, depriving them of fuel for the formation of stars and leaving them quiescent?
- How are galaxies fed star-forming fuel through cool and/or hot flows of gas from the CGM and IGM?

While some observational astronomy techniques are paramount to answering these questions, many salient features are more readily measurable within simulations than in the real Universe. Studying the results of different simulations and comparing them with observations informs scientists’ decisions as to which telescopes and instruments to use and where to point them for collecting new observations. For example, astronomers can identify systems with clear signatures of outflows, or “superbubbles”, blown out from energized galaxies; they can identify particular large-scale structure such as filaments and walls, or sheet-like structures; or they can find hot regions of the intergalactic medium within the cosmic web. That is, by consulting simulations, where it may be easier to identify and isolate a particular structure than in the real Universe, new tracer characteristics that are identifiable observationally can be tested and confirmed with observational data. Here, we identify four core tasks enabled by *CosmoVis*:

T1: Identify structural patterns — *CosmoVis* accelerates the science workflow for finding relevant cosmological structures and regions of interest within those structures, such as: the large-scale structure of the cosmic web, including knots, filaments, walls/sheets, and voids; individual galaxies, galaxy groups, and galaxy clusters; and evidence of galactic winds via superbubbles and biconical outflows.

T2: Simulate QSO sightlines and spectra — *CosmoVis* replicates the observational capabilities of a telescope by generating synthetic spectra from the material along 1D “virtual skewers” in the simulations, probing regions of interest in galaxies, the IGM, and the CGM. Multiple types of data are available from a single virtual skewer (also see Table 3.2): physical properties (temperature, density, metallicity and entropy); the column densities of neutral and dozens of heavier element ions; and synthetic absorption line spectra, which are directly comparable to those obtained via instruments such as the Cosmic

Origins Spectrograph on the Hubble Space Telescope or the High Resolution Echelle Spectrometer (HIRES) on the Keck Telescope.

T3: Connect IGM/CGM observables to gas physical conditions — *CosmoVis* overlays information about both the column densities of different elements and the thermodynamics/heavy element enrichment of the gas on “virtual skewers”, facilitating new insight into complex hydrodynamical processes. The former (column densities) are measurable in actual spectra in the ‘real’ Universe, but the thermodynamical quantities such as temperature and density must be inferred from those measurements with sophisticated modeling. Various astrophysical contexts exhibit different gaseous physical conditions, such as the very hot gas within galaxy clusters, and *CosmoVis* can be used to predict how these conditions can manifest in measurements of various ions accessible by astronomical instrumentation.

T4: Compare simulation results — Through a wide range of visualization functionality, *CosmoVis* enables researchers to discover and analyze selected features of the simulation datasets and to compare results either with empirical observational data or with those across multiple simulations in order to generate and/or validate hypotheses. Examples of these comparisons enabled by *CosmoVis* include comparisons between different temporal snapshots of the same simulation, e.g., multiple redshifts of IllustrisTNG, and comparisons between simulations, e.g., EAGLE and IllustrisTNG, to investigate differences in physical prescriptions such as black hole feedback.

Taken together, the tasks supported by *CosmoVis* empower researchers to establish connections between the different gas and galaxy properties, to discover relationships between local galactic regions and large-scale filamentary structures, and to more easily identify, extract, and interpret synthetic results that can inform observations of the Universe.

3.5.1 Task 1: Identify structural patterns in simulation data

Modern cosmological simulations are large enough to contain thousands of galaxies comparable to our own Milky Way. Usually, galaxies are selected by

programmatically querying galaxy catalog databases based on a set of criteria such as halo mass or star formation rate. Cosmological simulations contain features at the galaxy, CGM, and IGM scales as well as within the intracluster medium (ICM). Galaxy clusters, where thousands of galaxies densely populate a relatively small volume, exhibit unique properties in their galaxy populations *and* gas contents relative to those in sparser environments. On an even larger scales, clusters tend to form at nodes (intersections) of the emergent filamentary structure of the cosmic web.

A primary task for astrophysicists working with simulation datasets is to identify and investigate interesting structures at a range of scales. Identifying these structures can be extremely complex in and of itself if done algorithmically [172, 185] but often serves as a starting point for further analysis. However, these structures are often *readily apparent visually*. Depending on the specific use case, relevant structures could emerge through an exploration of the simulated universe, or via a targeted search for specific structures containing predetermined characteristics of interest. Once a particular structure is identified, its various features can be further analyzed to gain insight into how the galaxies within these structures interact with them (e.g., via galactic ‘superwinds’) or other scientific questions. As different simulation datasets—zoom-in versus large-box— may be run at different resolutions, it can be difficult to study the relationships between features at different cosmological scales.

For example, “cosmic sheets” (plane-like structures with two primary axes unlike filaments that have one and are quasi-cylindrical) are laboratories for studying complex hydrodynamical processes that may produce nearly metal-free gas clouds in the intergalactic medium [181]. The relevant hydrodynamical processes, however, require re-simulating smaller regions of interest (a subset of a larger simulation volume) at increased resolutions not afforded for a large-volume cosmological simulation. Such workflow is common practice in the simulator community where a ‘large box’ simulation is run first at a fiducial (lower) resolution and certain regions, e.g., the immediate surroundings of a particular galaxy, are rerun in a ‘zoom-in’ simulation at higher resolution, drawing initial conditions from the original ‘large box’ simulation [131]. However, identifying regions of interest within the large boxes is generally a slow,

difficult process; the ability to do so visually by freely exploring the simulation volume greatly increases the efficiency of this process. Scientific Use Case 1, described in subsection 3.7.1, provides an example of using *CosmoVis* to search for and identify properties of cosmic sheets.

3.5.2 Task 2: Simulate quasar absorption line spectroscopy

The circumgalactic environments of galaxies mediate the critical inflow and outflow processes that drive galaxy evolution [278]. Because the circumgalactic medium (CGM) is so diffuse, the gas does not generally emit light at levels bright enough to observe by imaging with current telescopes and instruments except for a few isolated cases [79, 238]. Therefore, the most reliable method for detecting this diffuse medium is via quasar absorption line spectroscopy, wherein a bright background source such as a quasar, or quasi-stellar object (QSO), is observed through spectroscopy, and the foreground material leaves its imprint on the QSO spectrum in the form of absorption lines [47]. This technique is highly sensitive to diffuse media and has the power to detect gas at densities several orders of magnitude below gas that emits light and can be imaged. However, this technique does have a major drawback: the sources that are bright enough to feasibly observe and measure their spectra are somewhat rare, and one must generally compile a statistical picture of the contents and gas motions of the CGM by compiling samples of multiple galaxies that have suitable probes of the CGM [54, 80, 151, 280].

Unfortunately, individual galaxies that contain multiple QSO probes of their CGM are exceedingly rare [69, 173]. Due to this inherent limitation in observational data, cosmological simulations can play a pivotal role in testing hypotheses that would otherwise be impossible. The Python package *Trident* [147] allows for generating synthetic versions of these spectra by physically modeling the propagation of light through the simulated cosmic medium and onto specialized instruments, such as those mounted on the Hubble Space Telescope. With data products from the simulations that match the instrumental response (e.g., sensitivity and resolution) of those that take the observational data, astronomers can compare the simulation snapshots to observations of the Universe using the same analytical techniques. Scientific Use Case 2, described

in subsection 3.7.2, provides an example of using *CosmoVis* to generate synthetic spectra to test the intrinsic variance of absorption properties within a galactic halo.

3.5.3 Task 3: Connect IGM/CGM observables to gas physical conditions

According to the accepted Big Bang Cosmology, elements heavier than helium (or an atomic number greater than two) were only present in very small trace amounts at the beginning of the Universe [211]. Thus, the large amounts of carbon, nitrogen, iron, etc., out of which humans and our planet are constructed must have been forged in the nuclear reactions within stars. These heavy elements (or ‘metals’ in the common parlance of astronomers) therefore are ejected as stars shed their gaseous envelopes and explode in supernovae. These elements are then transported vast distances from their points of origin into the CGM and IGM [287]. Absorption signatures of metals have been long detected in spectra of quasars, and absorption line surveys of the circumgalactic medium have revealed unambiguously that the halos of galaxies are ‘enriched’ with metals [63, 94]. In fact, researchers have posited that nearly all of the metal line systems detected to date arise from circumgalactic environments, even if the host galaxy (that presumably lies near the QSO line of sight) is unknown [224].

Observational astronomy indicates that the IGM is also enriched with metals via the processes of star formation. Throughout stars’ lifetimes, the enormous amount of heat and pressure in their cores facilitate nuclear reactions that convert low-atomic number elements such as hydrogen and helium into heavier elements. These heavier metals can be transported far away from their stellar origins via solar winds or ejected during supernovae [36]. Nevertheless, the level of heavy enrichment (or metallicity) as well as the actual physical thermodynamical properties of the gas are quite uncertain from observational data and are heavily model dependent, requiring column density measurements of multiple metal species at various ionization states. Thus, column densities serve as a primary measurable quantity from observational datasets to infer what scientists really want to know: the gas volumetric density, temperature,

and metallicity (collectively ‘physical conditions’). For example, neutral hydrogen (H I) is abundant in the interstellar medium and is observable across a range of redshifts [226]. The modeling required to infer the physical conditions of this gas critically depends on reliable measurements of H I and other metal ions. However, comparing the metal distribution and temperature/density structure between simulations and real environments in the Universe is crucial to verifying consistency or refuting the core physics employed by simulations [227]. For example, studies examining five-times-ionized oxygen (O VI) in the *Illustris* [126,252,288,289] and *EAGLE* [187] cosmological models found too low column densities as compared to observations, whereas the newer *IllustrisTNG* [215] simulations show more realistic distributions of highly ionized oxygen (O VI) [198].

Scientific Use Case 3 (subsection 3.7.3) provides an example of using *CosmoVis* to retrieve column density measurements in order to gain new insight into the metal enrichment of the IGM. Additionally, Scientific Use Cases 1 and 2 make use of *CosmoVis* to retrieve column densities in order to analyze physical attributes of large scale structures (subsection 3.7.1) and galaxy halos (subsection 3.7.2).

3.5.4 Task 4: Compare simulation results

One of the primary reasons that astronomers consult cosmological simulations is to bring to bear actual observations taken of our own Universe on the physical models in the simulations, which are often (necessarily) broad approximations of unresolved small scale processes. For example, Rasmussen et al. [229] compared halos in cosmological simulations to observed galaxies to demonstrate that environmental effects on the hot gas detected around spiral arms effectively halts star formation. It can be useful to compare multiple simulations to observational surveys and thus uncover the strengths and pitfalls of different simulation codes, which can lead to improving the constituent models for future generations of simulations [284]. A key consideration when comparing observational data and simulated data is taking into account observational biases. Donnari et al. [105] compared galactic observations from the Sloan Digital Sky Survey (SDSS) to the distribution of galaxies in the

IllustrisTNG simulations and found discrepancies when naively querying the simulation. These could be accounted for by considering observational biases during sample selection.

Modeling the physical processes involved in star formation is essential to producing accurate simulations in terms of reproducing the number density of galaxies at a given mass and other observables. For example, without accounting for gaseous outflows, simulations tend to overestimate the galactic stellar mass content when compared to observations [101]. Therefore, star-formation driven feedback is crucial in producing galaxies with realistic masses [171]. Large scale simulations such as *EAGLE* or *IllustrisTNG* can be useful for getting a statistical overview of the mass distribution of galaxies, but they lack the resolution of zoom-in simulations such as *FIRE* [143], which can be powerful to diagnose physical nuances of gas dynamics in local star forming regions. Modeling the dynamics of gas as it flows into and out of galaxies in different cosmic environments can help astrophysicists predict how star formation plays a role in galaxy quenching, morphology, and metal enrichment. For instance, Segers et al. [249] used the *EAGLE* simulation to study the contribution of recycled gas in star forming galaxies. Star-forming fuel can be stripped during galaxy mergers, or gas can be blown out and re-accreted or “recycled” by a galaxy. Certain models show that recycled gas from stellar mass loss can contain enough fuel to sustain star formation in older galaxies [171].

Eventually, galaxies can run out of fuel or no longer have ideal conditions for star formation and become quenched. Myriad observational studies as well as cosmological simulations have shown that galaxies within dense clusters are ‘quenched’, forming virtually no new stars compared to more isolated galaxies [50,314]. Although there are several plausible theories for why some galaxies quench star formation (particularly in dense environments such as those readily visually identified with *CosmoVis*), no consensus on the exact mechanisms has yet converged, and observations are limited especially for more distant galaxies at earlier cosmological times [180]. *CosmoVis* provides functionality to investigate multiple simulation datasets in order to generate new hypotheses about data collected via empirical observation.

Furthermore, the synthetic spectra and column densities can also be ex-

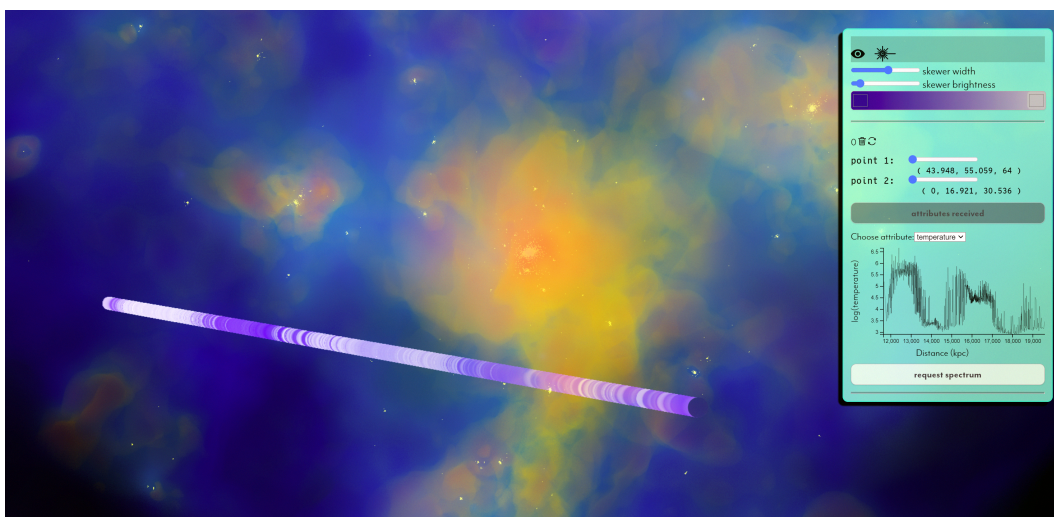


Figure 3.6: *Temperature* profile along a virtual skewer. Using the default color map, blue indicates low temperatures (10^4 K) and red indicates high temperatures ($> 10^6$ K). These temperature values along the skewer are also displayed in the line plot in the *Observe* panel on the right, where the user can also adjust the skewer position and define a custom color map. An additional skewer width option is available to adjust the radius of the cylindrical object.

ported and further examined using traditional observational analysis codes. Moreover, the 3D volume rendering itself illustrates key features of the cosmological structures and can be used to share results within collaborations or for inclusion in publications.

3.6 Application Design of *CosmoVis*

The interdisciplinary *CosmoVis* development team is made up of visualization designers, graphics researchers, and astrophysicists. We developed *CosmoVis* over an 18-month period following an iterative design process that was informed by continuous feedback from both observational and theoretical astrophysicists. We met (and continue to meet) on a weekly basis to discuss features and datasets, to conduct code reviews, and to prioritize efforts in order to provide an effective tool that will benefit the various astronomer communities working with simulation volumes. We interact with members of these communities in both informal and formal settings (see section 3.8) and are currently supporting the use of *CosmoVis* for a range of scientific investigations.

Our initial design was based on two main goals, which were to (1) create

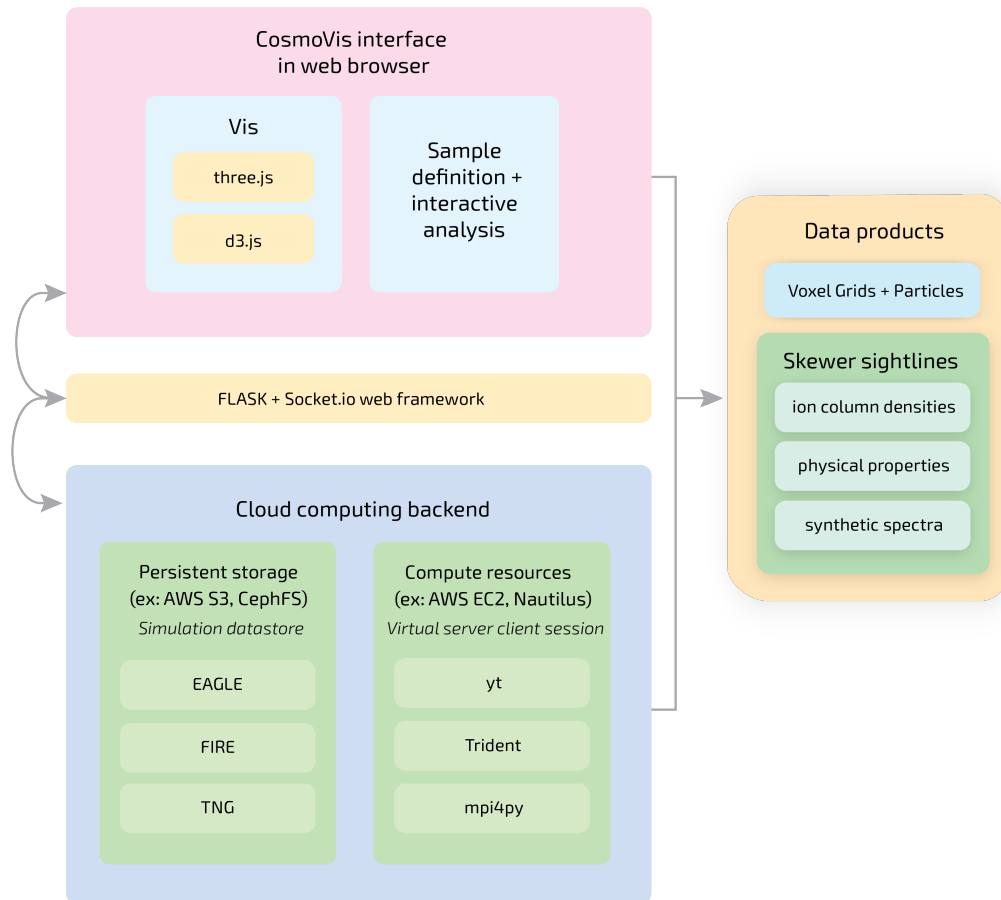


Figure 3.7: Architecture of the *CosmoVis* software. The *CosmoVis* client interface is accessible via modern web browsers, where the visualization is powered by *D3.js* for loading and displaying 2D graphs and *Three.js* for 3D volume rendering. The interface allows for interactive exploration of the simulation dataset and on-demand analysis via the placement of the virtual skewers. Requests are sent via *Socket.IO* between the client front-end and the server back-end, which can be hosted on an EC2 instance on Amazon Web Services or on the Nautilus distributed computing platform. Metadata describing simulation parameters, skewer placement, and the desired data product are sent to the server. The back-end runs the *Flask* webserver and hosts a series of custom Python scripts. When the client requests a simulation snapshot (e.g., *EAGLE* or *TNG*), it is retrieved using the *yt* package and rendered using a series of custom fragment shaders. When requested by the user (through the placement of skewers), ion column densities, physical properties, and synthetic spectra are generated by *Trident*. These data is then sent back to the client browser, where plots are dynamically created using *D3.js* and rendered atop the virtual skewers using the *Three.js* library.

an interactive volume rendering solution to visualize a range of cosmological simulation datasets, and (2) enable the placement of virtual skewer sightlines through the simulated volumes and use them to generate synthetic spectra. We surveyed existing browser-based volume visualization software tools to see if any were adaptable to include interactive “virtual skewer” placement. However, we ultimately determined that it was necessary to design and develop our own application to facilitate our novel workflow and to support tasks **T1–T4**.

Accessing and working with simulation datasets can be prohibitively complex and time consuming, so we opted to develop *CosmoVis* as a web application in order to maximize accessibility across operating systems and to avoid the need for compiling or installing software locally. Furthermore, contending with the enormous disk space and memory requirements to store and load simulation datasets require hardware resources beyond those available on even high-end laptops or workstations. Over the course of the development of the software, as we became more familiar with the needs of various astrophysics research communities working with simulation datasets, our design goals coalesced into the feature set detailed below. Here, we describe our data processing pipeline, our 3D volume rendering and particle rendering approach, the various features available via the user interface panels, and the implementation details of the client-server architecture that supports the interactive visual analysis of very large 3D datasets.

3.6.1 Data Preprocessing

Data from most cosmological simulations are distributed in the form of “tracer” particle clouds. This particle representation mirrors smoothed particle hydrodynamics, the prevalent simulation methodology. However, despite the name, these tracers in fact represent discretized field data (gas or dark matter density, gas temperature, and pressure) or agglomerations of discrete macroscopic objects (stars or black holes). A detailed list of example fields available in simulation datasets (and exposed in *CosmoVis*) can be found in Table 1 of our supplemental material. Even though such representation trivially lends itself to particle-based visualization, the sizes of these massive datasets—typically containing billions of tracers—quickly becomes unmanageable, especially

when it comes to their transfer and rendering. We therefore opted for a hybrid approach where tracers that inherently represent field quantities (gas and dark matter attributes) are converted into uniform voxel grids, while retaining the discrete agglomerates in their original particle form.

To perform the conversion we rely on the *yt* Python package. The optimal sampling rates for the voxelized grids depend on the size of the simulated domain, the number of contained tracers, as well as the actual details resolved by the simulation. Rather than attempting to determine this automatically, we preprocess several grid resolutions (between 64^3 and 512^3 , though potentially even higher resolutions could be used on systems with more capable GPUs), and we let the user choose the most appropriate size on demand. These texture resolutions are sufficient for distinguishing features in and around the cosmic web filaments in large-volume simulations, as well as for representing more fine-grained features in zoom-in simulations. By default, we initially load the low-resolution grid so that the application loads quickly. To further optimize the storage, transfer, and rendering, we compress the original Float64 voxels to UInt8 and unpack them during the rendering stage. The precision afforded by the Float64 value is not required for visualizing the data, and downsampling simulation data this way reduces file size substantially, greatly accelerating data transfer over the web. (More finely sampled data is retrieved on demand via the skewer interactions described in subsection 3.6.3.)

3.6.2 Interactive Rendering

To visualize the hybrid field and particle data, *CosmoVis* incorporates three separate rendering passes using custom shader programs: a star particle pass, a skewer sightline pass, and a volume pass using the standard emission-absorption model [186]. In the first two passes, the star particles and skewers are rendered separately and saved to off-screen position and color buffers. In the third pass, a three-channel 3D texture containing the values of each gas, dark matter, and density voxel is integrated through by a physically based ray marching shader. Here, the optical thickness of the gas and dark matter media are attenuated by the local hydrogen number density, making denser regions appear optically thicker. The depth buffers from the particle and skewer passes

are used as early stopping criteria for the integration, which allows for consistent compositing of the volume, particle, and surface colors (see Figure 3.10).

In this design, the volume is able to represent the field quantities effectively, including dark matter density, gas density, temperature, metallicity, and other attributes. Each of these quantities can be rendered using user-defined custom transfer functions. Examples showing multiple gas fields are highlighted in Figure 3.1 and Figure 3.4. Notably, we use two distinct color transfer functions in order to differentiate between the baryonic matter (gas) and dark matter filaments. The discrete star macro-particles then provide sufficient cues about the shape and orientation of galaxies. In the default configuration, in addition to rendering volumes of up to 512^3 , we can render on the order of 10^5 particles and dozens of active probing skewers at interactive rates.

Another design premise of *CosmoVis* is to provide the user with an un-hindered, real-time analysis of the visualized simulation dataset. To this end, we implemented standard modes of interaction— zooming, rotating, and panning— directly in the visualization canvas. Users can also filter and slice through the volume using sliders in the *Data Selection* panel to hone in on specific regions (such as the two clusters in Figure 3.5), interactively fine-tune the volume transfer functions within the *Layers* panel in the user interface, and place virtual skewers throughout the volume in the *Observe* panel (Sec. 3.6.4). This additional functionality helps to better localize and probe regions of interest in the data (see Figure 3.10).

3.6.3 Skewer Interaction

Virtual skewer objects can be dynamically placed throughout the simulation volume by pointing and clicking once the “drop skewer” mode is activated, as shown in Figs. 3.1, 3.6, and 3.10. The placement of skewers mimics how an astronomer might peer into the viewfinder of a telescope and observe the cosmos. Extending this metaphor to the rendered environment, we can think of the scene’s camera as a specialized sensor, e.g., the Cosmic Origins Spectrograph on the Hubble Space Telescope (HST), which captures linear samples (spectra) along its ‘pencil-beam’ line of sight. The endpoints of the skewer are initially determined by the intersection of the cursor and the boundaries of the

visualized region, and can be adjusted using sliders available in the *Observe* panel.

When the skewers are placed within the volume, additional data products become available for computation. As summarized in Table 2 of the supplementary material, these data include elemental ion column densities and physical properties along a skewer, as well as synthetic spectra directly comparable to actual absorption line spectroscopy one might obtain with HST. These retrieved results are superimposed upon the skewer as a banding pattern, indicating peaks and valleys in the data and presenting the data within the context of the structures that the skewer intersects. The incorporation of these data products using the skewer metaphor into a no-code visualization workflow is novel to the astronomy visualization community. These data are also displayed as 2D line plots in the graphical interface. A menu available in the *Observe* panel enables the user to switch between different properties such as temperature, mass, and density, as well as between different ion column densities, such as H I or O VI, which then updates both the 2D line plot and the skewer banding within the volume.

3.6.4 User Interface

The design of the user interface minimizes visual clutter while maintaining accessibility for core functionality. The full width and height of the web browser window are used as the interactive 3D visualization canvas, while four floating panels positioned to the right of the rendering canvas— *Data Selection*, *Layers*, *Observe*, and *Spectra*— provide the user with a wide range of control over various aspects of the simulation. These panels can be opened and closed as needed, so as not to obscure the main 3D volume view. Figure 3.8 shows an example of each of the user interface panels.

The *Data Selection* panel provides dropdown lists of available simulations and their preprocessed volume resolutions for gas attributes and dark matter, as well as sliders for interactively slicing the volume along the X , Y , and Z axes. A grid overlay with 1 Mpc spacing can also be toggled via this panel.

An array of options in the *Layers* panel show or hide the various data fields in the main rendering window. Selecting *Gas*, *Dark Matter* or *Stars* fields

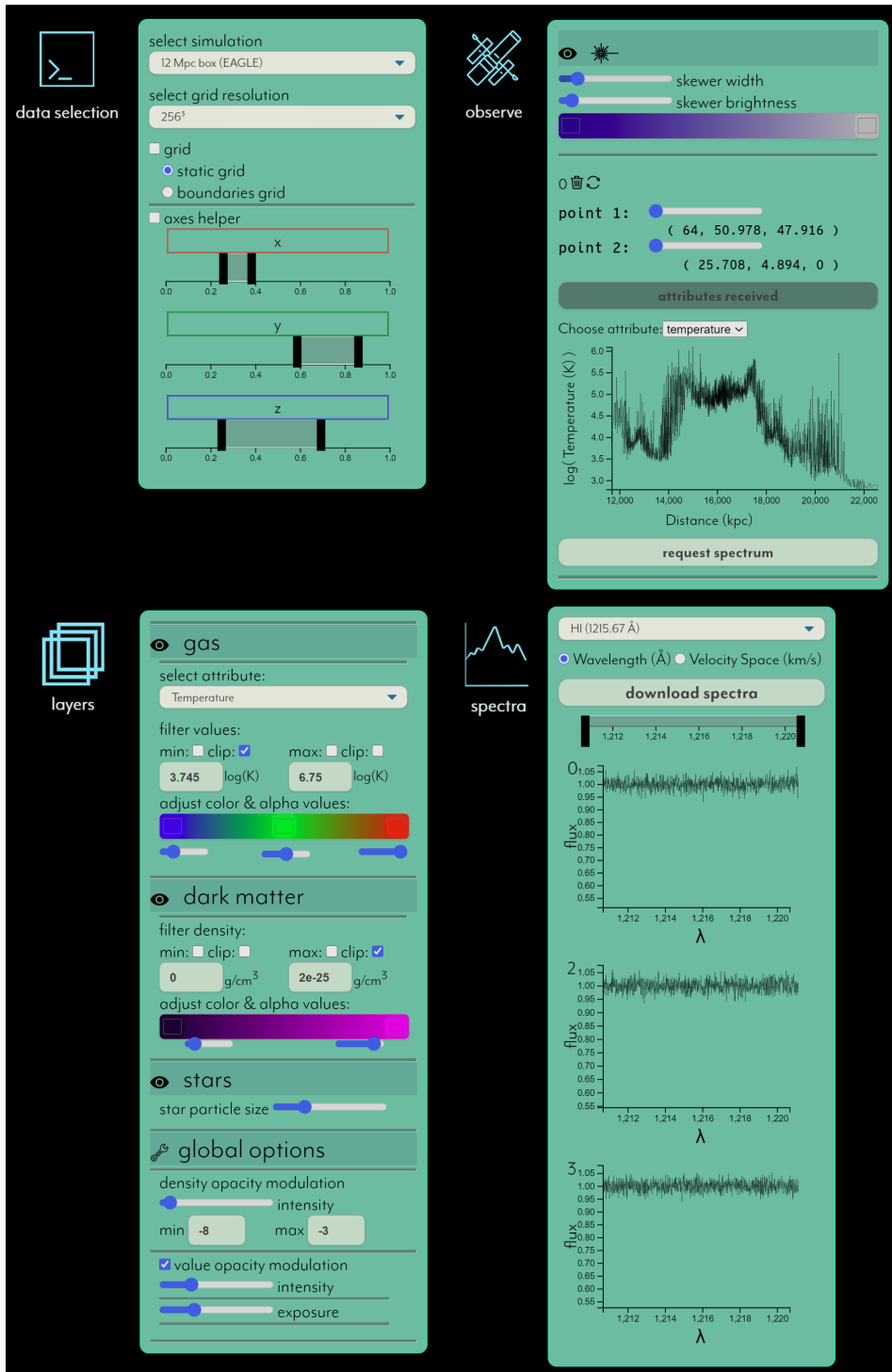


Figure 3.8: Four user interface panels can be accessed on demand to load in and interact with simulation datasets, including retrieving spectra and CGM/IGM gas physical conditions: Data Selection, Layers, Observe, and Spectra.

expands a menu for fine-tuning each field. For example, for *Gas*, the user can select the available attributes such as temperature, entropy, carbon, oxygen content, or metallicity. The user can also tune the maximum and minimum boundaries for controlling the 3-channel color transfer function, which can be customized with interactive color pickers and sliders controlling the density-modulated volume optical thickness. Similarly, the user can tune the *Dark Matter* density range and color transfer function. The *Star* menu allows for adjusting the size of the stars to accommodate different screen resolutions. (Hovering over star particles themselves provides tabular information about the corresponding sub-halo.) The user can also adjust the strength of the hydrogen gas number density modulation, which controls the optical thickness of the gas and dark matter volumes; the value modulation, which adjusts the optical thickness based on the magnitude of the active attribute; and the overall exposure of the scene.

The *Observe* panel provides functionality for placing skewers, requesting synthetic data products, and viewing data plots associated with the skewers. When clicking on the simulated universe in the canvas, skewers are placed along the camera axis, with endpoints automatically clipped to the active volume boundaries. After placement, the *Observe* panel provides controls for the skewer’s spatial extent, as well as options for generating synthetic data products: collecting column densities along the skewer for a wide variety of elements and ions detectable through absorption, as well as physical condition fields such as temperature, metallicity, and entropy. Once received by the client, the user can switch among the various synthesized data fields, updating the graph in the *Observe* panel, as well as the banding rendered along the skewer within the visualization. The second data product class that can be generated is synthetic spectra, which are analogous to physical observations captured by instruments on HST and ground-based telescopes.

The *Spectra* panel is populated with processed synthetic absorption line spectra generated through user interactions via the *Observe* panel, which can be cross-compared by centering on specific wavelengths and linked brushing. Once a spectrum is computed by the server, a corresponding new plot is displayed. Each spectrum is aligned such that when selecting a specific spectral

feature from the dropdown menu (such as the neutral hydrogen Lyman-alpha line), switching between unit spaces (either wavelength or velocity space), or using the brush slider, each spectrum can then be visually compared against one another vertically. The panel provides functionality to export all generated spectra as a FITS file (the industry standard among astrophysicists), which can be ingested into other tools for further analysis.

3.6.5 Client-Server Architecture

CosmoVis was designed using a client-server model to minimize the amount of data and processing on the client machine via the web browser by offloading the more computationally expensive simulation processing to the server. In so doing, *CosmoVis* starts up quickly and maintains interactive rates, even for large simulation datasets where a single snapshot is ~ 2 TB in size. Placing skewers within the simulation volume triggers computations that are performed remotely on the server and operate directly on these large datasets. This is necessary to retrieve physical properties and generate the synthetic spectra the locations traversed by the skewer. An overview illustrating the *CosmoVis* client-server architecture is shown in Figure 3.7.

Our data preprocessing pipeline is written in Python and uses the *yt* package to generate visualization files. These are hosted on a Python server with the *Flask* web framework, and are loadable upon request by the web client using *D3.js* on the front-end. The 3D visualization environment is rendered using *Three.js* and custom fragment shaders. Simulations are loaded into memory using *yt* on server startup, and on-demand skewer computations are enabled with *Trident*. *Gunicorn* is used to fork the data across workers such that multiple requests can be handled simultaneously. Communication between the front-end client and back-end server is managed using *Socket.io*. When a user requests spectra or column density information by placing a skewer in the volume or updating a selected skewer via the user interface panel, the skewer endpoint coordinates and the specified data types are sent via *Socket.io* to the server. On the server, synthetic data products are computed using *Trident*, and then returned to the web browser, where the data is displayed in 2D line plots using *D3.js* and rendered in 3D using *Three.js*.

CosmoVis has been deployed previously using the Elastic Compute Cloud (EC2) on Amazon Web Services (AWS). However, while this platform was cost-effective for hosting small volumes, supporting multiple large-volume simulations proved to be prohibitively expensive. Currently, we are using the Nautilus distributed computing platform, which provides free compute cluster resources to the scientific community [40]. *CosmoVis* is containerized using *Docker* and configured using *Kubernetes* to run on Nautilus. In addition to our publicly accessible web application, source code for *CosmoVis* is freely available for users who wish to install a custom configuration on their own servers. Table 3 and Table 4 in the supplementary material document provide information about the loading times and framerates of *CosmoVis* under different client and server configurations.

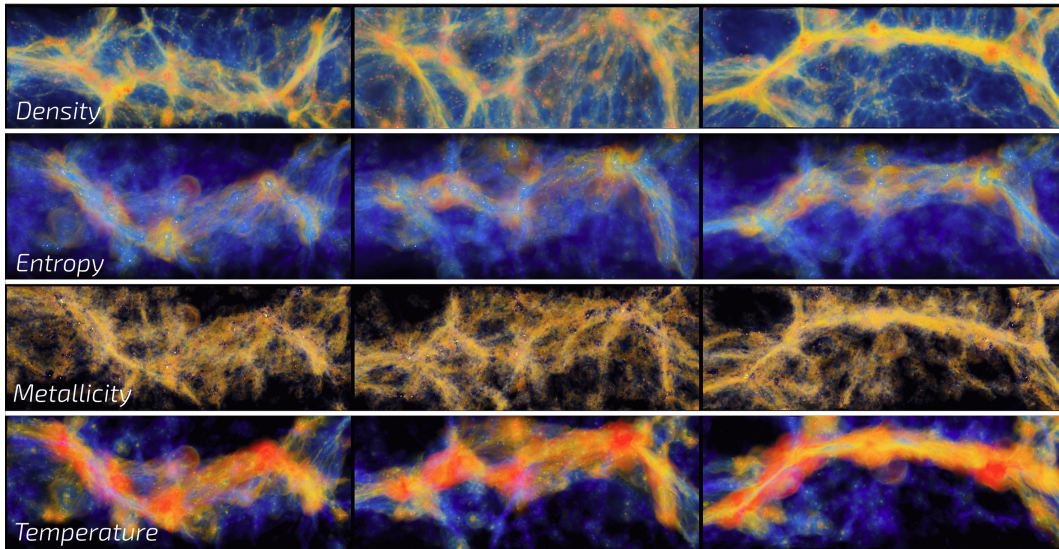


Figure 3.9: Example of a cosmic sheet found in the 100 Mpc *EAGLE* simulation shown from three different angles rotated around the same axis, arranged left to right. Different gas fields are presented in each row, from top to bottom: density, entropy, metallicity, and temperature. See Use Case 1 in Sec. 3.7.1 for more information about these cosmic structures.

3.7 Scientific Use Cases

In this section, we present three scientific use cases developed by our primary astronomy collaborators that represent real-world examples of lines of scientific inquiry that can be addressed using *CosmoVis*, and that highlight

how one might go about using the tool to answer these questions. *CosmoVis* is not limited to just these three lines of scientific inquiry, and could foreseeably be used to recreate a wide range of observational absorption line spectroscopy studies on simulated datasets. Note how these use cases are accomplished in the *CosmoVis* interface without the user needing to type any code, which greatly simplifies the traditional analysis workflow (see Fig. 3.3).

3.7.1 Case 1: Identifying and Examining Sheet Regions

A fundamental design motivation for *CosmoVis* was to provide an ability to easily explore the large volumes of cosmological hydrodynamical simulations to identify interesting cosmic structures [172]. These could be structures that are noted as interesting from a general, free-form exploration of the simulated universe or a targeted search for the structures with predetermined characteristics that could then be analyzed in greater detail. Here, we use *CosmoVis* to identify some cosmic sheets that can be followed up with simulations at zoom-in resolution. Mandelker et al. [181, 182] has analyzed zoom-in simulations of sheets to reveal that these are sites of important, observable hydrodynamical effects that elude coarser resolution simulations. This use case is directly inspired by those studies and will directly informing further similar investigations. We will also conduct some initial quantitative analysis on the sheet regions within *CosmoVis*.

When one initially loads *CosmoVis*, the cosmic web structure of the universe is readily apparent from the filamentary structures that contain the large majority of galaxies and intergalactic gas [64, 78, 90]. A somewhat closer inspection reveals that the not all of these structures are equal in their size and shape. Certain filaments are rather sparse with only a few galaxies and may appear to be merely offshoots of larger, more robust filaments that have much richer galaxy contents [98]. Furthermore, the initial visualization modality wherein the gas temperature is color coded reveals that the gas physical conditions can vary widely from depending on one’s location in the cosmic web, such as at the intersections of filaments (nodes).

Certain regions appear as plane-like structures where multiple filaments converge or appear to have formed from a larger coherent structure. Such a

‘cosmic sheet’ is shown in Fig. 3.9, and these regions have attracted heightened recent attention. As we discuss in **T1** (subsection 3.5.1), these sheet regions exhibit complex hydrodynamical processes as multiple cosmic structures coalesce. Generally, it is difficult to identify these sheet regions within a simulation, but *CosmoVis* offers a transformative improvement to this workflow.

For this use case, we employ the *EAGLE* 100 Mpc box visualized at 256^3 resolution [244] using the gas visualization modality with the temperature attribute. Larger simulated volumes by their nature contain greater numbers of the rarer, more massive structures, hence our choice of the largest volume. Sheets are plane-like structures with two primary dimensions, and are oriented in arbitrary directions. We proceed by using the slicing feature within the *Data Selection* panel and narrowing the thickness of the volume visualized in the x-direction to ~ 0.2 times the full volume (physically corresponding to 20 Mpc widths). The 3D rendering and interactivity of *CosmoVis* is critical for identifying sheets, as this slice of the simulated volume is inspected for sheet candidates by rotating it in several directions. Sheets are confirmed by finding (interactively) a camera orientation parallel to the structure. From this angle, the structure appears to collapse to one dimension. We log identified sheets by hovering over galaxies within the sheet and recording the approximate coordinates of the structure. We then view the next 20 Mpc slice of the simulation volume, inspect for sheet candidates, and proceed as above through the remaining volume in the x-direction. We then expand the slice to include the entire x-dimension and narrow the range to 20 Mpc in the y-direction, continuing the search as we did in the x-direction.

In all, we identified > 20 sheet candidates in the *EAGLE* 100 Mpc volume. Through the process of identifying sheets, we observed (from the colorization) that certain sheets seemed to contain much more high temperature gas ($T > 10^5$ K) than others. As a preliminary investigation of the temperature variation from sheet to sheet within individual sheets, we selected two sheets: one predominately filled with high temperature gas (orange-red in color) and another with predominately cool gas (blue-green in color). We oriented each sheet with the camera angle perpendicular its plane and placed three skewers through each using the functionality under the *Observe* panel

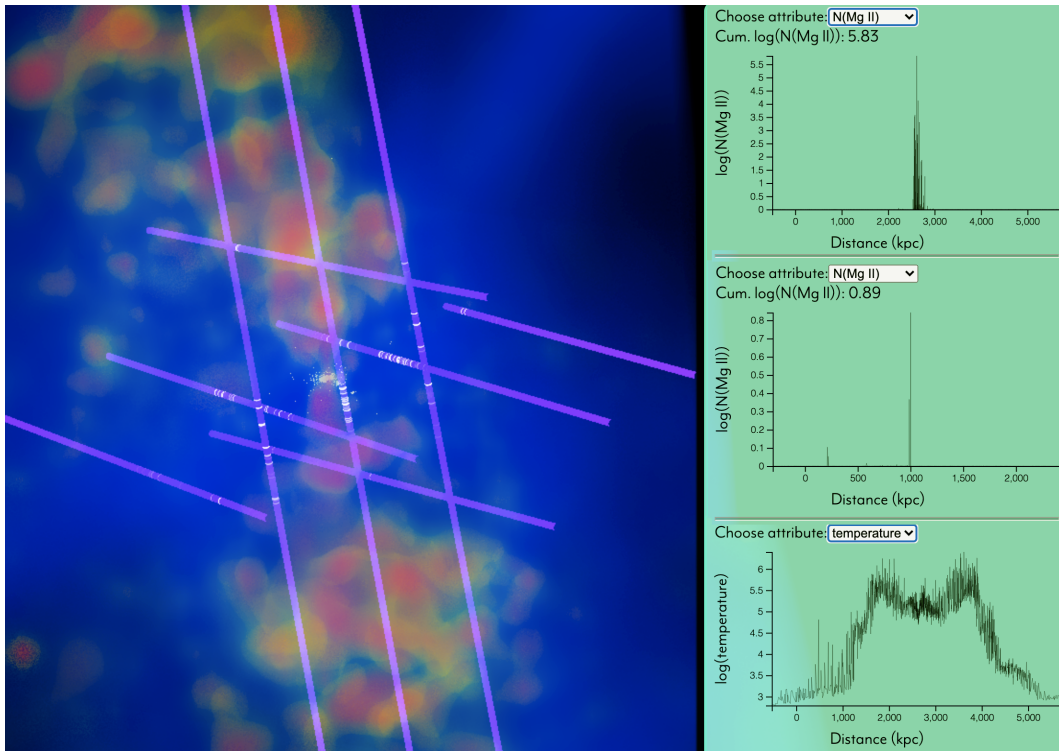


Figure 3.10: The CGM environment analyzed in Use Case 2 (Sec. 3.7.2). Here we have placed ten skewers through an apparent biconical galactic superwind to analyze its thermodynamic structure.

(see Sec. 3.6.3). Using the ‘request skewer attributes’ functionality, we inspected the temperature, density, metallicity, and $N(\text{H I})$ (neutral hydrogen column density) attributes. For the cooler-gas sheet, we found highly variable physical conditions along each skewer within the sheet. The cumulative $N(\text{H I})$ ranged from $\log N(\text{H I})/\text{cm}^{-2} = 11.77 - 15.25$. The gas probed by the highest column density skewer could be easily detected by routine Hubble Space Telescope observations, however the others would not be. The temperatures along the skewers varied from $\log T/\text{K} = 3.0 - 6.0$, with the higher temperature regions coinciding with density peaks along the skewer. We conclude that the intergalactic medium in such a sheet might be detectable only in small regions of enhanced density, whereas the entire structure is largely undetectable with current observational capabilities.

3.7.2 Case 2: Simulating a QSO absorption line study of the circumgalactic medium

Given the rarity of galaxies that have been measured using absorption line spectroscopy by multiple QSO probes of their CGM it can be difficult for astronomers to create a statistical model of the gas dynamics purely from observations, as discussed in **T2** (subsection 3.5.2). Therefore, a fundamental question haunts the interpretation of these studies: are hidden variables contaminating the statistical composite picture of the CGM owing to the selection of multiple galaxies with underlying different properties? In this use case, we attempt to address this question by performing an experiment that simply cannot be conducted in the real Universe with current instrumentation: by generating a suite of synthetic spectra probing individual galaxies at particular locations to test the intrinsic variance of absorption properties, i.e., column densities, within a single gaseous halo.

For our experiment, we identify a galaxy in a relatively isolated environment in the *EAGLE* 12 Mpc volume. The 12 Mpc box is ideal for this study because we seek a typical (not high-mass) galaxy that is fairly common in the Universe. Large statistical CGM studies are dominated by low-to-intermediate mass galaxies due to their prevalence among the galaxy population as a whole and it is in turn much more feasible to find bright QSOs (which are rare) to probe them. Many suitable candidates exist within the 12 Mpc volume. Our galaxy of interest is shown in Figure 3.10. Note the plumes of warm/hot gas extending to either sides of the galaxy. These result from gas outflows driven by supernovae and supermassive black hole activity [199, 287]. Clearly, the CGM of this galaxy contains inhomogeneities in all of these quantities. We first examine the temperature, density, entropy, and metallicity gas modalities within the ‘layers’ menu, noting that the hotter inner regions inside these ‘super-bubbles’ appear to be enhanced in metallicity, indicating that the outflow is clearly carrying enriched material. However, these same regions are *suppressed* in density. We then placed ten skewers through this region to investigate how the CGM substructure manifests in the sightlines. The skewers were placed to probe directly through both lobes of the plume structure, through the central regions the galaxy CGM, and its periphery. The skewers reveal quantitatively

the temperature distribution through the gas plumes, varying from $T \sim 10^6$ in one plume to 10^5 K in the central regions of the galaxy and back to 10^6 K in the other plume. The column densities of H I and other ions in these skewers are quite low, with $\log N(\text{H I})/\text{cm}^{-1} = 11.8\text{-}13.3$ and $\log N(\text{O VI})/\text{cm}^{-1} = 11.3\text{-}12.5$. Typical values from the CGM literature are $\log N(\text{H I})/\text{cm}^{-1} > 14.5$ and $\log N(\text{O VI})/\text{cm}^{-1} > 14.0$ [154, 280, 281]. These low column densities might indicate that the hot winds from the galaxy are simultaneously sweeping out the material and ionizing it to states beyond those we are measuring, e.g., to O VII or O VIII. Indeed, checking the column densities of those species in *CosmoVis* shows that they well exceed that of O VI.

3.7.3 Case 3: Metal enrichment of the intergalactic medium

As discussed in **T3** (Sec. 3.5.3), metal absorption signatures have been long detected in the spectra of quasar sightlines, and absorption line surveys of the circumgalactic medium have revealed without a doubt that galaxy halos are ‘enriched’ with metals [63, 94], and nearly all of the metal line systems detected to date arise from circumgalactic environments [224]. Posed differently, the question is “Have we yet detected truly IGM metal absorbers?”

In this use case, we turn *CosmoVis* to this question by using a combination of the 3D large-scale visualization and the virtual skewers to obtain column density measurements through various structures. Two main factors play into whether a metal absorption line system will be imprinted on a spectrum: 1) the medium must be metal enriched and 2) the medium must be dense enough to contain enough of a given species to leave a detectable absorption line.

Our general procedure as follows: We visualize the metallicity attribute of the gas layer and search for regions of the simulation volume that have relatively high metallicity values but that are well separated from the stars in galaxies. Then, visualizing the gas density layer of the same regions, we identify the highest-density subregions of the high-metallicity intergalactic environments. Using the skewer tool, we place skewers through this region, minding the extent of the skewer so that it only probes the region of question (using the sliders in the ‘observe’ window). Lastly, we retrieve the skewer attributes and scan the total column densities for a ions commonly detected in the literature,

including Mg II, O VI, and C IV.

Upon examining all simulation volumes, metal-enriched regions of intergalactic filaments are ubiquitous. Contrary to expectation, we find many regions immediately surrounding galaxies devoid of enriched gas. Similar to the scenario observed in Use Case 2, it is likely that these environments have had their gas evacuated by strong galactic outflows. In the *EAGLE* simulations, this phenomenon is largely driven by black hole feedback and can clear a galaxy’s CGM, eventually causing it to stop forming new stars [206]. Fig. 3.11 shows a large-scale view of the metallicity distribution in a slice of the *EAGLE* 25 Mpc volume, where several of these metal-poor cavities are evident. Identifying several candidate regions that might produce observable intergalactic metal absorption, we place skewers and inspect their ion column densities. We find no sightlines where, according to the simulation, we should detect metal line absorbers in truly intergalactic space (i.e., not in the CGM).

3.8 Evaluation

In addition to evaluating *CosmoVis* through the scientific use cases presented above, we gathered extensive feedback when presenting the software and describing its capabilities at a number of conferences and workshops. In October 2020, we presented a visual analysis based on a prototype of *CosmoVis* to the *IEEE VisAstro Workshop* as part of their Data Challenge, which sought novel visualizations of relationships between galaxies and the cosmic web in order to gain new insight into the physical processes that shape galaxies across cosmic time [76]. In December 2020, we introduced *CosmoVis* at the *RHyTHM: Research using yt Highlights* Meeting. In February 2021, we showcased the *CosmoVis* software and discussed selected use cases as part of the *Workshop on the Fundamentals of Gaseous Halos*, organized by the Kavli Institute for Theoretical Physics (KITP). At the KITP workshop, 172 astrophysicists attended our software demonstration, and 47 of them accepted our invitation to attend a more in-depth exploration of *CosmoVis*. Feedback from these presentations was very positive, with many researchers interested in learning how they could use *CosmoVis* to investigate research questions using

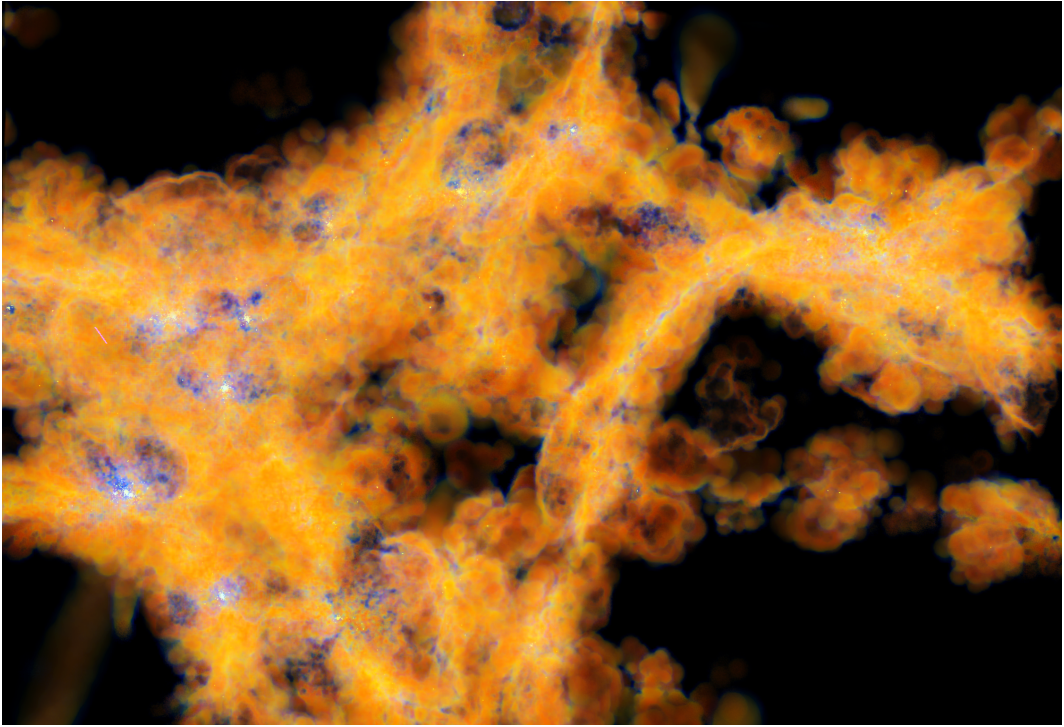


Figure 3.11: The large-scale metallicity distribution in the *EAGLE* 25 Mpc volume. This is a ~ 5 Mpc slice of the cosmic web, revealing that much of the intergalactic medium in cosmic web filaments is highly enriched with heavy elements (low to high metallicity colored blue to red). Also seen in here are a number of regions in the immediate vicinities of galaxies likely evacuated of their metal-rich gas due to galactic winds [206].

their own datasets.

Additionally, eight researchers recruited from the KITP workshop provided us with detailed qualitative feedback to an open-ended study. We invited participants to use *CosmoVis* to complete specific tasks, soliciting information about the effectiveness of our visualization and interaction techniques for a range of analysis tasks. All participants were members of astrophysics labs engaged in research involving simulation datasets, and included one professor, three PhD students, one Master’s student, and three undergraduate students. Surprisingly, the majority of the participants (five out of eight) indicated that they normally did not use any visualization software to conduct their research. Two of the remaining participants mentioned using custom python and/or Matlab scripts to plot data. One researcher uses a range of tools, including Jupyter notebooks, *yt*, and *Pynbody* [220]. Prior to attempting the tasks, participants had watched a live software demonstration that illustrated the

various elements of the *CosmoVis* user interface, and were given access to a recording of this. The tasks were chosen to represent a range of analysis activities that we expected to be familiar to workshop attendees, but difficult to carry out using existing software tools. They included the following: 1) Identify a region with ‘warm-hot’ 10^5 - 10^6 K gas; 2) Identify a region of high metallicity; 3) Find a signature of galactic winds; and 4) Use the virtual skewer tool to measure a distribution of temperature and ion species.

All participants were able to carry out the first two tasks, but two participants explained that they were not sufficiently comfortable with galactic winds research to carry out the third task, and another two participants had trouble controlling the skewers when attempting to complete the fourth task. After finishing the tasks, we asked users to provide feedback on their experience and to rate their interest in using *CosmoVis* for different activities. Six users indicated that they were very likely to incorporate *CosmoVis* in their own research workflow as well to use it as an exploratory tool for investigating simulation datasets, and all eight users believed that it would be a useful platform for outreach and disseminating results. Furthermore, all users indicated that they believed *CosmoVis* would be useful as an educational tool in various pedagogical contexts or for public outreach.

The participants commented positively on their experience using *CosmoVis*, each mentioning the overall responsiveness of the application even when working with large simulation datasets. We were especially interested in users’ reactions to using skewers within the simulation volumes. Two participants told us that they found the skewer functionality to be a little confusing at first, with one of them specifically mentioning that the interface made it easy to add additional skewers accidentally. Overall, however, the participants agreed that it was very beneficial to be able to generate spectra on-the-fly using the skewers. Three of the participants mentioned the importance of more extensive documentation (i.e., beyond on the video tutorial) to help users get more familiar with probing the simulations via the skewer functionality.

Most excitingly, users reported having new insights into their own research, even after working with *CosmoVis* for only a few hours. One participant expressed surprise when noticing that IGM filaments had a higher metallicity

than expected. Another was intrigued when they noticed that the hottest gases in the cube were found primarily at their edges. Yet another told us they found it interesting to see how individual stars were embedded in the clouds of metals. One participant told us that they looked forward to using *CosmoVis* in a current research project, explaining how they could aggregate multiple sightlines and compare them to ionization-modeled quantities from observed spectra. Multiple participants indicated that *CosmoVis* would be useful for their own data explorations, highlighting the importance of first getting an overview of their simulation snapshot and then being able to more thoroughly analyzing various data elements on demand.

Participants also expressed interest in having additional features made available that could support a range of use cases, such as: visualizing velocity fields to help identify superbubbles; incorporating custom data fields specific to their own research; adding additional sampling methods to compliment the skewers, such as radial shells; and providing more precise controls for some of the interactive sliders. Based on this feedback, we are planning to incorporate some of these ideas into future updates of *CosmoVis*, and we are currently in the process of revising the documentation in order to better support new users.

3.9 Conclusion

In this paper we have introduced *CosmoVis*, a novel interactive visualization software application for rendering and analyzing large-volume and zoom-in cosmological simulation datasets. We presented a series of scientific use cases demonstrating that *CosmoVis* is a useful tool for supporting a range of analysis tasks that enable new approaches to investigate a range of astrophysical phenomena, including identifying sheet regions, simulating a QSO absorption line to analyze the circumgalactic medium, and exploring metal enrichment within the intergalactic medium. We described the positive reception of *CosmoVis* by different communities of astrophysicists, and provided initial detailed feedback from eight domain experts at different career levels.

Future work will incorporate additional simulation datasets, and simplify

the pipeline for ingesting custom datasets. Additionally, while *CosmoVis* enables plotting skewer column densities and synthetic spectra, we plan to incorporate additional analysis functionality, such as generating equivalent width measurements from the synthetic spectra. We also are currently in the process of generalizing our representation of star particles to provide more information about the specific subhalos or galaxies in which they reside, and to explore rendering techniques such as ambient occlusion to make particle and volume depth cues more apparent. Another area for future exploration is extending *CosmoVis* to animate the evolution of simulations across a range of redshift snapshots.

CosmoVis is available via our open source GitHub code repository at <https://github.com/CreativeCodingLab/CosmoVis>, along with source code, detailed instructions on how to set up a custom server and load in custom datasets, and additional documentation.

3.10 Supplemental Material

This section includes additional material: four tables, each containing additional technical information about the *CosmoVis* software or the cosmological simulation datasets that it visualizes. Tables 1 and 2 specify details about the simulation data products available in *CosmoVis*, and Tables 3 and 4 provide performance metrics across different client and server configurations. *CosmoVis* is a scalable data visualization web application co-created by visualization designers, computer graphics researchers, and astrophysicists interested in making large-scale hydrodynamic cosmological simulations of the Universe more accessible to experts and non-experts alike. By streamlining the data analysis pipeline through a unified interface, users can interactively access the results of different simulations, view a myriad of particle fields, take quantitative measurements, and share their discoveries.

Table 1 lists the large number of different physical parameters that are encoded in different large scale hydrodynamic simulation suites that are currently supported by *CosmoVis*. This table is organized by particle type (gas, dark matter, star, and black hole), followed by the physical fields available for

each type, and finally the availability of these fields across *Illustris*, *EAGLE*, and *IllustrisTNG* simulations. See Sect. 2 of the main article for more details about these simulation datasets. Sects. 5.1 and 5.2 of the main article provide further information about how we preprocess and visualize these parameters.

In addition to the interactive 3D visualization, *CosmoVis* supports exploration of these simulations via virtual skewers that can be placed by users throughout the volume, acting as probes along a particular sightlines. Table 2 provides an overview of the different data products that can be retrieved via these skewers. They include: (1) *physical properties* along a skewer’s sightline (such as temperature, density, entropy, and metallicity), (2) dozens of different *ion column densities* (indicating the distributions of various metals), and (3) *synthetic absorption line spectra*. For more information about these data products and how the skewers are implemented and used in *CosmoVis*, see Sects. 5.3 and 6.2 and Figs. 1, 6, and 10 in the main article.

We designed *CosmoVis* following a client-server architecture to ensure its accessibility to users with a variety of computing hardware. Visualization textures are fetched from the server on demand and rendered directly in the client’s web browser. Data products related to the skewers are computed in the cloud and then returned to the web application. For both our development purposes and to support users with access to a high performance computing cluster, *CosmoVis* can be run locally from a single Python script and accessed on the local network. While this works on a small scale, *CosmoVis* has also been deployed using Amazon Web Services Elastic Compute Cloud (AWS EC2), and, more recently, using the Pacific Research Platform’s Nautilus compute cluster, which makes high performance computers freely available to academic researchers. For more information about the software architecture, see Sects. 5 and 5.5, as well as Figs. 3 and 7 in the main article.

Table 3 contains performance metrics about the initial load time and performance of the *CosmoVis* application. Here we show our tests across two deployments (locally on a Windows 10 PC and on the Nautilus Pacific Research Platform) accessed via three different machines (a consumer MacBook Air (2017), a M1 Macbook Pro (2020), and a custom built Windows 10 PC containing a 12 GB NVIDIA Titan Xp GPU). Within a few seconds of startup

time, the application can be immediately interacted with and runs at 60 fps on each of the platforms, including on the consumer laptop.

Table 4 compares the performance (load times, and framerates) across computers (the same one as listed in Table 3), the visualization volume resolution (128^3 , 256^3 and 512^3), and the deployment method (local vs. cloud). Each machine is able to render the 128^3 and 256^3 resolution volumes, however, the Macbook Air is unable to render the 512^3 volume successfully.

Table 3.1: Example particle types and fields made available in cosmological simulation snapshots that can be retrieved and plotted using *CosmoVis*. Simulation datasets typically organize their data in terms of different particle types (gas, dark matter, stars, and black holes), each expressing a variety of physical quantity fields.

Particle Type	Fields	Simulations		
		Illustris (2013)	EAGLE (2017)	IllustrisTNG (2018)
<i>Gas</i>	AGN Radiation (Bolometric intensity)	x		x
	Center Of Mass			x
	Cooling Rate	x		x
	Coordinates	x	x	x
	Density	x	x	x
	Electron Abundance	x		x
	Element/Metal Abundances		x	x
	Energy Dissipation			x
	Entropy		x	
	Expansion Factor at Maximum Temperature		x	
	Gravitational Potential Energy	x		x
	Host Halo Mass		x	x

Particle Type	Fields	Simulations		
		Illustris (2013)	EAGLE (2017)	IllustrisTNG (2018)
	Internal Energy	x	x	x
	Iron Mass Frac From SNIa		x	
	Mach Number			x
	Magnetic Field			x
	Magnetic Field Divergence			x
	Mass	x	x	x
	Maximum Temperature		x	
	Metal Mass Frac From AGB, SNII and SNIa Stars		x	
	Metallicity	x	x	x
	Metals Tagged			x
	Neutral Hydrogen Abundance	x		x
	Star Formation Rate	x	x	x
	Subfind DM Density			x
	Subfind Density, HSML, and Velocity Dispersion	x		x
	Temperature		x	
	Total Mass From AGB, SNII and SNIa Stars		x	
	Velocity	x	x	x
	Volume	x		
	Wind Dark Matter Velocity Dispersion	x		x
<i>Dark matter</i>	Coordinates	x	x	x
	Potential	x		x
	Subfind DM Density			x
	Subfind Density, HSML, and Velocity Dispersion	x		x

Particle Type	Fields	Simulations		
		Illustris (2013)	EAGLE (2017)	IllustrisTNG (2018)
	Velocity	x	x	x
<i>Star particles</i>	Birth Density		x	
	Birth Position and Velocity			x
	Coordinates	x	x	x
	Element/Metal Abundances		x	x
	Expansion Factor at Maximum Temperature		x	
	Feedback Energy Fraction		x	
	Host Halo TVir Mass		x	
	Initial Mass	x	x	x
	Iron Mass Frac From SNIa		x	
	Mass	x	x	x
	Maximum Temperature		x	
	Metal Mass Frac From AGB, SNII and SNIa Stars		x	
	Metallicity	x	x	x
	Metals Tagged			x
	Number of Tracers	x		
	Potential	x		x
	Previous Stellar Enrichment		x	
	Stellar Enrichment Counter		x	
	Stellar Formation Time	x	x	x
	Stellar Photometrics	x		x
Stellar Hsml			x	
Subfind DM Density			x	
Subfind Density, HSML, and Velocity Dispersion	x		x	

Particle Type	Fields	Simulations		
		Illustris (2013)	EAGLE (2017)	IllustrisTNG (2018)
	Total Mass From AGB, SNII and SNIa Stars		x	
	Velocity	x	x	x
<i>Black holes</i>	Black Hole Mass	x	x	x
	Black Hole Mass Accretion Rate	x	x	x
	Bondi/Eddington Accretion Rate			x
	Coordinates	x	x	x
	Cumulative Thermal/Kinetic AGN Energy Injection	x		x
	Cumulative Accreted Mass	x	x	x
	Cumulative Number of BH Seeds Swallowed		x	
	Density	x	x	x
	Expansion Factor When BH last accreted another BH		x	
	Formation Time		x	
	Gravitational Potential	x		x
	Host Halo Mass	x	x	x
	Mean Magnetic Pressure			x
	Number of Black Hole Mergers	x		x
	Number of Tracers	x		
	Pressure	x	x	x
Sound Speed		x		
Surrounding Gas Velocity		x		

Particle Type	Fields	Simulations		
		Illustris (2013)	EAGLE (2017)	IllustrisTNG (2018)
	Subfind DM Density			x
	Subfind Density, HSML, and Velocity Dispersion	x		x
	Thermal Energy in QSO-Heated Bubbles	x		x
	Velocity	x	x	x

Table 3.2: Virtual skewer data products available in *CosmoVis*. *Physical properties* and *ion column densities* linearly align along skewers, and can be dynamically plotted as banding patterns directly onto the skewer, as well as a plot in the *Observe* panel. *Synthetic spectra* correspond to observational absorption line spectroscopy measurements taken with specialized telescope sensors.

Skewer Data Products	Details
<i>Physical Properties</i>	Temperature, density, entropy, and metallicity
<i>Ion Column Densities</i>	H (I, II), C (I, II, III, IV, V, VI), He (I, II, III), Mg (I, II, X), N (II, III, IV, V, VI, VII), Na (I, IX), Ne (III, IV, V, VI, VIII), O (I, II, III, IV, V, VI, VII, VIII), S (II, III, IV, V, VI), and Si (II, III, IV, XII)
<i>Synthetic Spectra</i>	Absorption line spectroscopy samples that emulate observational techniques

Table 3.3: Comparison of first contentful paint (s), time to interactive (s), and framerate (fps) across three different machines and two server configurations as measured using Google Lighthouse. The local server is running on a Windows 10 PC with an Intel Core i7-8086K CPU @ 4.00GHz, 32 GB Memory, and a 12 GB NVIDIA Titan Xp GPU, and was tested on the same network using a MacBook Pro (13-inch, Apple M1 chip, 2020, 16 GB Memory) and a MacBook Air (13-inch HD+, Intel 1.8 GHz Dual-Core Intel Core i5, 8 GB Memory, Intel HD Graphics 6000 1536 MB). The same machines were used to test the performance of *CosmoVis* deployed in an Ubuntu Docker container on the Nautilus distributed compute cluster. Running *CosmoVis* locally results in marginally faster initial load times than the Nautilus deployment. All machines support realtime framerates of 60fps for the smallest volume resolution (64^3).

Server	Machine	Volume Res.	First Content Paint (s)	Time to Interact (s)	Frame-rate (fps)
Local (Windows 10PC)	Windows 10 PC	64	1.2	1.3	60
	M1 MacBook Pro (2020)	64	1.2	1.2	60
	MacBook Air (2017)	64	1.6	2.5	60
Nautilus (Ubuntu Container)	Windows 10 PC	64	1.8	1.8	60
	M1 MacBook Pro (2020)	64	1.8	1.8	60
	MacBook Air (2017)	64	2.2	3.3	60

Table 3.4: Performance comparison of loading times and framerates for larger volume resolutions (128^3 , 256^3 and 512^3) across three different machines accessing two different server configurations. The Windows machine containing the dedicated GPU is able to run all default resolutions at interactive rates or better, and even the older consumer laptop with integrated graphics is able to support higher resolutions up to 256^3 .

Machine	Volume Resolution	Load Time (s)		Framerate (fps)	
		Local Server	Nautilus Cluster	Local Server	Nautilus Cluster
Windows 10 PC	128	0.78	3.11	60	60
	256	2.77	25.65	53	53
	512	22.74	174.28	35	33
M1 MacBook Pro (2020)	128	4.10	3.71	60	60
	256	12.99	19.38	43	38
	512	61.75	133.06	18	20
MacBook Air (2017)	128	3.01	3.49	30	30
	256	12.81	26.34	10	10
	512	-	-	-	-

Chapter 4

Visualizing the Temporal Evolution of Cosmological Simulations

4.1 Abstract

We introduce a temporal analysis extension to the cosmology simulation visualization tool, *Cosmo Vis*. This new functionality empowers astronomers to explore and analyze the temporal evolution of galaxies, halos, and the Cosmic Web's large-scale structure within cosmological simulations. With interactive 3D volume animations, astronomers can now make insightful comparisons across time, tracking individual galaxy halo mergers and examining synthetic absorption line spectra of galaxies at different stages of formation. We introduce a new interactive small multiples "film strip" view to accompany sequence data and help contextualize temporal changes between snapshots. To showcase the capabilities of this extension, we demonstrate a range of astronomy analysis tasks and provide detailed insights into the design and implementation of the tool. Furthermore, we include real-world Science Use Cases that highlight the practical value of this extension, along with a thorough performance evaluation and astronomer feedback.

4.2 Introduction

Hydrodynamic cosmological simulations are key to understanding the physical processes underlying the universe's evolution and empirical observations. These simulations are generated on high-performance computers using code that models the formation of cosmological features, ranging from the Cosmic Web's large-scale structure (LSS), black holes, and dark matter. The largest of these simulations contains upward of tens of millions of galaxy halos. They can range in size on the order of tens of megaparsecs (10 Mpc) to hundreds of Mpc (> 100 Mpc) and are referred to as big-box simulations. In contrast, zoom-in simulations may be a few hundred kiloparsecs across and contain information about a single halo.

Each snapshot from a given simulation contains the system's state for a single point in time. These are large multidimensional datasets containing a distribution of particle tracers that carry information about many gas physical conditions (such as mass, temperature, density, metallicity and much more), dark matter, stars, and black holes. The cosmological simulations this tool is designed to visualize typically have starting conditions defined around the end of the cosmological dark ages (see Table 4.1). As the name suggests, the universe was completely devoid of light during this period and composed mostly of homogeneous hydrogen gas before forming the first stars and galaxies. However, other simulations model the physics of the early universe or even the evolution of specific structures such as stars.

These simulation snapshots are ordered in terms of the cosmological redshift, or as commonly referred to simply as z . In observations, redshift is a measure of how light wavelengths "stretch out" and elongate as they traverse the expanding universe, shifting more "red" in the electromagnetic spectrum the longer they travel. In this context, redshift simply refers to the amount of time that has passed. Redshift can approximate the age of the material in the universe, such as galaxies, halos, and quasars. This is done by measuring how far the wavelengths of known ion emissions have shifted. A low redshift (i.e., $z = 0$) indicates "present day" or local conditions, whereas galaxies with

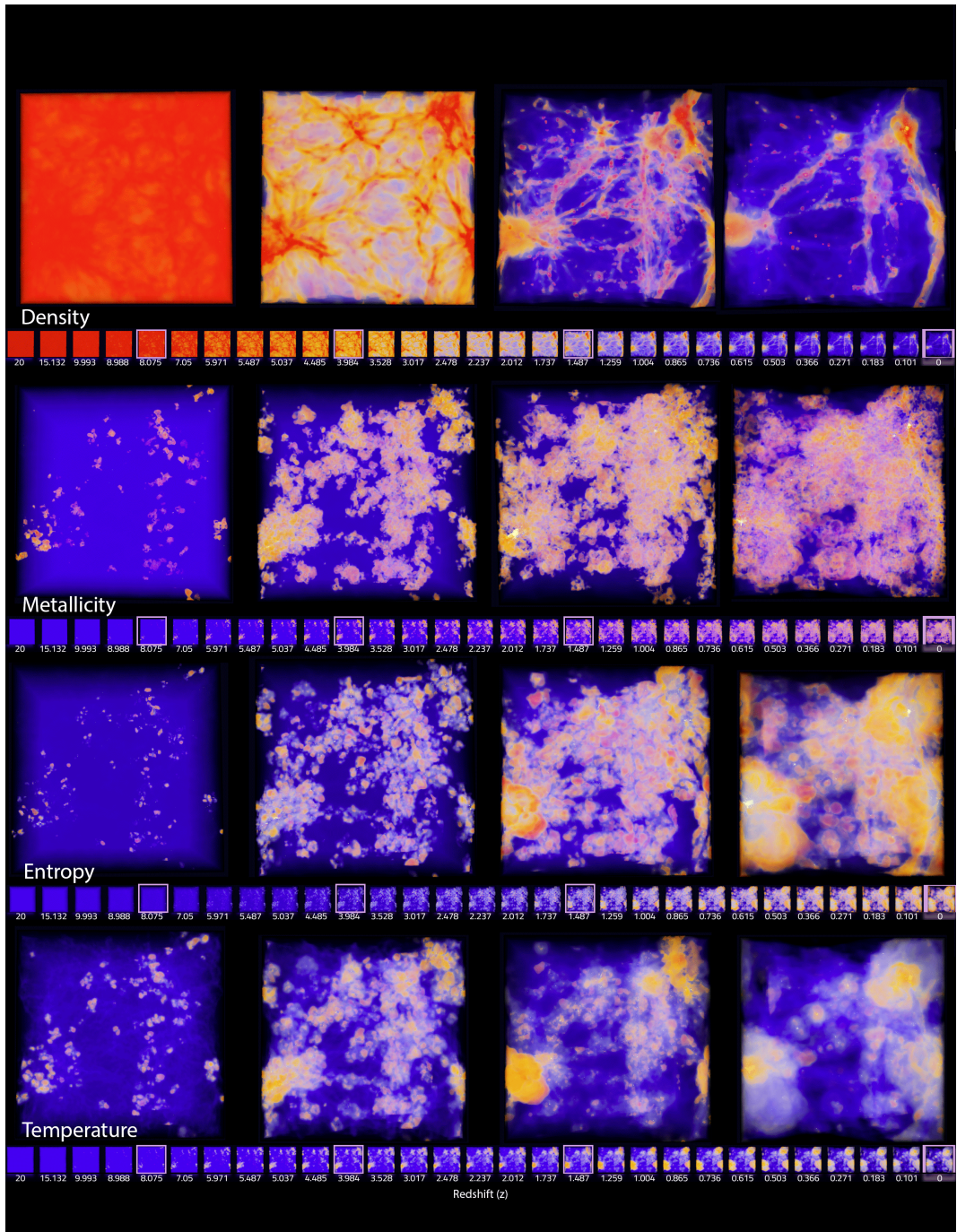


Figure 4.1: Fields are shown from the 12 Mpc EAGLE simulation from top to bottom: density, metallicity, entropy, and temperature. From left to right: snapshots arranged from high redshift ($z = 20$) to low redshift ($z = 0$). Four enlarged views ($z = 0, 1.487, 3.984,$ and 8.075) are shown above each film strip, with their thumbnail highlighted in the sequence. Each visualization uses the same blue (low value) to yellow (middle value) to red (high value) color scheme. However, the numbers the transfer function aligns to depend on the specific dataset and type of field. All thumbnails in the film strip have synchronized color scales.

a higher redshift (i.e., $z > 10$) were born closer to the start of the universe. These high-redshift objects are more difficult to observe and require sensors more sensitive to infrared light. Table 4.1 shows a chronology of cosmological epochs regarding redshift and years since the Big Bang.

Table 4.1: A chronological list of cosmological epochs regarding the universe’s age, redshift, and what galaxies instrumentation can detect.

Time			Detectable galaxies		
Epoch	Age of the Universe	Redshift (z)	HST	JWST	Sims
<i>Present Day</i>	13.8 Billion Years	0	x	x	x
<i>Reionization Ends</i>		6 [311] [104] [233]	x	x	x
<i>Oldest Detected Galaxies</i>	325 Million Years	13.2 [232]		x	x
<i>Reionization Begins</i>	200 Million Years	20		x	x
<i>Dark Ages</i>		20-1100			x
<i>Recombination</i>	370 Thousand Years	1100			
<i>Inflation</i>		1200			
<i>Big Bang</i>	0	1200			

Absorption line spectroscopy is one of the primary tools astronomers use to study the structure, composition, age, and physical properties of galaxies, stars, gas, and other structures in our universe, along with imaging and emission line spectroscopy. Using this technique, astronomers can measure the absorption profile of dozens of elemental ions, the redshift of intervening and absorbing gas (which can also indicate distances), and gas dynamics. This method dates back to the beginnings of photography when physical archives of star catalogs were first produced. Since then, observation equipment has had much higher

sensitivity to far infrared frequencies that allows the detection of high redshift galaxies. Observational astronomers use telescopes equipped with specialized sensors such as the Cosmic Origins Spectrograph (*COS*) on the Hubble Space Telescope and *NIRSpec* or *MIRI* on the James Webb Space Telescope to study the composition of galaxies and intervening material. By pointing the telescope at a bright distant light source such as a quasar, astronomers can measure different physical properties such as the source's velocity, temperature, and metal ion abundance. The spectrum can also capture intervening material (gas of the IGM/CGM) through scattering and absorption. While these spectra are essential for astronomers to study the universe, we are physically constrained with when and where we can take measurements. Galaxy surveys often capture one or a few spectra ($n = 1$) of many galaxies. This does not create an excellent statistical picture of what happens within galaxies but rather what the average galaxy is like. On the other hand, astrophysicists can use the same spectroscopic techniques to compare the results of simulations with observations of the universe.

Herein lies one significant difference between tracking galaxy evolution from observational techniques and through simulations. Given the short existence of humanity and observational astronomy versus the age of the universe, it is impossible to track the entire life of a single galaxy, let alone from any vantage point other than Earth. On the other hand, large-scale cosmological simulations contain the histories of multitudes of galaxies from the beginning of the universe to the present day. These simulation codes are seeded with initial physical conditions and tuning parameters that reflect our understanding of the conditions of the early universe before galaxies began to form. With each time step of the simulation, a snapshot is saved, which results in a series of large n-dimensional data files.

There are many strengths of studying simulations for cosmologists compared solely to observations. In simulations, you can look at a galaxy from any angle, not just from Earth. Each data point in a simulation contains information about many physical conditions (metal abundance, density, temperature, star formation rate, etc.), some of which are not directly observable. Table 4.2 shows a list of physical properties quantified in simulations that are

either not directly observable via a telescope or are inferred properties based on other observables. Simulations provide access to a wide range of galaxy and halo morphology, as well as the ability to track the history of galaxies between time points and combination events through a data structure called a *merger tree*. Astronomers can use simulation results to plan new galaxy surveys [290].

Table 4.2: A list of several physical properties that can be directly observed, estimated from observations, and available in cosmological simulation snapshots. Only a few measurements are directly observable in our universe (and limited to emission sources), whereas simulations contain multitudes of fields directly accessible and ready to visualize.

Physical Property	Directly Observable	Estimated Measure	Included in Simulations	Derivable from Simulation
<i>Column Density</i>		X		X
<i>Density</i>		X	X	
<i>Entropy</i>		X	X	
<i>Gravity</i>		X	X	
<i>Luminosity</i>	X		X	
<i>Ly alpha emission</i>	X			X
<i>Magnetic Field</i>		X	X	
<i>Mass</i>		X	X	
<i>Metallicity</i>		X	X	
<i>Pressure</i>		X	X	
<i>Star Formation Rate</i>		X	X	
<i>Temperature</i>		X	X	
<i>Velocity</i>		X	X	

This paper introduces an extension to *CosmoVis* that enables temporal

visualization and animation of cosmological simulation datasets. The main contributions of this paper are:

- a framework for animating cosmological simulation data;
- a query tool for filtering galaxies based on different properties (i.e., mass, star formation);
- zoom-in to individual galaxy halos and visually investigate temporal evolution and mergers;
- generates column densities and synthetic absorption line spectra across multiple time points.

In the following sections, we discuss related visualization and astronomy tools in section 4.3, introduce a set of science analysis tasks that this extension to *CosmoVis* enables in section 4.4, the design and implementation in section 4.5 and section 4.6, a set of relevant scientific use cases in section 4.7, followed by a conclusion in section 4.8.

4.3 Background and Related Work

Astronomy visualization is a rich area of research where many tools have been created to study both observational and simulation data. The survey by Lan et al. [164] identifies many state-of-the-art interactive astronomy visualization software created in the last decade and categorizes them based on data analysis tasks, visualization techniques, and topics in astronomy. This section focuses on relevant examples of time series visualization and temporal animation of astronomy data, emphasizing cosmological simulations. Compared to other topics in astronomy visualization, real-time animation and inspection of time-series simulation volume data have been relatively under-explored.

As discussed in the previous section, cosmological simulations are a powerful tool for studying the universe's evolution. Abramov et al. introduce the general purpose simulation visualization tool *CosmoVis* [35] that allows users to visually inspect individual snapshots, visualize gas, dark matter, and star attributes, place "virtual skewers" to create synthetic absorption line spectra and column densities, but before this release did not include animations be-

tween different snapshots. To study changes over time, other tools use merger trees to study galaxy evolution in cosmological simulations. Merger trees are a type of data structure used to keep track of the history of galaxy halos and their progenitors. Halos are gravitationally bound structures containing galaxies, and progenitors are the precursor galaxy halos from the previous snapshot that have merged over time into the current halo. There might be multiple, singular, or no progenitors for a given halo. With no progenitors, this might be a halo at the beginning of the simulation (e.g., $z = 20$) that has just reached a critical density and emerged from the homogeneous structure of the early universe. Merger trees can be drawn as a node-link diagram graph where the root node is a halo connected to one or more progenitor nodes from the previous snapshot at a higher redshift. Many different algorithms have been developed to identify, tag, and track individual galaxies and halos within simulations [161]. A common method for generating merger trees is the Friends of Friends (FoF) algorithm that identifies halos within the raw data. The Rockstar algorithm is another one that is commonly used. One challenge with comparing merger tree data is various methods and simulations may store this information differently. To address this issue, *ytree* [255] can be used for analyzing merger tree data from various simulations and formats, allowing for cross-comparison and graph visualization. *ytree* [255] is an extension to the Python module *yt* [282], a popular astronomy tool that supports loading, visualization and analysis of many different simulation codes.

Some examples of astronomy visualization tools utilize merger trees for tracking galaxy evolution. [118] [39] [139] [251] [222] Fritschi et al. [118] animated particle simulation data with trajectory trail annotations to model temporal evolution of the cosmic environment and halo mergers within the large-scale structure of the HACC simulation using *VTK* [247] and *Qt*. Preston et al. [222] used a distributed computing rendering technique in their cosmological simulation visualization tool that consists of three data views: 3D particle rendering, merger trees exploration, and quantitative plots. Almryde et al. [39] built a web application that animates thousands of different dark matter halo merger trees in the HACC simulation. Scherzinger et al. [245] also created an interactive exploration tool for dark matter cosmology simulations.

Work by Hazarika et al. [139] used multiple technologies (WebGL/Three.js, Paraview/Python/C++/vtk) to visualize changes in dark matter simulations using parallel coordinate plots and merger tree browsing. Shan et al. [251] use 2D lasso brushing to select halos in 3D space and explore them further by tracing their paths in 3D while watching the merger tree in 2D.

Another tool, TULIPS (Tool for Understanding the Lives, Interiors, and Physics of Stars) [165] is a visualization software written in python that works with one-dimensional MESA stellar evolution code [210]. This tool takes advantage of the spherical symmetry of the simulation data to create circular, 2D plots and animations to showcase the properties, such as mass, radius, and chemical profile, of stellar objects as they evolve. Haroz et al. [136] introduced a visualization technique for measuring time variant uncertainties in cosmological particle simulations.

While this section focuses on simulations, there are a few relevant examples of time-series analysis visualization of observational data. TimeTubes [121] [241] [242] is a software that is used to monitor variance in quasars over time. Montage [57] is a tool that can be used to animate observation data cubes. The Gaia sky catalog [239] visualizes real-time star movements in their visualization tool. [219]. NASA World Wind is an SDK that enables researchers to build temporal visualization models around Earth [217], and the Globe Browsing project [58] built in OpenSpace [62] allows for spatial-temporal visualization of other celestial bodies such as Mars. Polyphorm [110] is a visualization tool for Cosmic Web reconstruction that uses a Monte Carlo Physarum Machine (MCPM) computational model (inspired by the growth of *Physarum polycephalum* slime mold) to reconstruct transport networks from sparse 2D and 3D data, such as simulated and observational cosmological datasets.

Outside of astronomy, other domains use visualization of large n-dimensional datasets—for example, weather models and medical imaging. In the realm of temporal medical data, TempoCave [306] is a VR tool that allows the comparison of brain connectomes before and after psychological treatment. Rudenko et al. summarize tools and methods for visualizing 4D meteorological data [236]. Chen et al.'s [96] survey further contextualizes time-domain visualizations into three specific categories, which this tool fits into: simulation space explo-

ration (with *glyph-based visual mapping*), parameter visual exploration (with *focus+context exploration*, *overview-to-detail exploration*, and *visual steering*), and feature definition, extraction and tracking in simulation data (*data sampling and reduction in feature space* and *interactive feature exploration*).

Regardless of domain, creating compelling and meaningful visualizations from temporal data is a challenge in of itself, let alone adding spatial and volumetric constraints to the mix. One of the issues is that we perceive time as it passes by us – we cannot stop time and look at the continuum of everything that has happened [262]. One obvious solution to visualize temporal data in a format fit for human consumption is through animation. However, this is not without its own critiques. As Streit and Gehlenborg note, while animation can be aid in visual comprehension, it can be difficult "to detect recurring patterns and compare across multiple time points" [262]. In this extension to *CosmoVis*, this challenge of comparing temporal changes between time points is mitigated through a multifaceted approach. For starters, the introduction of a linear, small-multiples "film-strip" view, as introduced in section 4.5, users can watch and interact with a 3D animation, but also view each frame in the context of the entire sequence. Additionally, it acts as a way of quick navigation between time points, whereby cross-comparisons can be swiftly made.

4.4 Analysis Tasks

Large-scale cosmological simulations are a powerful tool for validating assumptions and physical theories about the universe. One of the ultimate goals of creating these simulations is to verify physical theories underlying the universe's structure and test the results based on empirical observations. While the predictions made by simulations can also be used to drive new star and galaxy surveys, simulations are simply important tools for interpreting observational data and gaining physical insight into measurements. While simulations are not perfect recreations of the universe (they make certain assumptions about the underlying physics, have finite resolution, particle tracers are not as granular as the name suggests), they can be a helpful tool for contextualizing measurements and finding patterns. One of the strengths of simulations

is the temporal aspect, whereby the mechanisms underlying galaxy formation and evolution can be studied in a controlled environment. In this section, we introduce the following temporal cosmology analysis tasks we identified in collaboration with astronomers:

- **Task 1: Identify key phases of structure formation**

When do structures such as galaxies, voids, and filaments emerge from the cosmic medium?

How do they evolve?

- **Task 2: Track galactic life cycles**

How do galaxies evolve, and what affects the availability of fuel for sustaining star formation?

How do older galaxies compare to younger galaxies?

- **Task 3: Compare galaxy evolution across environments**

How does the environment around a galaxy impact its evolution and physical properties?

What impact do galactic inflows and outflows have on the composition of the nearby intergalactic medium?

- **Task 4: Examine galaxy morphology transformation**

How does galaxy shape and geometry changes over time as they age and interact with their neighbors?

Are there any correlations between physical properties such as mass, temperature, or metallicity with morphology?

This extension to *CosmoVis* enables these tasks and more for comparing simulations across multiple redshifts. Several usage scenarios of this tool can be found in section 4.7.

4.4.1 T1: Identify key phases of structure formation

Through observations, astronomers can identify structures such as galaxies, halos, voids, and filaments. Still, it isn't easy to fully know when they formed

or how an individual galaxy has changed over its lifetime. By looking at time-series visualizations of cosmological simulations, astronomers can identify different stages of evolution and when specific structures emerge.

Soon after the Big Bang and continuing through the inflation period and dark ages, the universe was non-linearly expanding. The structure of the early universe was primarily homogenous, although quantum fluctuations became amplified during cosmic inflation. The gravitational force pulled matter together, forming slightly denser clumps, which eventually grew larger more rapidly. At some point, the collapse of matter into halos becomes nonlinear, characterized by the formation of large-scale structures such as clusters of galaxies, filaments, and voids. Key epochs for studying early galaxies include the reionization of Hydrogen ($z = 5 - 7$) and Helium ($z = 2 - 3$). During these epochs, the first stars and galaxies formed and emitted enough radiation to ionize the neutral gas in the universe, making it transparent to light. Astronomers can observe the virialization phase in the more recent universe, where galaxy clusters are fully formed. At this stage, the gravitational potential energy of the matter in the cluster is so high that it exceeds the kinetic energy of the particles, causing them to become trapped in the cluster. As the gas in the cluster becomes denser, it heats up, leading to the emission of X-rays. Another visually identifiable temporal feature is high temperature and high entropy shock bubble formation around massive halos and intergalactic medium filaments.

4.4.2 T2: Track galactic life cycles

Galaxies are complex, dynamic structures that vary greatly in morphology, properties, and composition. As the universe ages, galaxies emerge, grow, recirculate material, sometimes combine, or become quiescent and run out of the fuel required for star formation. Galaxies that emerged towards the universe's beginning may have vastly different characteristics than later low redshift galaxies, which have more "recycled" material and heavier metal elements. For instance, the early universe was composed of cold, low-entropy non-metallic gas (i.e., devoid of elements heavier than hydrogen helium). Cold, dense pockets in the medium of the early universe are where the first stars and

galaxies emerged. These early stars began to warm the medium of the cosmos, emitting radiation into the universe. Only after the first stars formed and died did heavier elements (metals in this context) become prevalent in the universe. Tracking galaxies to earlier snapshots allows astronomers to compare the shape of high and low redshift galaxies. Astronomers present a statistical picture of galaxy evolution by comparing galaxies from different cosmological epochs. The limitation of this sampling is that it does not provide information about how a single galaxy has formed and evolved, only an average of all the galaxies in the universe across a range of redshifts (albeit with a sampling bias towards low-redshift). With simulations, galaxy halos can be tracked over their entire life span by traversing the merger tree, and we can also monitor the evolution of their physical properties. Are there features (shape, mass, morphology, composition, temperature) of young, high redshift galaxies that correlate with the present-day epoch ($z = 0$)? By investigating changes between snapshots, astronomers can better understand the progenitors of certain types of objects and track the salient physical processes throughout their transformation.

While simulations can somewhat accurately physically model galactic systems, utilizing simulations to make meaningful cross-comparisons with observation datasets (including absorption and emission spectra, optical imaging, etc.) is a non-trivial task. *Trident* [147] is an open-source post-processing tool for hydrodynamic simulation codes that mimics observational techniques from standard telescope instruments, returning absorption-line spectra along linear trajectories through simulation volumes. The data products from *Trident* can help correlate spectral features, such as absorption lines, physical gas properties (temperature, pressure, entropy, energy, etc.), and ion column density maps to simulated physical structures. Placing these synthetic sight-lines through the same feature in multiple sequential snapshots can be used to help find correlations that provide interesting markers for stages of galaxy evolution, and then compared back to observational data.

4.4.3 T3: Compare galaxy evolution across environments

The local environment in which a galaxy forms significantly impacts how it evolves. After the cosmic dark ages, the distribution of the universe was

mostly uniform. However, some regions were denser than others, leading to the formation of galaxy clusters, groups, and individual galaxies. In dense clusters, the proximity of neighboring galaxies can result in gravitational interactions such as galaxy mergers, whereby more massive galaxies are formed through combination, and tidal stripping, where a larger galaxy steals stars and other material from a smaller neighbor. These events can affect star formation and other processes that alter the galaxy's structure and properties. Through studying emission spectra lines and absorption profiles of galaxies in observational catalogs such as the Sloan Digital Sky Survey (SDSS), astronomers can compare galaxies in different environments, such as central ones in the core of large clusters, or further removed satellite galaxies [208] [158]. Through the use of *Trident* [147] and the virtual skewers introduced in chapter 3, the temporal effects can be measured visually and analytically by probing galaxies in different environments (i.e., sparse or dense) through multiple snapshots.

Environment plays a crucial role in regulating the gas supply required for star formation. Observable galaxies exhibit various star formation rates (SFRs) that environmental factors can impact. Some galaxies may exhibit no active star formation ($0 M_{star}/yr$), while others might produce the equivalent of one thousand suns per year [156]. Throughout a galaxy's life, it might use all the gas sustaining its star formation, and in others, abundant gas supply in their environment allows them to maintain high SFRs. Galaxies in dense environments may experience a more hostile intergalactic medium, which can prevent gas from flowing into the galaxy and lead to a lower SFR.

Galaxies can "interact" with their environment by transporting material through active inflows and outflows. Inflows allow for the accretion of material into the galaxy and provide fuel for star formation. Galaxies at the center of dense clusters can exhibit large cool gas inflows enriched with metals while having low SFR. Galaxies can also exhibit powerful outflows due to AGN or supernovae activity, which can profoundly impact the surrounding environment. Some galaxies have an active galactic nucleus (AGN) powered by supermassive black holes that spew material from their core in powerful outflows of gas and energy. Outflows remove gas and other materials from the galaxy, which can be recycled back into the galaxy's circumgalactic medium (CGM)

or disperse and become part of the intergalactic medium (IGM). Outflows can impact the evolution of galaxies by limiting the amount of gas available for future star formation and altering the chemical composition of the remaining gas. These outflows can alter the composition of the IGM through metal enrichment, where heavy elements formed in stars, such as carbon, nitrogen, and oxygen, are spread outside the galaxy. In dense environments such as galaxy clusters, inflow and outflow interactions between galaxies can lead to mixing gas with different metallicities, enhancing metal enrichment within the galaxy and the surrounding intergalactic medium. Overall, the environment can significantly impact galaxy evolution, and understanding these effects is crucial to understanding the formation and evolution of galaxies in the universe.

4.4.4 T4: Examine galaxy morphology transformation

Galaxies exist on a continuum of various shapes, structures, and sizes called their morphology. Many galaxy morphologies have been identified, such as spiral, S0, elliptical (strong relationship to stellar mass and halo mass) [300], spheroidal, irregular, lenticular, and dwarf galaxies. Galaxy morphology can be correlated to certain measurable features, such as their size (more giant galaxies may have more complex structures), luminosity (brighter galaxies tend to produce more stars), color (red vs. blue, quiescent vs. star-forming), black hole mass, rotation, and velocity [86]. Many factors, including mergers, AGN activity, and interactions between the IGM and CGM can influence the evolution of galaxy morphology. Observations suggest that galaxies have evolved significantly since the early universe. In the early universe, galaxies were much smaller and less massive than today. They also formed stars at a much higher rate than they do now. As the universe aged, galaxies grew in size and mass through mergers with other galaxies, which can dramatically alter their morphology.

Simulations help predict that some morphologies, such as elliptical galaxies, arise due to galaxy mergers [300]. Spiral galaxies, on the other hand, are thought to form from the collapse of a rotating gas cloud, which leads to the formation of a disk structure with spiral arms. The interplay between star formation and feedback processes also shapes galaxy morphology. For example,

star formation can drive the growth of a galaxy’s disk. At the same time, feedback from supernovae explosions and active galactic nuclei can regulate SFR and redistribute gas and stars within a galaxy. In addition to these physical processes, the morphology of galaxies is also influenced by the environment in which they reside. Galaxies in dense clusters, for example, are more likely to be elliptical due to the higher frequency of mergers and interactions. Overall, the evolution of galaxy morphology is a complex process influenced by various physical processes and environmental factors. Understanding this evolution is a major area of research in modern astronomy, as it can provide insight into the formation and development of the universe. Simulation visualizations make it easier to discern galaxy and halo shapes when given the full range of camera motion around a celestial object. In contrast, we can only observe other galaxies from a small range of angles around Earth at a single point in time (in cosmological timescales).

4.5 Application Design

This temporal analysis extension to *CosmoVis* [35] enables users to visually interact with, inspect, and animate cosmological simulations animated across multiple redshift snapshots. In this section, we outline the design of several key features, which together enable the four analysis tasks introduced in section 4.4. The main features introduced here include real-time playback of large-scale cosmological simulation animations, a linked small multiples "film strip" of each snapshot and/or halo merger tree, a galaxy query tool with linked brushing, and zooming into galaxies and halos from big-box simulations. Users can visualize multiple fields from each snapshot through a hybrid volume and particle visualization, such as gas, dark matter, and star properties. Changes to the transfer function are applied to each snapshot synchronously to aid in visual comparisons. These design decisions, taken together, reduce the cognitive load users require to understand these complex, volumetric, time-series datasets. The "virtual skewers" introduced in *CosmoVis* [35] have also been updated to operate on multiple snapshots loaded in a sequence, enabling further exploration. Examples of the skewer tool used across snapshots are shown

in Figure 4.4, Figure 4.5, and the science use cases in section 4.7.

4.5.1 Temporal Playback

The temporal playback panel, shown in figure [add figure], directly enables the user to accomplish the first two analysis tasks introduced in the previous section: (T1) *identify critical phases of structure formation* and (T2) *track galaxy life cycles*. In the user interface, there is an option to investigate a single snapshot from the dropdown or to load in a sequence of snapshots. When investigating a single snapshot, the temporal playback panel is hidden and becomes visible when the animation mode is selected. This timeline panel includes several buttons for navigating between snapshots: start, stop, previous and next frame. Users can adjust the playback speed of the animation from 1 to 60 frames per second. Users can visually inspect and notice when particular structures emerge or dissipate by manually switching between frames, watching an animation, or glancing at the film strip. The temporal sampling and the time between each snapshot are determined by the simulation suite that the data comes from. For example, the EAGLE simulations provide twenty-eight snapshots, whereas Illustris-TNG provides one hundred. Users can playback and watch the entire simulation evolve or zoom in and track an individual galaxy halo sequence using the galaxy query panel. When users inspect an individual galaxy, they can track its evolutionary history and how its features have changed.

When the temporal playback mode is activated, a film strip panel appears at the bottom of the screen, as shown in Figure 4.1, Figure 4.3, and Figure 4.4. This panel contains a series of horizontally arranged linked-view thumbnails showing each animation frame in the simulation. The thumbnails are arranged from high redshift on the left to low redshift on the right. As the user skips between frames or watches the animation, the background of the thumbnail of the current snapshot becomes highlighted. Additionally, the camera angle, color transfer function, and clipping planes for each thumbnail are linked to the main visualization and are updated in real-time. The film strip feature helps astronomers complete the four analysis tasks we outlined in the previous section: (T1) *identify critical phases of structure formation*, (T2) *track galaxy life*

cycles, (T3) compare galaxy evolution across environments, and (T4) examine galaxy morphology transformation.

4.5.2 Galaxy Query

The galaxy query tool helps astronomers quickly explore the dataset and locate interesting galaxies to study. The query panel allows users to search for galaxies within a snapshot based on physical parameters: halo mass, stellar mass, star formation rate, gas mass, and black hole mass, as shown in Figure 4.2. Users can toggle a checkbox on and off for each field to enable filtering of that attribute. Additionally, users can select and query a range of values using a double-ended slider. This allows, for example, focusing on only very massive galaxies or ones with high star formation rates. When one or more filters are selected, the query results panel populates a table with a list of galaxies, their host halo ID, and mass. Clicking on the Galaxy ID immediately updates the view such that the big box is sliced down to show just that galaxy. Clicking on the halo ID loads a higher-resolution version of the halo for further investigation.

This query feature helps accomplish the four analysis tasks we introduced in the previous section: (T1) *identify critical phases of structure formation*, (T2) *track galaxy life cycles*, (T3) *compare galaxy evolution across environments*, and (T4) *examine galaxy morphology transformation*. For starters, galaxies and halos exist in various phases and environments, and filtering based on properties such as mass or star formation rate can help detect specific galaxies. In temporal mode, selecting a halo from the query allows the user to explore the entire history of that halo by following its largest progenitors in the merger tree from in proceeding snapshots (i.e., $z > 0.0$).

4.5.3 Halo Close-up

While it is interesting to observe and note the large-scale structure of big-box simulations, the diversity of galaxies within them makes it worthwhile to zoom in to see the finer details. Using the galaxy query tool described above

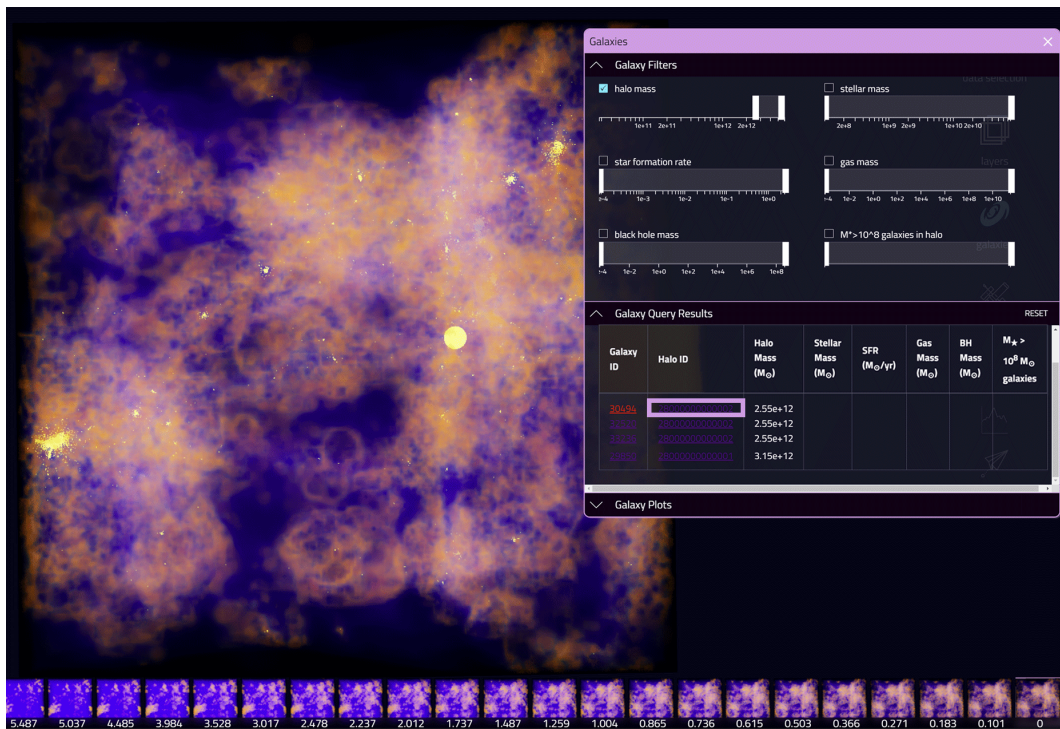


Figure 4.2: The galaxy query tool allows users to filter, highlight and zoom in upon galaxies and halos in the simulation. Here, the metallicity field is shown in the main visualization and a portion of the film strip along the bottom. Hovering over a galaxy ID in the list in the open galaxy panel places a yellow sphere in the volume, indicating its size and position. Clicking on the galaxy ID slices the volume centered upon that galaxy. Clicking on a halo ID loads a higher fidelity zoomed-in volume, as shown in the next figure, Figure 4.3.

and shown in Figure 4.2, the location of individual galaxies can be highlighted from the list, and halos can be selected for higher detailed viewing. Close-ups can be loaded from the table of halos populated in the galaxy query panel to investigate individual halos more closely. When a Halo ID is selected in the galaxy query results, higher-resolution visualization files are loaded for more detail. Users can interact with these close-ups in the same manner as the whole volume. Users can inspect a halo in a single snapshot or track its evolution across multiple time points. A sequence of the heaviest progenitor halos for a given halo sequence is loaded into memory, and the user can switch between redshift frame by frame, watch an animation play out, or switch back and forth between the close-ups and big box by clicking a frame in the film strip. Along the bottom of the screen above the film strip for the entire volume, a film strip containing close-up halos is also shown. The camera rotation, volume slicing tool, and transfer functions are synchronized for each thumbnail on the film strip and linked to the main view to maintain the orientation and slicing. Furthermore, the center location of the selected galaxy is highlighted in the full-box film strip with a green sphere with a diameter of two virial radii, which can be seen in Figure 4.3.

Zooming into halos from the big box simulations helps accomplish the four analysis tasks we outlined in the previous section: (T1) *identify critical phases of structure formation*, (T2) *track galaxy life cycles*, (T3) *compare galaxy evolution across environments*, and (T4) *examine galaxy morphology transformation*.

4.5.4 Side-by-side Comparisons

In addition to the film strips panel and galaxy query zoom-ins, two different snapshots can be selected to make synchronized side-by-side comparisons. As the application runs natively in the web browser, it is simple to open *CosmoVis* in another tab, load the sequence with a different field selected, and view the same sequence at the same or different time points. These two snapshots can be the big box at two different time points, a big box and zoom-in, or two galaxy zoom-ins. This further facilitates cross-comparisons between two galaxies in different environments, such as a galaxy within a dense cluster with another

galaxy in a void or at the edge of an extensive system. This functionality helps directly support (T3) *Compare galaxy evolution across environments*.

4.6 Implementation

Same as the original *CosmoVis* application, this tool is available over the web and accessible via a modern web browser. Users can run this tool locally using several Python scripts or access a live demo version hosted on Indiana University’s *Jetstream2* super-computing platform [134]. The temporal evolution features introduced here are implemented on top of the original *CosmoVis* application codebase [35]. The new functionality requires loading multiple preprocessed snapshots on the client side for visualization as well as the raw simulation data on the server for handling skewer requests. This section discusses the implementation of relevant new preprocessing steps, the visualization interface, and technologies used for deployment and hosting.

4.6.1 Data Processing

One of the major challenges while building this tool is making the data accessible via the web browser at interactive and real-time rates. Due to the large size of the size of the raw simulation data, snapshots are not directly sent to the web browser. Instead, select fields from the simulation are preprocessed at multiple resolutions and hosted as static JSON files in the web application to achieve reasonable loading times for real-time interactions and visualizations. Preprocessed data includes the gas and dark matter particles (gas attributes, hydrogen number density) converted into uniformly gridded volumes, as well as star particles. The preprocessing steps are carried out using the *yt* package for Python [282] in two separate scripts, one for high-resolution galaxy zoom-ins, and another for the full simulation box. The main modifications made to the preprocessing scripts are adding a loop to cycle through multiple snapshots in the same run and saving the files in an appropriate folder hierarchy that can be loaded from the front end. This way, the values remain consistent between snapshots.

To create the galaxy catalog, metadata is collected from the raw simulation

data about each galaxy’s position, unique identifier, host halo ID, progenitor galaxies, mass, star formation rate, black hole mass, virial radius, and spin axis. These attributes are used to add a spin axis arrow glyph to the scene, build a catalog of halo sequences, and populate the galaxy query results list as shown in Figure 4.2. Since galaxy and halo identifiers are unique per snapshot, the merger tree must be traversed to build a sequential list of halo IDs by selecting the largest progenitor halo, starting from the lowest redshift snapshot.

4.6.2 Visualization Interface

The visualization interface can be accessed via any web browser supporting WebGL 2.0. The layout and styling of the web application are defined using HTML and CSS. Javascript facilitates real-time user interactions and data visualization generation. Visualization data (gas volumes, star particles, and metadata) are stored as JSON files and loaded using D3.js. To facilitate the analysis of a series of snapshots, data loaded from preprocessed JSON files using *python* and *yt* [282], and subsequently populate an array with data texture and uniforms for each snapshot. In the user interface, playback buttons for play, start, stop, looping, and switching between the next and previous frames updates the view to display the selected redshift. The data cubes for the volume raymarching of gas attributes, density, and dark matter are stored in the red, green, and blue color channels of a 3D data texture. These textures, cameras, scenes, controllers, and other parts of the 3D graphics visualization are created and managed using the THREE.js javascript library. All interactive graphics are displayed on a single WebGLRenderer canvas. Different views, such as the default volume, side-by-side comparisons, galaxy zoom-ins, and film strip thumbnails, are each supported by their Scene object but share the same controls. This framework supports real-time interactivity with many simultaneous interactive volume renderings.

4.6.3 Deployment and Hosting

This software is deployed as a web application from a virtual machine instance on Indiana University’s *Jetstream2* super-computing platform [134]

running Ubuntu version 20.04. It has also been hosted on AWS EC2 and PRP Nautilus (source). The server hosts the static visualization, *HTML*, *CSS*, and *Javascript* files transferred to the front end and the raw simulation data for on-demand absorption line spectra and column densities generation. Rendering is performed locally on the client’s computer using *THREE.js WebGL*. The server code is written in *Python* using *Flask* as the microservice framework. *Socketio* is used to communicate between the back end and the web client. *RabbitMQ* and *Celery* are configured on the server to create and manage message queues efficiently. Code for processing column densities and synthetic absorption line spectra are written in a separate Python script using *yt* to load the data, *Trident* to process the rays, and *Celery* to handle the task requests from the main *Flask* application. On the server, web ports are configured using *NGINX*, and *Flask* and *Celery* service distributions are controlled with *Gunicorn*.

4.7 Science Use Cases

In collaboration with astronomers, we have developed several use cases that highlight how the application design and implementation outlined in section 4.5 and section 4.6 can be used to address the analysis tasks identified in section 4.4. The data presented in this science use case are from the 12 Mpc EAGLE simulation [244], with the figures all generated using *CosmoVis* [35].

4.7.1 Use Case 1: Identify the time of structure emergence

Upon loading the temporal mode for the dataset in *CosmoVis*, an animation begins looping through the evolution of the big box in the main visualization panel, along with the film strip described in the previous section. The current frame is highlighted in the film strip as the animation plays. In Figure 4.1,

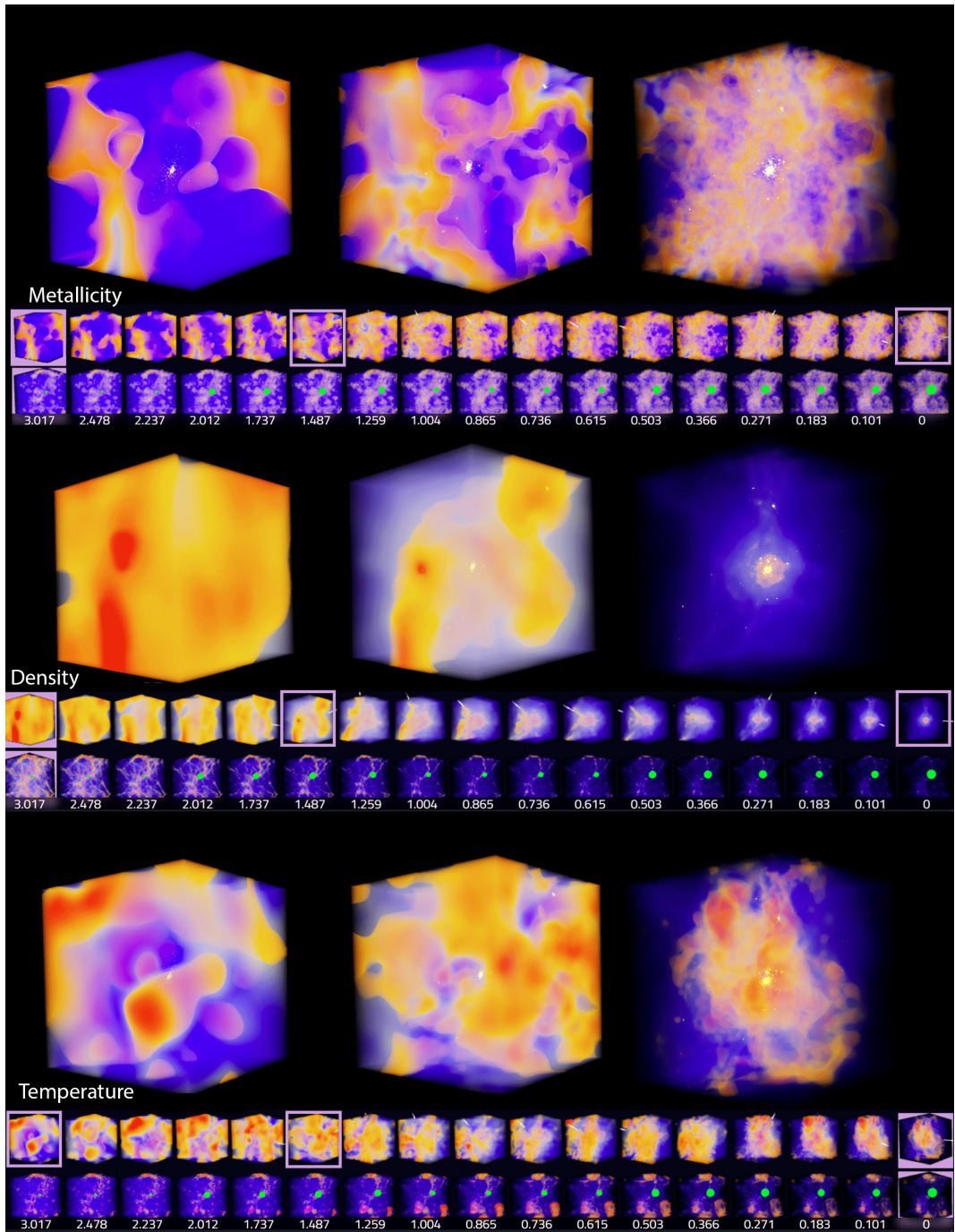


Figure 4.3: Three fields are shown from a galaxy halo in the 12 Mpc EAGLE simulation from top to bottom: metallicity, density, and temperature. For each field, three zoom-in views of the galaxy are shown from different points in time ($z = 3.017, 1.487, 0.000$). Below the large views are two rows of film strips. On top is the zoom-in, and below is the full volume with the position and size of the halo indicated by a green sphere. The blown-up views are highlighted in the zoom-in film strip. The next figure, Figure 4.4, shows a skewer piercing through a cross-section of this halo.

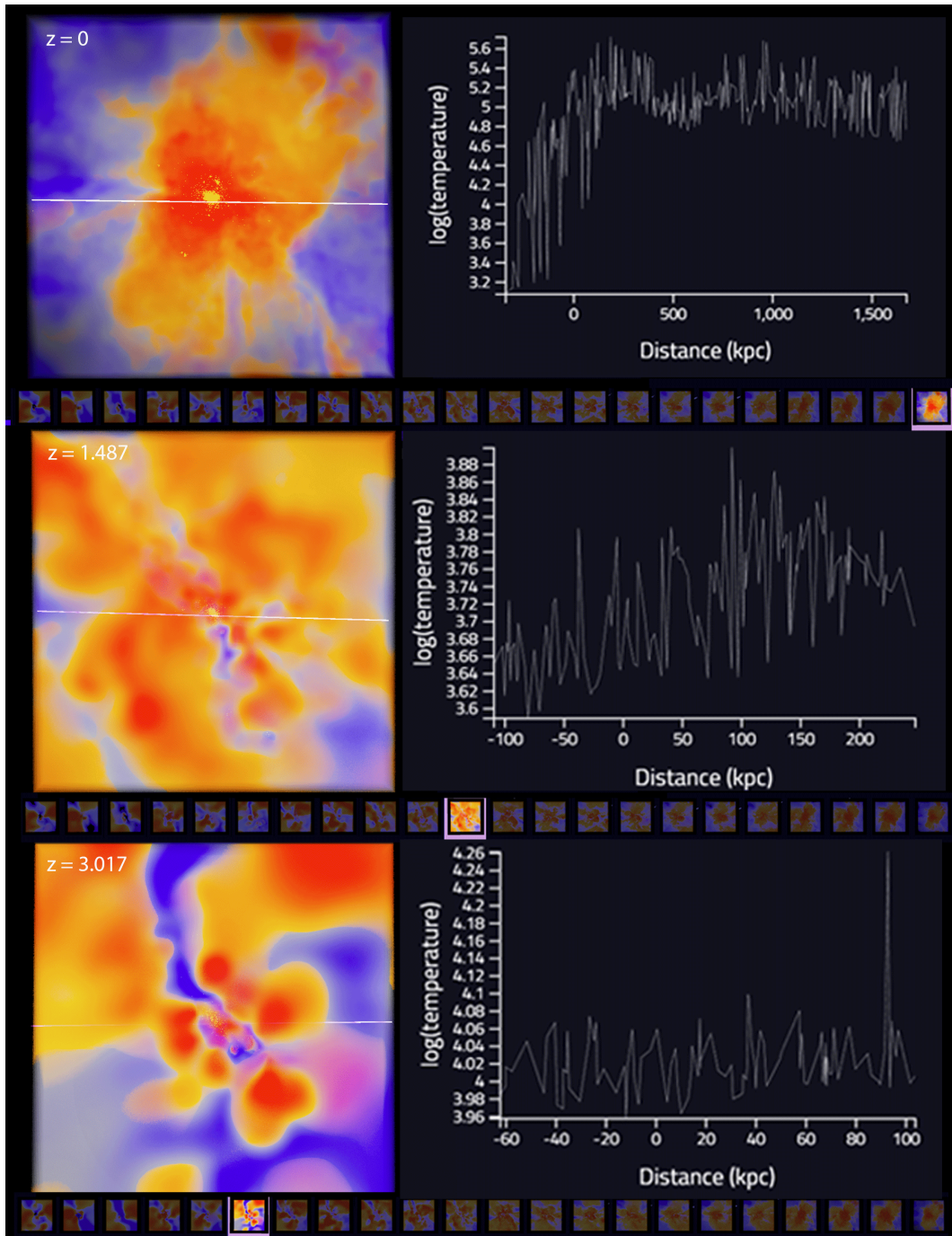


Figure 4.4: Temperature is shown for the same galaxy halo cross-section across three-time points ($z = 0.000, 1.487, 3.017$) arranged top to bottom. The cross-section cuts through the core of the galaxy to highlight interior features. In each frame, a skewer placed horizontally pierces through the center of the view. Skewers contain information about physical parameters (i.e., temperature, density, metallicity, and entropy) and column densities of element and metal ions in the medium. While not shown here, metals were detected using the skewers in the $z = 0.000$ snapshot but not in the earlier snapshots ($z = 1.487$ and 3.017). In Figure 4.4,

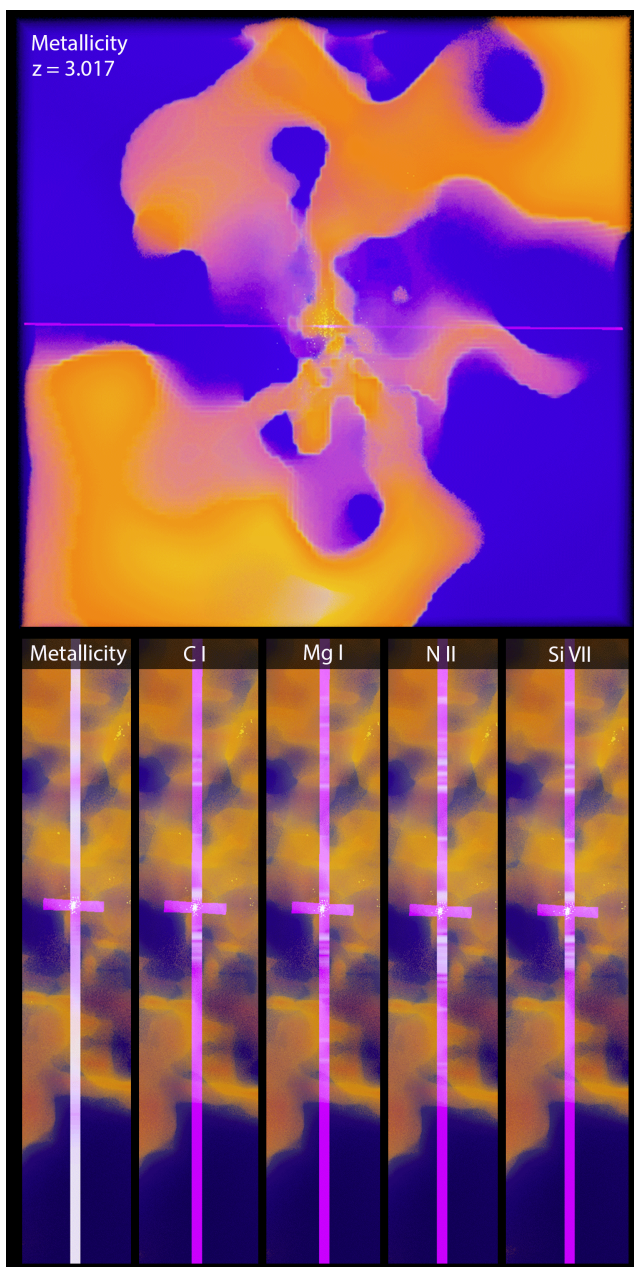


Figure 4.5: Metallicity is shown for the galaxy halo cross-section at $z = 3.017$. The cross-section in the top frame cuts through the galaxy halo's core to highlight interior features, continuing the investigation from Figure 4.4. The purple color along the skewer indicates non-detection, whereas white highlights indicate the presence of that feature. While no metallicity was detected in the zoomed-in region around the halo, we can zoom back out to reveal the neighborhood around the halo, as shown on the bottom. A skewer has been placed perpendicular to the one in the upper frame, and a few fields (metallicity, Carbon I, Magnesium I, Nitrogen II, and Silicon VII) are shown on the skewer. This skewer confirms a deficit of metals within the galaxy halo but shows metals in the IGM outside the halo, as indicated by the white lines.

we see several fields (temperature, entropy, metallicity, and density) with the same color scaling across each snapshot. Looking closely at the film strip row of the density panel in Figure 4.1, we can see the solid red coloration on the left, indicating that the physical structure of the simulation at an early state ($z = 20$) was more-or-less uniformly dense when compared to a lower redshift snapshot. As our eyes pan to the right, we can see the emergence of nuances in the cosmic structure – knots, filaments, and voids. By the end of the simulation ($z = 0$), the densest red regions are scattered throughout the sparse blue volume, indicating the presence of halo clusters embedded within the large-scale structure of the Cosmic Web. If we now look at temperature, we can see the emergence, growth, and dispersion of "bubbles" of hot gas from the center of dense galaxy clusters, spreading outwards into the medium. This directly addresses T1: *Identify key phases of structure formation.*

The new temporal animation tooling in *CosmoVis* makes it simple to select a galaxy in the $z = 0$ snapshot and look back in time to see it interactively visualized in every previous snapshot to its earliest occurrence when the structure was just dense and large enough to be considered a galaxy halo. Activating the temporal animation mode and opening the galaxy query tool, we can easily query galaxies based on several parameters in the $z = 0$ snapshot (mass, star formation rate, etc.) and open a sequence of 'the most giant' progenitor galaxies in the previous snapshots, as shown in Figure 4.2. Only the most massive progenitor halo is shown in the film-strip view to reduce visual clutter, even though several galaxies may merge between snapshots. Using the playback feature, users can watch the evolution of the simulation unfold in the main 3D view while keeping track of where it lies in time by looking at the highlighted frame in the film strip at the bottom of the screen. This temporal zoom-in feature enables T2: *Track galactic life cycles.*

4.7.2 Use Case 2: Compare low redshift halo with progenitors

As introduced in the original *CosmoVis* paper [35] and chapter 3, the skewer tool allows users to take 1D samples along a line of sight that returns column densities of multiple physical parameters along with absorption spectra.

These skewers are comparable to how astronomers take spectral observations of the universe with the Hubble Space Telescope. This skewer tool also works with multiple snapshots loaded in a sequence. The combination of the temporal animation, film strips, and galaxy query zoom-in tool makes it easy to navigate spatially within a dataset and through time in multiple sequential snapshots. Coupled with skewers, users can quantitatively sample the volumes and uncover details that can ultimately be compared with observational measurements.

In the new temporal mode, placing skewers in multiple snapshots and comparing the results is as easy as clicking on a different frame in the film strip and activating the skewer placement mode. In Figure 4.3, we show the metallicity, density, and temperature of the same halo in seventeen sequential snapshots in a zoom-in film strip aligned with a film strip of the whole simulation with a green sphere indicating the position of the halo. For each field, three close-up views of the halo are shown from the sequence ($z = 0.000, 1.487, \text{ and } 3.017$). In Figure 4.4, skewers are placed horizontally through a temperature cross-section of the same halo and temperature plots along the skewer. The skewer contains additional information, such as physical properties (density, entropy, metallicity) and column densities of various elements and metal ions. Interestingly for this halo, the skewer returned a metallicity measurement for the $z = 0.000$ snapshot but not the other two ($z = 1.487$ and 3.017). This could be due to the early age of the galaxy halo (i.e., metals have not been adequately formed yet) or that metals have already been ejected into the IGM and are no longer in the CGM. Looking more closely at the top panel in Figure 4.5, it looks like there are plenty of metals in the galaxy halo; they are just inhomogeneously distributed within it, and the skewer pierced a relatively sparse region.

Another feature to note is the length of the skewer between the halo in each snapshot: nearly 2,000 kpc at $z = 0.0$, about 400 kpc at $z = 1.487$, and only around 150 kpc at $z = 3.017$. Considering the size of the zoom-in halo box has a radius of $2 r_{vir}$ (virial radius). Two main factors contribute to this increase: the expansion of the universe and halo mergers. Despite the visualization's depiction of a smooth, uniformly sized animation between time steps, the physical distance across the volume is smaller at high redshifts when

the universe was young and largest at low redshifts.

Just below the three zoom-ins are multiple fields from the column densities of each line (temperature, entropy, metallicity, light elements, and heavy elements). The peaks in the graph in the *Observe* panel match the highlighted regions along the skewer, decreasing the cognitive load between interpreting the plot and geometrically understanding the configuration of the skewer through space and the morphology of the galaxy halo. This helps solve T4: *Examine galaxy morphology transformation*.

4.7.3 Use Case 3: Compare halo conditions to its local environment

In Figure 4.5, we can see the same halo and skewer as before in the $z = 3.017$ snapshot in the top frame, this time visualizing metallicity in the volume. When queried by the skewers, the halo at this redshift did not exhibit any metallicity. Looking just at the temperature profile, it is unclear where the metals are, only whether the gaseous medium is hot or cold. However, looking at the skewer in relation to the volume's metallicity, it is clear that the skewer pierces through a region devoid of metals (indicated by the blue color) and that metals are instead present (possibly ejected) from the halo perpendicular to the skewer. We can zoom back out to the full view by clicking on the corresponding big-box frames in the filmstrip, which is represented in the bottom panel of Figure 4.5. Here, the original skewer in the top panel is visible as a short horizontal skewer, indicating no metal detection. In the tool, skewers placed within the localized halos are also shown within the context of the big-box snapshot visualizations. When dropped in the zoom-in, the skewer length is limited to the extent of the zoom-in box (a virial radius of two). When we switch back to the full view, the endpoints are subsequently transformed back into the coordinate system of the big box.

In the big-box view, we can place more skewers outside the halo to see how the local environment (the intergalactic medium, or IGM) compares to the interior of the galaxy halo (the circumgalactic medium, or CGM). In the bottom panel of Figure 4.5, a second skewer has been placed vertically, perpendicular to the original skewer, to investigate the local neighborhood around

the halo. Here we can see the following fields mapped upon the skewer: metallicity, Carbon I, Magnesium I, Nitrogen II, and Silicon VII. Along the skewer, white bands indicate the detection of that feature, and purple indicates non-detection. Of course, the skewer data product contains many more ions that can be interactively switched and explored in the user interface. In this case, heavy metals are not located within the halo, but somewhat further outside, which suggests that they may have been ejected. It is interesting to look at metallicity when considering the temporal evolution of galaxies. The early universe consisted of primarily homogenous neutral hydrogen gas. However, as the first stars emerged, died, and went supernova, they spewed heavier elements, or metals, into the environment around them. When we compare the IGM around halos and the CGM within them, we can see that heavy metals are ejected from the centers of clusters. Being able to sample halos across simulations – both spatially and temporally – solves T4: *Examine galaxy morphology transformation*

4.8 Conclusion

In this chapter, we introduce an extension to *CosmoVis* that enables real-time visual analysis and animation of cosmological simulations at both large and small scales. This work allows multiple simulation snapshots to be loaded in series and played back in real time in the web browser. Users can loop, skip ahead, and inspect simulations at different time steps to better understand galaxy evolution in the Cosmic Web. Additionally, the interactive, clickable small multiples film strip view makes it easy to navigate sequential snapshots. Plus, analytic measurements using the same "virtual skewers" can be made in multiple snapshots and subsequently compared. The examples in this chapter come from the EAGLE 12 Mpc box simulation. Still, the framework described is also designed to handle other cosmological simulation codes that can be loaded using *yt* [282], such as Illustris-TNG. This software helps astronomers study the histories and take measurements of galaxies and halos within cosmological simulations that can inspire new sky surveys and be compared to observations of our universe. Skewer measurements taken within *CosmoVis*

can be contextualized within the volume through which they are placed, in a visual scheme similar to how they were introduced in *IGM-Vis*. The difference between the two tools is that one works with observational data (*IGM-Vis*), and the other works with hydrodynamic cosmological simulations (*CosmoVis*). While simulations are powerful tools, they are by no means replacements for making observations of the universe due to their underlying physical assumptions (i.e., prioritizing large-scale modeling over small-scale features), they can be a visual aid for identifying features (e.g., galaxies, filaments, voids, etc.) from multiple angles or redshifts (time points).

In future work, several additions could be incorporated to further the utility of the software and streamline the user experience. Most immediately, more simulations can be preprocessed to be used with this temporal extension to *CosmoVis*. The 12 Mpc EAGLE simulation box shown in this chapter is relatively lightweight (roughly 28 GB of storage for 28 snapshots), which makes it great for quickly prototyping features without waiting too long for data to process and load. However, more state-of-the-art cosmological simulations, such as Illustris-TNG, have intensive storage requirements (50 - 300 Mpc box, 100 snapshots, 0.5 - 2 TB of storage per snapshot). The framework and preprocessing workflow presented here can be applied to larger datasets and visualized in this application. Although, having on-demand processing of synthetic skewer products on demand for the largest simulations is still a technical challenge that can be overcome with the investment of dedicated hardware.

Regarding the skewer tool, one requested feature has been highlighting or retrieving galaxies nearby a sightline. Similar functionality was introduced in *IGM-Vis* [77], where an "impact parameter" slider can be used to sample galaxies encompassed by a specific radius around the skewer. This could help locate interesting local galaxies hidden within the large-scale structure of the volume. Another querying tool could be a filament tracker or highlighting feature. One of the challenges with observational and even simulations is seeing the filamentary structure between galaxies, halos, and clusters. A click-and-drag tool could allow the user to highlight the densest gaseous or dark matter voxel along the cursor's line of sight. Users could hone in using the existing volume slicing feature, and adjusting the radius of the new highlight tool based

on kpc to annotate the volume and return spatial coordinates and more precise metadata from the simulation. This could be useful to find minimum and maximum values based on criteria such as temperature, density, entropy, etc., in the dataset.

When tracking galaxy halo histories through the merger tree, it could be helpful to show more information. Only the largest progenitor halo is shown to keep the film strip panel from being too cluttered. Since mergers result from two or more halos collapsing into the same structure, a decision was made to only follow the sequence of the largest halo back in time. There could be a feature to indicate at least how many halos merged between snapshots or expand upon the thumbnails to show all of the progenitors and their trajectories across time. A Python package such as *ytree* could be incorporated into the data processing pipeline to accommodate different merger tree codes.

Chapter 5

Conclusion

In this thesis, I have introduced three novel interactive astronomy visualization tools designed and created in close collaboration with astronomers. Together, these tools streamline the analytic workflow of astronomers studying observations of our universe and cosmological simulations. This is done by transforming otherwise cumbersome, coding-intensive tasks into accessible interactive visualization tools that run directly in the web browser. Interweaving challenges exist for visualizing both these types of data, observations, and simulations, whereby galaxies are embedded within a volume, and absorption line spectra are superimposed on cylindrical "skewer sightlines." These three tools use the "skewer sightline" as a means to spatially contextualize cosmological measurements within the cosmic environments through which they pierce. The visualization challenge of large, cosmological datasets inspired these tools. Part of what ties these programs together is their ability to take multiple pieces of information otherwise hidden within complex data structures and place them into unified, linked-view interfaces that are accessible via the web browser.

Immediate future work will be to prepare chapter 4 for submission to a visualization conference venue such as *IEEE VIS*, as chapter 2 and chapter 3 have already been published in *Eurographics Eurovis* and *IEEE Transactions in Visualization and Computer Graphics*. Before the paper submission and publication, the code and online software demonstration for the temporal analysis features introduced in chapter 4 will be made publicly available to the community. Future work can further bridge the gap between how observational measurements can be compared to simulation results. One avenue in particu-

lar that is underexplored would be to be bringing observations and simulations together into the same tool. Both *IGM-Vis* and *CosmoVis* more-or-less visualize the same types of information, just in different contexts (observations vs. simulations) and with varying levels of fidelity. On the one hand, observations serve as the basis for our understanding of the universe despite not being able to easily see the complete picture (for example, derived measurements that are not optically observable, but describe the physics of the universe), which can be statistically compared to the outputs of galaxies in simulations (e.g., comparing the mass distribution). On the other hand, simulations contain multitudes of fields that are otherwise difficult to observe thoroughly in our actual universe but are fundamentally going to be imperfect, and at best, serve as an approximation of the underlying mechanisms. The particle tracers in cosmological simulations are already a statistical approximation of the contents of a region of space. They are not as granular as the name suggests (i.e., a star "particle" contains more than a single star).

To address some of the challenges with preprocessing entire simulations snapshots to be used in *CosmoVis* (especially those 100+ Mpc across), perhaps some computation can be democratized and conducted on power users' local computers rather than as a preprocessing step or on a server. Sometimes astronomers may be interested in studying one particular redshift snapshot of a simulation suite, such as an Illustris-TNG box, and already have the data stored on their machine powerful enough to open it (one of the main bottlenecks in the analysis pipelines is acquiring and opening the datasets). But there might be a case where the dataset has not yet been preprocessed for use in *CosmoVis*. A future version of *CosmoVis* could leverage advances in running Python scripts natively in the web browser through Anaconda's new experimental open-source package PyScript. Given the sheer number of galaxies and scale contained in the largest, high-fidelity big-box simulations, it may be worthwhile and now possible to generate and preprocess data locally using the clients' hardware rather than as an offline step or running it on the server. This would require the client to store the simulation data on their computer. Still, often researchers may be most interested in studying a particular region of a specific snapshot and may not need the entire sequence in full fidelity

but would like to harness the speed of their computing hardware. This could also be used to process skewer sightlines locally, reducing latency. This power user mode could be synchronized to the cloud server, where any preprocessed visualization halo data or skewers are uploaded into a shared repository for other users to explore. To further streamline efforts to reduce coding in the preprocessing workflow, a universal simulation loading utility with a graphical user interface could be developed to make it easier to prepare different simulations for *CosmoVis*. This tool could still rely on the Python package *yt* as it supports many simulation codes. Users could drag and drop the simulation folder, either a single snapshot or a sequence, into the application. Once selected, the tool can load metadata about each snapshot, such as redshift, box size, and number of particles for example. From there, the user can select from all available particle types and fields within the dataset, which are automatically retrieved with the metadata and the units. After that, users can choose the voxel resolutions they want to preprocess (i.e., 64, 128, 256, 512 voxels per edge). Pressing "run" would open a console window that runs the preprocessing script, and the utility GUI is updated with a progress bar. Something like this would significantly reduce the complexity of loading in a new simulation. Plus, it could be configured to synchronize preprocessed data with a cloud repository for other users to enjoy.

In 2021, the "Pathways to Discovery in Astronomy and Astrophysics for the 2020s" [197] was released, outlining future efforts and goals across astronomy for the next few decades and generations of researchers. Technical advances in a wide range of observational astronomy capabilities will shine a new light on imaging of exoplanets and stars, other solar systems, and physical phenomena such as gravitational waves and particles, to characterize and understand the origins, formation, and evolution of cosmic structures such as black holes, neutron stars, galaxies, the filaments of the cosmic web, dark matter and dark energy. With new observational data sources, methods, and results, effectively communicating those new findings to a broader audience will be needed. As time has shown, dynamic interactive visualization methods go hand-in-hand with astronomers conducting science and sharing knowledge. New methods of presenting and interacting with data will need to be developed, which will

probably be driven not by advances in astronomy, but by market-driven advances in the capabilities of consumer computer graphics hardware to support real-time rendering of games and animation, the performance and accessibility of new displays (AR/VR/holograms), and novel analysis techniques driven by interactive visualizations. Given the extensive range of scales astronomers deal with, both temporally and spatially, and the amount of data produced by observations and simulations, there may not ever be one visualization tool or method that will work in every situation. Still, it is not outside the realm of possibility.

Chapter 6

Appendix

6.1 Glossary

Absorption line spectroscopy— a widely-used observational technique for probing the spectral properties of gas in intra-galactic environments and the intergalactic medium. To date, dozens of characteristic absorption signatures for specific element ions have been identified in astronomical spectra. Cross-correlating these known signatures in the spectra of distant emitters (e.g., quasars) can reveal the presence of gas between the source and the observer. One can measure the redshift of the absorbing gas, indicating its distance, as well as the dynamics and chemical makeup of the intervening material.

Active galactic nucleus (AGN)— a highly luminous, compact and energetic region at the center of "active" galaxies. Likely powered by a super-massive black hole that accretes matter and emits intense radiation across the electromagnetic spectrum.

C IV— triply ionized carbon. Tracer of ionized, heavy element-enriched CGM and IGM gas within IGM-Vis' default dataset

Circumgalactic medium (CGM)— baryonic matter such as gas, plasma, and dust, gravitationally bound to a particular galaxy but existing outside the central regions where most stars reside.

COS, Cosmic Origins Spectrograph— COS is the instrument aboard HST that obtained QSO spectra for IGM-Vis' default dataset

Cosmic epoch— Refers to the different stages of the universe as its structure and composition have evolved from its early state from the Big Bang to

now.

Cosmic web— the totality of knots, filaments, walls and voids, which constitute the Universe’s large-scale structure. This structure has evolved from the original mostly-uniform matter distribution through the influence of gravity, electromagnetism, and the theorized dark energy. The distribution of its characteristic features has been first predicted by analytic theory [106, 149, 312] and more recently substantiated by wide-field galaxy surveys and large-scale cosmological simulations, which we analyze through *CosmoVis*.

Cosmological redshift z — a standardized cosmological measure indicating both time and distance, discovered by Hubble [144]. The value $z = 0$ corresponds to the present day. Redshift results from the expansion of space over which light emitted by a distant object at an earlier cosmic epoch travel towards an observer at present time [74]; as a result, light rays appear by the observer at longer (redder) wavelengths than they were emitted.

Cosmological simulations— first-principles physics simulations performed at super-galactic scales, e.g., 50-500 Mpc. Such simulations typically consider tens of billions ‘particles’ (macro tracers of conventional baryonic and dark matter) and timescales spanning the entire duration of the Universe.

Column density— a measure of the amount of matter (often an element or ion) along an observer’s line of sight, typically in units of particles/cm². A shorthand notation is often used to denote element or ion column density. For example, “the column density of neutral hydrogen” can be written “N(H I)”.

DEC, declination— A celestial coordinate that, along with RA, defines a position on the sky

Equivalent width— A measure of absorption strength for a spectral line that indicates the amount of material along the line of sight

Filament— These are the largest features of the cosmic web and are composed of diffuse gas, galaxies, and dark matter.

H I— neutral hydrogen. Primary tracer of IGM and CGM gas within IGM-Vis’ default dataset

Halo— gravitationally bound structure containing galaxies

HST, Hubble Space Telescope— HST is an ultraviolet/optical/infrared observatory launched in 1990, and is the source of all QSO spectra for IGM-Vis’

default dataset

Intergalactic medium (IGM)— baryonic matter such as gas, plasma, and dust that is not gravitationally bound to a particular galaxy, but instead belongs to the cosmic web.

Kiloparsec (kpc)— an astrophysical unit of distance, approximately 3,260 light years or 3×10^{16} km.

Large-scale structure (LSS)— the high-level organizational structure of the Universe, starting at scales larger than individual galaxies. Embedded in it are ecosystems of galaxies whose evolution is regulated by the matter and energy budget surrounding them. Also see “cosmic web” above.

Ly α , Lyman α — Spectral transition from the ground state to the first excited state of H I. Ly α is a spectral feature indicated in IGM-Vis’ skewers and measured for equivalent width profiles

Megaparsec (Mpc)— an astrophysical unit of distance, equal to 1000 kpc.

Merger Tree— A branching graphical or hierarchical representation that tracks the merging of smaller galaxies to form larger ones over cosmic time.

Outflow— Process where material such as gas is ejected from a galaxy into its surrounding environment and intergalactic medium due to an active galactic nucleus (AGN) or intense star formation within the galaxy.

QSO, Quasi-stellar object— Quasar or other object used as background source for absorption line spectroscopy. Skewers in IGM-Vis are generated from QSO spectra

RA, right ascension— A celestial coordinate that, along with DEC, defines a position on the sky

r_{vir} **virial radius**— A fiducial radius from the center of a galaxy to which the CGM extends. Within this radius, gas is considered likely to be gravitationally bound to the galaxy

SDSS, Sloan Digital Sky Survey— SDSS is a large imaging and spectroscopic survey that provides all galaxy information within IGM-Vis’ default dataset

SFR, star formation rate— The star formation rate is typically expressed in units of the mass of the Sun per year. Galaxies are colored in

IGM-Vis according to their SFR relative to their mass

Void— Vast, low-density cavities that lie between the filaments of the cosmic web and are lacking in galaxies and other material.

Bibliography

- [1] 1930's Woman with spinning wheel. Moon worshippers. | Huntley Film Archives.
- [2] Almagestū Cl. Ptolemei. Opus ingens ac nobile omnes celorū motus continens.
- [3] Astronomy - lovely 1930's animation the solar system | Huntley Film Archives.
- [4] Astronomy in ancient and medieval china | Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences.
- [5] Brahe's Mural Quadrant.
- [6] From the Scale Model of the Sky to the Armillary Sphere: The Physics Teacher: Vol 52, No 7.
- [7] Galileo and the Telescope | Modeling the Cosmos | Articles and Essays | Finding Our Place in the Cosmos: From Galileo to Sagan and Beyond | Digital Collections | Library of Congress.
- [8] The High Energy Astrophysics Observatory 2 (Einstein).
- [9] Museo Galileo - In depth - Ecliptic coordinates.
- [10] Ptolemaic system | Definition & Facts | Britannica.
- [11] Secrets Of Nature.
- [12] This Month in Astronomical History: 50 Years of CCDs | American Astronomical Society.

- [13] Tycho Brahe's Observations and Instruments | High Altitude Observatory.
- [14] Among The Stars 1921, 1921.
- [15] The Eclipse, 1921.
- [16] Solar Eclipse, 1921.
- [17] Partial Eclipse Sweeps Across England, 1925.
- [18] Total Eclipse Of The Sun, 1925.
- [19] Astronomers View Mars Through Telescope, 1926.
- [20] Science's latest wonder. The Planetarium - with its 119 lenses - enables onlooker to observe the planets in their courses, 1926.
- [21] The Mystery Of The Moon, 1932.
- [22] Planetarium (1932), 1932.
- [23] A Trip To The Moon, 1932.
- [24] Searching The Heavens, 1933.
- [25] Through Galileo's Telescope, 1933.
- [26] How The Earth Was Born, 1934.
- [27] Eclipses To Order (1938), 1938.
- [28] OP BEZOEK BIJ EEN AMATEUR-ASTRONOOM, 1943.
- [29] HET PLANETARIUM, 1944.
- [30] Great Universe, 1945.
- [31] Astronomy, 1940's - Film 18316, Apr. 2015.
- [32] Galileo and the Telescope, Feb. 2022. Australia Telescope National Facility.

- [33] S. J. Aarseth and F. Hoyle. Dynamical Evolution of Clusters of Galaxies, I. *Monthly Notices of the Royal Astronomical Society*, 126(3):223–255, June 1963. doi: 10.1093/mnras/126.3.223
- [34] B. Abolfathi, D. S. Aguado, G. Aguilar, C. Allende Prieto, A. Almeida, T. T. Ananna, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, and et al. The fourteenth data release of the Sloan Digital Sky Survey: First spectroscopic data from the extended Baryon Oscillation Spectroscopic Survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. *The Astrophysical Journal Supplement Series*, 235:42, 2018.
- [35] D. Abramov, J. N. Burchett, O. Elek, C. Hummels, J. X. Prochaska, and A. G. Forbes. Cosmovis: An interactive visual analysis tool for exploring hydrodynamic cosmological simulations. *IEEE Transactions on Visualization and Computer Graphics*, 28(8):2909–2925, 2022.
- [36] A. Aguirre, L. Hernquist, J. Schaye, N. Katz, D. H. Weinberg, and J. Gardner. Metal enrichment of the intergalactic medium in cosmological simulations. *The Astrophysical Journal*, 561(2):521, 2001.
- [37] J. Ahrens, B. Geveci, and C. Law. ParaView: An end-user tool for large data visualization. In *The Visualization Handbook*, chap. 36, pp. 717–732. Elsevier, 2005.
- [38] B. a.-Z. b. a. R. al Jazari. "Design for the Water Clock of the Peacocks", from the *Kitab fi ma'rifat al-hiyal al-handasiyya* (Book of the Knowledge of Ingenious Mechanical Devices) by Badi' al-Zaman b. al Razzaz al-Jazari, 715.
- [39] K. R. Almryde and A. G. Forbes. Halos in a Dark Sky: Interactively exploring the structure of dark matter halo merger trees. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)*, pp. 73–77, 2015.
- [40] I. Altintas, K. Marcus, I. Nealey, S. L. Sellars, J. Graham, D. Mishin, et al. Workflow-driven distributed machine learning in CHASE-CI: A cognitive hardware and software ecosystem community infrastructure.

In *Proceedings of the IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 865–873, 2019.

- [41] J. An, K. Won, Y. Kim, J.-Y. Hong, H. Kim, Y. Kim, H. Song, C. Choi, Y. Kim, J. Seo, A. Morozov, H. Park, S. Hong, S. Hwang, K. Kim, and H.-S. Lee. Slim-panel holographic video display. *Nature Communications*, 11(1), Nov. 2020. Number: 1 Publisher: Nature Publishing Group. doi: 10.1038/s41467-020-19298-4
- [42] S. Anderl. Implicit Changes of Model Uses in Astrophysics, Illustrated on the Paris-Durham Shock Model. *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin*, 27(4):515–546, Dec. 2019. doi: 10.1007/s00048-019-00225-8
- [43] D. Arnett. Mass Dependence In Gravitational Collapse Of Stellar Cores. *Canadian Journal of Physics*, 45(5):1621–1641, May 1967. doi: 10.1139/p67-126
- [44] D. Arnett and C. Meakin. Turbulent Mixing in Stars: Theoretical Hurdles. *Proceedings of the International Astronomical Union*, 5, Dec. 2009. doi: 10.1017/S174392131000030X
- [45] U. Ayachit. *The ParaView guide: A parallel visualization application*. Kitware, Inc., 2015.
- [46] J. N. Bahcall and P. J. E. Peebles. Statistical tests for the origin of absorption lines observed in quasi-stellar sources. *The Astrophysical Journal Letters*, 156:L7, 1969.
- [47] J. N. Bahcall and L. Spitzer, Jr. Absorption lines produced by galactic halos. *The Astrophysical Journal Letters*, 156:L63, 1969.
- [48] D. Baines, F. Giordano, E. Racero, J. Salgado, B. L. Martí, B. Merín, M.-H. Sarmiento, R. Gutiérrez, I. O. de Landaluce, I. León, et al. Visualization of multi-mission astronomical data with esasky. *Publications of the Astronomical Society of the Pacific*, 129(972):028001, 2016.

- [49] I. K. Baldry, K. Glazebrook, J. Brinkmann, Z. Ivezić, R. H. Lupton, R. C. Nichol, and A. S. Szalay. Quantifying the bimodal color-magnitude distribution of galaxies. *The Astrophysical Journal*, 600:681–694, 2004.
- [50] M. L. Balogh, J. F. Navarro, and S. L. Morris. The origin of star formation gradients in rich galaxy clusters. *The Astrophysical Journal*, 540(1):113, 2000.
- [51] C. Beaumont, A. Goodman, and P. Greenfield. Hackable user interfaces in astronomy with Glue. In *Astronomical Data Analysis Software and Systems XXIV*, vol. 495, p. 101, 2015.
- [52] S. A. Bedini and F. R. Maddison. Mechanical Universe: The Astrarium of Giovanni de’ Dondi. *Transactions of the American Philosophical Society*, 56(5):1–69, 1966. Publisher: American Philosophical Society. doi: 10.2307/1006002
- [53] M. J. Berger and J. Olinger. Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics*, 53(3):484–512, 1984.
- [54] J. Bergeron and P. Boissé. A sample of galaxies giving rise to mg II quasar absorption systems. *A&A*, 243:344–366, Mar. 1991.
- [55] T. Berners-Lee. World-Wide Web: The Information Universe. (1):7, 1992.
- [56] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001. Publisher: Scientific American, a division of Nature America, Inc.
- [57] G. B. Berriman and J. Good. The application of the Montage Image Mosaic Engine to the visualization of astronomical images. *Publications of the Astronomical Society of the Pacific*, 129(975):058006, 2017.
- [58] K. Bladin, E. Axelsson, E. Broberg, C. Emmart, P. Ljung, A. Bock, and A. Ynnerman. Globe browsing: Contextualized spatio-temporal planetary surface visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):802–811, 2017.

- [59] P.-A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W.-Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian. Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature*, 468(7320):80–83, Nov. 2010. Number: 7320 Publisher: Nature Publishing Group. doi: 10.1038/nature09521
- [60] M. R. Blanton, E. Kazin, D. Muna, B. A. Weaver, and A. Price-Whelan. Improved background subtraction for the Sloan Digital Sky Survey images. *Astronomical Journal*, 142(1):31, 2011.
- [61] T. Boch and P. Fernique. Aladin Lite: Embed your sky in the browser. In *Astronomical Data Analysis Software and Systems XXIII*, vol. 485, p. 277, 2014.
- [62] A. Bock, E. Axelsson, J. Costa, G. Payne, M. Acinapura, V. Trakinski, et al. OpenSpace: A system for astrographics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):633–642, 2019.
- [63] A. Boksenberg and W. L. W. Sargent. The existence of CA II absorption lines in the spectrum of the quasar 3c 232 due to the Galaxy NGC 3067. *The Astrophysical Journal*, 220:42–46, 1978.
- [64] J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575):603–606, Apr 1996.
- [65] F. Bonnarel, P. Fernique, O. Bienaymé, D. Egret, F. Genova, M. Louys, F. Ochsenbein, M. Wenger, and J. G. Bartlett. The ALADIN interactive sky atlas—a reference tool for identification of astronomical sources. *Astronomy and Astrophysics Supplement Series*, 143(1):33–40, 2000.
- [66] R. Bordoloi, A. J. Fox, F. J. Lockman, B. P. Wakker, E. B. Jenkins, B. D. Savage, S. Hernandez, J. Tumlinson, J. Bland-Hawthorn, and T.-S. Kim. Mapping the nuclear outflow of the Milky Way: Studying the kinematics and spatial extent of the northern Fermi Bubble. *The Astrophysical Journal*, 834:191, 2017.

- [67] S. Borthakur, T. Heckman, J. Tumlinson, R. Bordoloi, C. Thom, B. Catinella, D. Schiminovich, R. Davé, G. Kauffmann, S. M. Moran, and A. Saintonge. Connection between the circumgalactic medium and the interstellar medium of galaxies: Results from the COS-GASS Survey. *The Astrophysical Journal*, 813:46, 2015.
- [68] M. O. Boussejra, R. Uchiki, Y. Takeshima, K. Matsubayashi, S. Takekawa, M. Uemura, and I. Fujishiro. aflak: Visual programming environment enabling end-to-end provenance management for the analysis of astronomical datasets. *Visual Informatics*, 2019.
- [69] D. V. Bowen, D. Chelouche, E. B. Jenkins, T. M. Tripp, M. Pettini, D. G. York, and B. L. Frye. The structure of the circumgalactic medium of galaxies: Cool accretion inflow around NGC 1097. *The Astrophysical Journal*, 826:50, 2016.
- [70] T. Brahe. Tychonis Brahe Astronomiæ instauratæ mechanica, 1602. Publisher: apud L. Hvlsivm, 1602.
- [71] K. Brasch. A Short History of Astrophotography (Part 1). *The Journal of the Royal Astronomical Society of Canada*, 111:52, 2017.
- [72] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
- [73] T. Buick. *Orrery: A Story of Mechanical Solar Systems, Clocks, and English Nobility*. Springer Science & Business Media, Oct. 2013. Google-Books-ID: IU24AQAAQBAJ.
- [74] E. F. Bunn and D. W. Hogg. The kinematic origin of the cosmological redshift. *American Journal of Physics*, 77(8):688–694, 2009.
- [75] J. N. Burchett. Surveying the CGM and IGM across 4 orders of magnitude in environmental density. HST Proposal #15009, Aug. 2017. <http://adsabs.harvard.edu/abs/2017hst..prop15009B>.

- [76] J. N. Burchett, D. Abramov, O. Elek, and A. G. Forbes. Volumetric reconstruction for interactive analysis of the cosmic web. In *Proceedings of the IEEE VIS Workshop on Visualization in Astrophysics (VisAstro)*, pp. 1–15, 2020.
- [77] J. N. Burchett, D. Abramov, J. Otto, C. Artanegara, J. X. Prochaska, and A. G. Forbes. IGM-Vis: Analyzing intergalactic and circumgalactic medium absorption using quasar sightlines in a cosmic web context. *Computer Graphics Forum*, 38(3):491–504, 2019.
- [78] J. N. Burchett, O. Elek, N. Tejos, J. X. Prochaska, T. M. Tripp, R. Bordoloi, and A. G. Forbes. Revealing the dark threads of the Cosmic Web. *The Astrophysical Journal Letters*, 891(2):L35, 2020.
- [79] J. N. Burchett, K. H. R. Rubin, J. X. Prochaska, A. L. Coil, R. R. Vaught, and J. F. Hennawi. Circumgalactic Mg II emission from an isotropic starburst galaxy outflow mapped by KCWI. *The Astrophysical Journal*, 909(2):151, 2021.
- [80] J. N. Burchett, T. M. Tripp, R. Bordoloi, J. K. Werk, J. X. Prochaska, J. Tumlinson, et al. A deep search for faint galaxies associated with very low-redshift C IV absorbers: III. The mass- and environment-dependent circumgalactic medium. *The Astrophysical Journal*, 832(124):1–23, 2016.
- [81] J. N. Burchett, T. M. Tripp, R. Bordoloi, J. K. Werk, J. X. Prochaska, J. Tumlinson, C. N. A. Willmer, J. O’Meara, and N. Katz. A deep search for faint galaxies associated with very low-redshift C IV absorbers: III. The mass- and environment-dependent circumgalactic medium. *The Astrophysical Journal*, 832(124):124, 2016.
- [82] J. N. Burchett, T. M. Tripp, J. X. Prochaska, J. K. Werk, J. Tumlinson, J. M. O’Meara, R. Bordoloi, N. Katz, and C. N. A. Willmer. A deep search for faint galaxies associated with very low-redshift C IV absorbers: II. Program design, absorption-line measurements, and absorber statistics. *The Astrophysical Journal*, 815(2):91, 2015.
- [83] J. N. Burchett, T. M. Tripp, Q. D. Wang, C. N. A. Willmer, D. V. Bowen, and E. B. Jenkins. Warm-hot gas in X-ray bright galaxy clusters and

- the H I-deficient circumgalactic medium in dense environments. *Monthly Notices of the Royal Astronomical Society*, 475:2067–2085, 2018.
- [84] A. Burrows, J. Nordhaus, I. Hubeny, J. M. Stone, K. MacGregor, and K. Werner. Core-Collapse Supernova Explosion Simulations: The Path to and Necessity for 3D Models. In *AIP Conference Proceedings*, pp. 273–283. AIP, Boulder (Colorado), 2009. doi: 10.1063/1.3250066
- [85] A. L. Busby. Picturing the cosmos : Surrealism, astronomy, astrology, and the Tarot, 1920s-1940s. Dec. 2013. Accepted: 2014-02-19T15:55:08Z.
- [86] R. J. Buta, J. Falcón-Barroso, and J. Knapen. Galaxy morphology. *Secular Evolution of Galaxies*, p. 155, 2013.
- [87] Z. Cai, X. Fan, S. Peirani, F. Bian, B. Frye, I. McGreer, J. X. Prochaska, M. W. Lau, N. Tejos, S. Ho, and D. P. Schneider. Mapping the most massive overdensity through hydrogen (MAMMOTH) I: Methodology. *The Astrophysical Journal*, 833:135, 2016.
- [88] C. Canavas. Al-Jazarī’s Compendium of Ingenious Devices: A model of representing and communicating technical knowledge in a medieval Islamicate context. In L. Hilaire-Pérez, V. Nègre, D. Spicq, and K. Vermeir, eds., *Le livre technique avant le xxe siècle : À l’échelle du monde*, CNRS Alpha, pp. 71–82. CNRS Éditions, Paris, Nov. 2019. Code: Le livre technique avant le xxe siècle : À l’échelle du monde.
- [89] M. Caspar. *Kepler*. Courier Corporation, Oct. 2012. Google-Books-ID: cYQUAwwAAQBAJ.
- [90] R. Cen, J. Miralda-Escudé, J. P. Ostriker, and M. Rauch. Gravitational collapse of small-scale structure as the origin of the Lyman-alpha forest. *The Astrophysical Journal Letters*, 437:L9–L12, 1994.
- [91] R. Cen and J. P. Ostriker. Where are the baryons? *The Astrophysical Journal*, 514:1–6, 1999.
- [92] M. R. Chartrand. A Fifty Year Anniversary of a Two Thousand Year Dream. *The Planetarian*, 1973.

- [93] G. Chaucer. *A Treatise on the Astrolabe*. University of Oklahoma Press, 2002. Google-Books-ID: lLzYeLlhMdgC.
- [94] H.-W. Chen, K. M. Lanzetta, and J. K. Webb. The origin of C IV absorption systems at redshifts $z < 1$: Discovery of extended C IV envelopes around galaxies. *The Astrophysical Journal*, 556(1):158–163, 2001.
- [95] H.-W. Chen, J. X. Prochaska, B. J. Weiner, J. S. Mulchaey, and G. M. Williger. Probing the intergalactic medium-galaxy connection toward PKS 0405-123. II. A cross-correlation study of Ly α absorbers and galaxies at $z < 0.5$. *The Astrophysical Journal Letters*, 629:L25–L28, 2005.
- [96] X. Chen, L. Shen, Z. Sha, R. Liu, S. Chen, G. Ji, and C. Tan. A survey of multi-space techniques in spatio-temporal simulation data visualization. *Visual Informatics*, 3(3):129–139, 2019.
- [97] N. R. Council. *The Decade of Discovery in Astronomy and Astrophysics*. Nov. 1. doi: 10.17226/1634
- [98] M. Crone Odekon, G. Hallenbeck, M. P. Haynes, R. A. Koopmann, A. Phi, and P.-F. Wolfe. The Effect of Filaments and Tendrils on the H I Content of Galaxies. *The Astrophysical Journal*, 852(2):142–1–14, 2018.
- [99] M. Crosta. Astrometry in the 21st century. from hipparchus to einstein. *La Rivista del Nuovo Cimento*, 42(10):443–510, 2019.
- [100] T. Dang, P. Murray, and A. G. Forbes. BioLinker: Bottom-up exploration of protein interaction networks. In *Proceedings of the 10th IEEE Pacific Visualization Symposium (PacificVis)*, pp. 265–269. Seoul, Korea, 2017.
- [101] R. Davé, B. D. Oppenheimer, and K. Finlator. Galaxy evolution in cosmological simulations with outflows – I. Stellar masses and star formation rates. *Monthly Notices of the Royal Astronomical Society*, 415(1):11–31, 2011.
- [102] W. Davis. The Evolution of the Universe. *Current History (New York)*, 35(3):414–416, Dec. 1931. Num Pages: 3 Place: New York, United States Publisher: C-H publishing corporation, etc.

- [103] E. Dekker. Featuring the first Greek Celestial Globe. *Globe Studies*, (55/56):133–152, 2009. Publisher: International Coronelli Society for the Study of Globes.
- [104] S. Djorgovski, S. Castro, D. Stern, and A. Mahabal. On the threshold of the reionization epoch. *The Astrophysical Journal*, 560(1):L5, 2001.
- [105] M. Donnari, A. Pillepich, D. Nelson, F. Marinacci, M. Vogelsberger, and L. Hernquist. Quenched fractions in the illustrating simulations: comparison with observations and other theoretical models. *Monthly Notices of the Royal Astronomical Society*, 506(4):4760–4780, 2021.
- [106] A. G. Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6(4):320–330, 1970.
- [107] H. Draper. On photographing the spectra of the stars and planets. *Memorie della Societa Degli Spettroscopisti Italiani*, 8:A81–A86, Jan. 1879. ADS Bibcode: 1879MmSSI...8A..81D.
- [108] M. Edmunds. The Antikythera mechanism and the mechanical universe. *Contemporary Physics*, 55(4):263–285, Oct. 2014. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/00107514.2014.927280>. doi: 10.1080/00107514.2014.927280
- [109] A. Einstein. Ether and the Theory of Relativity. In M. Janssen, J. D. Norton, J. Renn, T. Sauer, and J. Stachel, eds., *The Genesis of General Relativity*, pp. 1537–1542. Springer Netherlands, Dordrecht, 2007. doi: 10.1007/978-1-4020-4000-9_34
- [110] O. Elek, J. N. Burchett, J. X. Prochaska, and A. G. Forbes. Polyphorm: Structural analysis of cosmological datasets via interactive Physarum polycephalum visualization. *IEEE Transactions of Visualization and Computer Graphics*, 27(2):806–816, 2021.
- [111] R. Etemadpour, M. Bomhoff, E. Lyons, P. Murray, and A. G. Forbes. Designing and evaluating scientific workflows for big data interactions. In *Proceedings of the IEEE Symposium on Big Data Visual Analytics (BDVA)*, pp. 1–8, 2015.

- [112] C. -. A. d. t. Flammarion. *Astronomie populaire : description générale du ciel... Tome 1 / Camille Flammarion*. 1880.
- [113] A. G. Forbes, A. Burks, K. Lee, X. Li, P. Boutillier, J. Krivine, and W. Fontana. Dynamic influence networks for rule-based models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):184–194, 2018.
- [114] A. J. Fox, B. P. Wakker, K. A. Barger, A. K. Hernandez, P. Richter, N. Lehner, J. Bland-Hawthorn, J. C. Charlton, T. Westmeier, C. Thom, and et al. The COS/UVES absorption survey of the Magellanic Stream. III. Ionization, total mass, and inflow rate onto the Milky Way. *The Astrophysical Journal*, 787:147, 2014.
- [115] R. Fraedrich, J. Schneider, and R. Westermann. Exploring the millennium run-scalable rendering of large-scale cosmological datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1251–1258, 2009.
- [116] T. Freeth, Y. Bitsakis, X. Moussas, J. H. Seiradakis, A. Tselikas, H. Mangou, M. Zafeiropoulou, R. Hadland, D. Bate, A. Ramsey, M. Allen, A. Crawley, P. Hockley, T. Malzbender, D. Gelb, W. Ambrisco, and M. G. Edmunds. Decoding the ancient Greek astronomical calculator known as the Antikythera Mechanism. *Nature*, 444(7119):587–591, Nov. 2006. Number: 7119 Publisher: Nature Publishing Group. doi: 10.1038/nature05357
- [117] M. Friendly. A Brief History of Data Visualization. In *Handbook of Data Visualization*, pp. 15–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-33037-0_2
- [118] L. Fritschi, I. B. Rojo, and T. Günther. Visualizing the temporal evolution of the universe from cosmology simulations. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)*, pp. 1–11, 2019.
- [119] C. L. Fryer. Using Computers to See Inside Supernovae.

- [120] C.-W. Fu and A. J. Hanson. A transparently scalable visualization architecture for exploring the Universe. *IEEE Transactions on Visualization and Computer Graphics*, 13(1):108–121, 2007.
- [121] I. Fujishiro, N. Sawada, M. Nakayama, H.-Y. Wu, K. Watanabe, S. Takahashi, and M. Uemura. Timetubes: Visual exploration of observed blazar datasets. In *Journal of Physics: Conference Series*, vol. 1036, p. 012011, 2018.
- [122] H. G. Funkhouser. A Note on a Tenth Century Graph. *Osiris*, 1:260–262, 1936. Publisher: [Saint Catherines Press, The University of Chicago Press, The History of Science Society].
- [123] P. Galison. Minkowski’s Space-Time: From Visual Thinking to the Absolute World. p. 33.
- [124] A. Geller and A. Gurvich. Firefly: A WebGL tool to explore particle-based data, 2021. <https://agurvich.github.io/Firefly/>. Accessed March 30th, 2021.
- [125] M. J. Geller and J. P. Huchra. Mapping the Universe. *Science*, 246:897–903, 1989.
- [126] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder, V. Rodriguez-Gomez, P. Torrey, and L. Hernquist. Introducing the illustris project: the evolution of galaxy populations across cosmic time. *Monthly Notices of the Royal Astronomical Society*, 445(1):175–200, 2014.
- [127] J. Gray and A. S. Szalay. The World Wide Telescope: An archetype for online science. *Communications of the ACM*, 45(11):50–54, 2002.
- [128] R. Gray and J. Dunning-Davies. A review of redshift and its interpretation in cosmology and astrophysics. June 2008.
- [129] J. C. Green, C. S. Froning, S. Osterman, D. Ebbets, S. H. Heap, C. Leitherer, J. L. Linsky, B. D. Savage, K. Sembach, J. Shull, and et al. The Cosmic Origins Spectrograph. *The Astrophysical Journal*, 744:60, 2012.

- [130] E. Greisen. AIPS, the VLA, and the VLBA. In *Astrophys. Space Sci. Library*, vol. 285, pp. 109–125. May 2006. Journal Abbreviation: Astrophys. Space Sci. Library. doi: 10.1007/0-306-48080-8_7
- [131] J. Guedes, S. Callegari, P. Madau, and L. Mayer. Forming Realistic Late-type Spirals in a Λ CDM Universe: The Eris Simulation. *The Astrophysical Journal*, 742(2):76, 2011.
- [132] R. Haber. Visualization idioms: A conceptual model for scientific visualization systems. *Visualization in Scientific Computing*, p. 21, 1990.
- [133] S. Habib, A. Pope, H. Finkel, N. Frontiere, K. Heitmann, D. Daniel, et al. HACC: Simulating sky surveys on state-of-the-art supercomputing architectures. *New Astronomy*, 42:49–65, 2016.
- [134] D. Y. Hancock, J. Fischer, J. M. Lowe, W. Snapp-Childs, M. Pierce, S. Marru, J. E. Coulter, M. Vaughn, B. Beck, N. Merchant, et al. Jetstream2: Accelerating cloud computing via jetstream. *Practice and Experience in Advanced Research Computing*, pp. 1–8, 2021.
- [135] P. Hanula, K. Piekutowski, C. Uribe, K. Almryde, A. Nishimoto, J. Aguilera, and G. E. Marai. Cavern Halos: Exploring spatial and nonspatial cosmological data in an immersive virtual environment. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)*, pp. 87–99. IEEE, 2015.
- [136] S. Haroz, K.-L. Ma, and K. Heitmann. Multiple uncertainties in time-variant cosmological particle data. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 207–214, 2008.
- [137] I. Hayes. Copperplate Engravings, 2019.
- [138] D. Hayton. An Introduction to the Astrolabe. p. 32.
- [139] S. Hazarika, T.-H. Wei, R. Mukherjee, and A. Barbur. Visualizing the life and anatomy of dark matter. In *2015 IEEE Scientific Visualization Conference (SciVis)*, pp. 101–106. IEEE, 2015.

- [140] C. Hesse-Edenfeld, M. J. Steinke, N. Santalidis, K. Huesmann, S. Leistikow, and L. Linsen. Interactive multi-level visualization of combined particle and multi-field cosmology data. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)*, pp. 12–30. IEEE, 2019.
- [141] L. Hogben, J. Huxley, R. A. W. Watt, and B. A. Keen. Films and Science. *Sight and Sound*, 5(20):150–151, 1936. Num Pages: 2 Place: London, United Kingdom Publisher: British Film Institute.
- [142] P. F. Hopkins. A new class of accurate, mesh-free hydrodynamic simulation methods. *Monthly Notices of the Royal Astronomical Society*, 450(1):53–110, 2015.
- [143] P. F. Hopkins, A. Wetzel, D. Kereš, C.-A. Faucher-Giguère, E. Quataert, M. Boylan-Kolchin, et al. FIRE-2 simulations: Physics versus numerics in galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 480(1):800–863, 2018.
- [144] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Science*, 15:168–173, 1929.
- [145] D. W. Hughes and C. Stott. The planisphere: a brief historical review. *Journal of the British Astronomical Association*, 105:35–39, Feb. 1995. ADS Bibcode: 1995JBAA..105...35H.
- [146] C. B. Hummels, B. D. Smith, P. F. Hopkins, B. W. O’Shea, D. W. Silvia, J. K. Werk, et al. The impact of enhanced halo resolution on the simulated circumgalactic medium. *The Astrophysical Journal*, 882(2):156, 2019.
- [147] C. B. Hummels, B. D. Smith, and D. W. Silvia. Trident: A universal tool for generating synthetic absorption spectra from astrophysical simulations. *The Astrophysical Journal*, 847:59, 2017.
- [148] Z.-H. Hwang, H.-S. Yan, and T.-Y. Lin. Historical development of water-powered mechanical clocks. *Mechanical Sciences*, 12(1):203–219, Feb. 2021. doi: 10.5194/ms-12-203-2021

- [149] V. Icke. Formation of galaxies inside clusters. *Astronomy and Astrophysics*, 27:1–21, 1973.
- [150] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the ACM Workshop on BEyond time and errors: novel evaluation methods for Information Visualization*, p. 6, 2008.
- [151] S. D. Johnson, H.-W. Chen, and J. S. Mulchaey. On the possible environmental effect in distributing heavy elements beyond individual gaseous haloes. *Monthly Notices of the Royal Astronomical Society*, 449:3263–3273, 2015.
- [152] S. F. Johnston. Reconstructing the history of holography. p. 455. Santa Clara, CA, May 2003. doi: 10.1117/12.478431
- [153] A. Jones. *A Portable Cosmos: Revealing the Antikythera Mechanism, Scientific Wonder of the Ancient World*. Oxford University Press, Jan. 2017. Google-Books-ID: SF3ODQAAQBAJ.
- [154] G. G. Kacprzak, J. R. Vander Vliet, N. M. Nielsen, S. Muzahid, S. K. Pointon, C. W. Churchill, et al. The relation between galaxy ISM and circumgalactic O VI gas kinematics derived from observations and Λ CDM simulations. *The Astrophysical Journal*, 870(2):137, Jan. 2019.
- [155] R. C. Kennicutt and N. J. Evans. Star formation in the Milky Way and nearby galaxies. *Annual Review of Astronomy and Astrophysics*, 50:531–608, 2012.
- [156] R. C. Kennicutt Jr. Star formation in galaxies along the hubble sequence. *Annual Review of Astronomy and Astrophysics*, 36(1):189–231, 1998.
- [157] L. J. Kewley and M. A. Dopita. Using strong lines to estimate abundances in extragalactic H II regions and starburst galaxies. *The Astrophysical Journal Supplement Series*, 142:35–52, 2002.
- [158] L. J. Kewley, D. C. Nicholls, and R. S. Sutherland. Understanding galaxy evolution through emission lines. *Annual Review of Astronomy and Astrophysics*, 57:511–570, 2019.

- [159] S. Khan, M. El-Gomati, M. Abattouy, and D. S. Ayduz. The Armillary Sphere: A Concentrate of Knowledge in Islamic Astronomy. p. 14, 2007.
- [160] R. P. Kirshner. Supernovae, an accelerating universe and the cosmological constant. *Proceedings of the National Academy of Sciences*, 96(8):4224–4227, Apr. 1999. doi: 10.1073/pnas.96.8.4224
- [161] A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand, et al. Haloes gone mad: the halo-finder comparison project. *Monthly Notices of the Royal Astronomical Society*, 415(3):2293–2318, 2011.
- [162] S. Kwok. *Our Place in the Universe: Understanding Fundamental Astronomy from Ancient Discoveries*. Springer, 2017.
- [163] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
- [164] F. Lan, M. Young, L. Anderson, A. Ynnerman, A. Bock, M. A. Borkin, et al. Visualization in astrophysics: Developing new methods, discovering our universe, and educating the earth. *Computer Graphics Forum*, 40, 2021.
- [165] E. Laplace. Tulips: A tool for understanding the lives, interiors, and physics of stars. *Astronomy and Computing*, 38:100516, 2022.
- [166] R. B. Larson. Numerical Calculations of the Dynamics of a Collapsing Proto-Star*. *Monthly Notices of the Royal Astronomical Society*, 145(3):271–295, Aug. 1969. doi: 10.1093/mnras/145.3.271
- [167] R. B. Larson, B. M. Tinsley, and C. N. Caldwell. The evolution of disk galaxies and the origin of S0 galaxies. *The Astrophysical Journal*, 237:692–707, 1980.
- [168] K.-G. Lee, A. Krolewski, M. White, D. Schlegel, P. E. Nugent, J. F. Hennawi, T. Müller, R. Pan, J. X. Prochaska, A. Font-Ribera, and et al.

- First data release of the COSMOS Ly α mapping and tomography observations: 3D Ly α forest tomography at $2.05 < z < 2.55$. *The Astrophysical Journal Supplement Series*, 237:31, 2018.
- [169] N. Lehner, J. M. O’Meara, A. J. Fox, J. C. Howk, J. X. Prochaska, V. Burns, and A. A. Armstrong. Galactic and circumgalactic O VI and its impact on the cosmological metal and baryon budgets at $2 < z < \sim 3.5$. *The Astrophysical Journal*, 788:119, 2014.
- [170] N. Lehner, B. D. Savage, P. Richter, K. R. Sembach, T. M. Tripp, and B. P. Wakker. Physical properties, baryon content, and evolution of the Ly α forest: New insights from high-resolution observations at $z < \sim 0.4$. *The Astrophysical Journal*, 658:680–709, 2007.
- [171] S. N. Leitner and A. V. Kravtsov. Fuel efficient galaxies: Sustaining star formation with stellar mass loss. *The Astrophysical Journal*, 734(1):48, 2011.
- [172] N. I. Libeskind, R. van de Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel, et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473:1195–1217, 2018.
- [173] S. Lopez, N. Tejos, C. Ledoux, L. F. Barrientos, K. Sharon, J. R. Rigby, M. D. Gladders, M. B. Bayliss, and I. Pessa. A clumpy and anisotropic galaxy halo at redshift 1 from gravitational-arc tomography. *Nature*, 554:493–496, 2018.
- [174] P. Lu. Bhāskara I on the Construction of the Armillary Sphere. *History of Science in South Asia*, 3:1–19, Feb. 2015. doi: 10.18732/H2MW2W
- [175] B. P. Luce. Modeling and analysis of the chaco canyon sun dagger site. *Archaeoastronomy*, 23, 2010.
- [176] T. B. Luciani, B. Cherinka, D. Oliphant, S. Myers, W. M. Wood-Vasey, A. Labrinidis, and G. E. Marai. Large-scale overlays and trends: Visually mining, panning and zooming the observable Universe. *IEEE Transactions on Visualization and Computer Graphics*, 20(7):1048–1061, 2014.

- [177] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao. Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18(10):1039–1065, 2006.
- [178] C. Ma, A. G. Forbes, D. A. Llano, T. Berger-Wolf, and R. V. Kenyon. SwordPlots: Exploring neuron behavior within dynamic communities of brain networks. *Journal of Imaging Science and Technology*, 60(1):10405–1–13, 2016.
- [179] G. Magli. At the other end of the sun’s path: A new interpretation of machu picchu. In *Recalling Eero Saarinen 1910–2010*, pp. 321–341. Springer, 2010.
- [180] A. Man and S. Belli. Star formation quenching in massive galaxies. *Nature Astronomy*, 2(9):695–697, 2018.
- [181] N. Mandelker, F. C. van den Bosch, V. Springel, and F. van de Voort. Shattering of Cosmic Sheets due to Thermal Instabilities: A Formation Channel for Metal-free Lyman Limit Systems. , 881(1):L20, Aug. 2019.
- [182] N. Mandelker, F. C. van den Bosch, V. Springel, F. van de Voort, J. N. Burchett, I. S. Butsky, D. Nagai, and S. P. Oh. Thermal Instabilities and Shattering in the High-Redshift WHIM: Convergence Criteria and Implications for Low-Metallicity Strong HI Absorbers. *arXiv e-prints*, p. arXiv:2107.03395, July 2021.
- [183] J. Marche. *Theaters of Time and Space: American Planetaria, 1930-1970*. Rutgers University Press, June 2005. Google-Books-ID: fA9YJxUKjMkC.
- [184] K. Mars. 410 Years Ago: Galileo Discovers Jupiter’s Moons, Jan. 2020.
- [185] D. Martizzi, M. Vogelsberger, M. C. Artale, M. Haider, P. Torrey, F. Marinacci, D. Nelson, A. Pillepich, R. Weinberger, L. Hernquist, J. Naiman, and V. Springel. Baryons in the Cosmic Web of IllustrisTNG - I: gas in knots, filaments, sheets, and voids. *Monthly Notices of the Royal Astronomical Society*, 486(3):3766–3787, 2019.

- [186] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.
- [187] S. McAlpine, J. C. Helly, M. Schaller, J. W. Trayford, Y. Qu, M. Furlong, et al. The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89, 2016.
- [188] T. McPhillips, S. Bowers, D. Zinn, and B. Ludäscher. Scientific workflow design for mere mortals. *Future Generation Computer Systems*, 25(5):541–551, 2009.
- [189] E. Meijering. A chronology of interpolation: from ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3):319–342, 2002.
- [190] J. R. Millburn. William Stukeley and the early history of the Orrery. *Annals of Science*, 31(6):511–528, Nov. 1974. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/00033797400200431>. doi: 10.1080/00033797400200431
- [191] K. P. Moesgaard. Ancient ephemeris time in babylonian astronomy. *Journal for the History of Astronomy*, 14(1):47–60, 1983.
- [192] J. J. Monaghan. Smoothed particle hydrodynamics. *Annual Review of Astronomy and Astrophysics*, 30(1):543–574, 1992.
- [193] A. Morrow. Hubble Team Breaks Cosmic Distance Record, Mar. 2016.
- [194] J. Naiman. AstroBlend: An astrophysical visualization package for Blender. *Astronomy and Computing*, 15:50–60, 2016.
- [195] NASA-Sloan Atlas. Data release 15, v1_0_1, 2019. <https://sdss.org/dr15/manga/manga-target-selection/nsa/>.
- [196] O. W. Nasim. *Observing by Hand: Sketching the Nebulae in the Nineteenth Century*. University of Chicago Press, Chicago, UNITED STATES, 2014.
- [197] E. National Academies of Sciences, Medicine, et al. *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. 2021.

- [198] D. Nelson, G. Kauffmann, A. Pillepich, S. Genel, V. Springel, R. Pakmor, et al. The abundance, distribution, and physical nature of highly ionized oxygen O VI, O VII, and O VIII in IllustrisTNG. *Monthly Notices of the Royal Astronomical Society*, 477(1):450–479, 2018.
- [199] D. Nelson, A. Pillepich, V. Springel, R. Pakmor, R. Weinberger, S. Genel, et al. First results from the TNG50 simulation: Galactic outflows driven by supernovae and black hole feedback. *Monthly Notices of the Royal Astronomical Society*, 490(3):3234–3261, 2019.
- [200] S. Nelson. The Harvard computers. *Nature*, 455(7209):36–37, Sept. 2008. Number: 7209 Publisher: Nature Publishing Group. doi: 10.1038/455036a
- [201] P. Neophytou, R. Gheorghiu, R. Hachey, T. Luciani, D. Bao, A. Labrinidis, G. E. Marai, and P. K. Chrysanthis. AstroShelf: Understanding the Universe through scalable navigation of a galaxy of annotations. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 713–716, 2012.
- [202] O. Neugebauer. The History of Ancient Astronomy Problems and Methods. *Journal of Near Eastern Studies*, 4(1):1–38, 1945. Publisher: University of Chicago Press.
- [203] B. D. Nguyen, N. V. Nguyen, V. Pham, and T. Dang. Visualization of data from hacc simulations by paraview. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)*, pp. 31–32, 2019.
- [204] R. P. Norris, C. Norris, D. W. Hamacher, and R. Abrahams. Wurdiyouang: an australian aboriginal stone arrangement with possible solar indications. *Rock Art Research: The Journal of the Australian Rock Art Research Association (AURA)*, 30(1):55–65, 2013.
- [205] J. M. O’Meara, N. Lehner, J. C. Howk, J. X. Prochaska, A. J. Fox, M. A. Swain, C. R. Gelino, G. B. Berriman, and H. Tran. The first data release of the KODIAQ Survey. *The Astronomical Journal*, 150:111, 2015.

- [206] B. D. Oppenheimer, J. J. Davies, R. A. Crain, N. A. Wijers, J. Schaye, J. K. Werk, et al. Feedback from supermassive black holes transforms centrals into passive galaxies by ejecting circumgalactic gas. *Monthly Notices of the Royal Astronomical Society*, 491(2):2939–2952, 2019.
- [207] N. Owano. Scotland lunar-calendar find sparks stone age rethink. *Phys.org*, 27, 2013.
- [208] A. Pasquali and V. Nachname. How environment drives galaxy evolution: Lessons learnt from satellite galaxies. *Astronomische Nachrichten*, 336(5):505–510, 2015.
- [209] B. Pathé. The First Hundred Years!, 1939.
- [210] B. Paxton, L. Bildsten, A. Dotter, F. Herwig, P. Lesaffre, and F. Timmes. Modules for experiments in stellar astrophysics (mesa). *The Astrophysical Journal Supplement Series*, 192(1):3, 2010.
- [211] P. J. E. Peebles. *Principles of Physical Cosmology*. Princeton University Press, 1993.
- [212] M. Peeples, J. Tumlinson, A. Fox, A. Aloisi, S. Fleming, R. Jedrzejewski, C. Oliveira, T. Ayres, C. Danforth, B. Keeney, et al. The hubble spectroscopic legacy archive. *Instrument Science Report COS*, 4, 2017.
- [213] M. S. Peeples, L. Corlies, J. Tumlinson, B. W. O’Shea, N. Lehner, J. M. O’Meara, J. C. Howk, N. Earl, B. D. Smith, and J. H. Wise. Figuring Out Gas & Galaxies in Enzo (FOGGIE). I. Resolving Simulated Circumgalactic Absorption at $2 \leq z \leq 2.5$. , 873(2):129, Mar 2019.
- [214] J.-J. Peters. The History of Television. *European Broadcasting Union*, p. 28, 2000.
- [215] A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, et al. Simulating galaxy formation with the IllustrisTNG model. *Monthly Notices of the Royal Astronomical Society*, 473(3):4077–4106, 2018.

- [216] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, vol. 5, pp. 2–4, 2005.
- [217] F. Pirotti, M. A. Brovelli, G. Prestifilippo, G. Zamboni, C. E. Kilsedar, M. Piragnolo, and P. Hogan. An open source virtual globe rendering engine for 3d applications: Nasa world wind. *Open Geospatial Data, Software and Standards*, 2(1):1–14, 2017.
- [218] S. K. Pointon, N. M. Nielsen, G. G. Kacprzak, S. Muzahid, C. W. Churchill, and J. C. Charlton. The impact of the group environment on the O VI circumgalactic medium. *The Astrophysical Journal*, 844:23, 2017.
- [219] D. Pomarède, H. M. Courtois, Y. Hoffman, and R. B. Tully. Cosmography and data visualization. *Publications of the Astronomical Society of the Pacific*, 129(975):058002, 2017.
- [220] A. Pontzen, R. Roškar, G. S. Stinson, R. Woods, D. M. Reed, J. Coles, and T. R. Quinn. pynbody: Astrophysics Simulation Analysis for Python, 2013. Astrophysics Source Code Library, ascl:1305.002.
- [221] U. Popov, E. Chandra, K. Heitmann, S. Habib, J. Ahrens, and A. Pang. Analyzing the evolution of large scale structures in the Universe with velocity based methods. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 49–56, 2012.
- [222] A. Preston, R. Ghods, J. Xie, F. Sauer, N. Leaf, K.-L. Ma, E. Rangel, E. Kovacs, K. Heitmann, and S. Habib. An integrated visualization system for interactive analysis of large, heterogeneous cosmology data. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 48–55, 2016.
- [223] D. J. Price. A collection of armillary spheres and other antique scientific instruments. *Annals of Science*, 10(2):172–187, June 1954. doi: 10.1080/00033795400200204

- [224] J. X. Prochaska, B. Weiner, H.-W. Chen, J. Mulchaey, and K. Cooksey. Probing the intergalactic medium/galaxy connection. V. On the origin of Ly α and O VI absorption at $z < 0.2$. *The Astrophysical Journal*, 740:91, 2011.
- [225] D. Punzo, J. Van der Hulst, J. Roerdink, T. Oosterloo, M. Ramatsoku, and M. Verheijen. The role of 3-D interactive visualization in blind surveys of HI in galaxies. *Astronomy and Computing*, 12:86–99, 2015.
- [226] A. Rahmati, A. H. Pawlik, M. Raičević, and J. Schaye. On the evolution of the HI column density distribution in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 430(3):2427–2445, 2013.
- [227] A. Rahmati, J. Schaye, R. A. Crain, B. D. Oppenheimer, M. Schaller, and T. Theuns. Cosmic distribution of highly ionized metals and their physical conditions in the eagle simulations. *Monthly Notices of the Royal Astronomical Society*, 459(1):310–332, 2016.
- [228] P. M. P. Raposo. ‘Moving the universe’: the Atwood Sphere and the Adler Planetarium in Chicago. *Early Popular Visual Culture*, 15(2):268–272, Apr. 2017. Publisher: Routledge _eprint: <https://doi.org/10.1080/17460654.2017.1318518>. doi: 10.1080/17460654.2017.1318518
- [229] J. Rasmussen, J. Sommer-Larsen, K. Pedersen, S. Toft, A. Benson, R. G. Bower, and L. F. Grove. Hot gas halos around disk galaxies: Confronting cosmological simulations with observations. *The Astrophysical Journal*, 697(1):79, 2009.
- [230] M. Rauch. The Lyman alpha forest in the spectra of quasistellar objects. *Annual Review of Astronomy and Astrophysics*, 36:267–316, 1998.
- [231] P.-K. H. Robert, R. H. Thibadeau, C. B. Cox, and R. H. P. Dunn. Time Dilation Visualization in Relativity, 1990.
- [232] B. Robertson, S. Tacchella, B. Johnson, K. Hainline, L. Whitler, D. Eisenstein, R. Endsley, M. Rieke, D. Stark, S. Alberts, et al. Identifi-

- cation and properties of intense star-forming galaxies at redshifts $z > 10$. *Nature Astronomy*, pp. 1–11, 2023.
- [233] B. E. Robertson, R. S. Ellis, S. R. Furlanetto, and J. S. Dunlop. Cosmic reionization and early star-forming galaxies: A joint analysis of new constraints from planck and the hubble space telescope. *The Astrophysical Journal Letters*, 802(2):L19, 2015.
- [234] P. Rosenfield, J. Fay, R. K. Gilchrist, C. Cui, A. D. Weigel, T. Robitaille, O. J. Otor, and A. Goodman. AAS WorldWide Telescope: A seamless, cross-platform data visualization engine for astronomy research, education, and democratizing data. *The Astrophysical Journal Supplement Series*, 236(1):22, 2018.
- [235] K. H. R. Rubin, J. M. O’Meara, K. L. Cooksey, M. Matuszewski, L. Rizzi, G. Doppmann, S. Kwok, D. C. Martin, A. M. Moore, P. Morrissey, and J. D. Neill. Andromeda’s Parachute: A bright quadruply lensed quasar at $z = 2.377$. *The Astrophysical Journal*, 859:146, 2018.
- [236] R. Rudenko, I. M. Pires, M. Liberato, J. Barroso, and A. Reis. A brief review on 4d weather visualization. *Sustainability*, 14(9):5248, 2022.
- [237] C. Ruggles, B. Cunliffe, and C. Renfrew. Astronomy and stonehenge. In *Proceedings-British Academy*, vol. 92, pp. 203–230. OXFORD UNIVERSITY PRESS INC., 1997.
- [238] D. S. N. Rupke, A. Coil, J. E. Geach, C. Tremonti, A. M. Diamond-Stanic, E. R. George, et al. A 100-kiloparsec wind feeding the circumgalactic medium of a massive compact galaxy. *Nature*, 574(7780):643–646, 2019.
- [239] A. Sagristà, S. Jordan, T. Müller, and F. Sadlo. Gaia Sky: Navigating the Gaia catalog. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1070–1079, 2019.
- [240] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692, 2019.

- [241] N. Sawada, M. Uemura, J. Beyer, H. Pfister, and I. Fujishiro. Time-tubesx: a query-driven visual exploration of observable, photometric, and polarimetric behaviors of blazars. *IEEE Transactions on Visualization and Computer Graphics*, 2020.
- [242] N. Sawada, M. Uemura, and I. Fujishiro. Multi-dimensional time-series subsequence clustering for visual feature analysis of blazar observation datasets. *Astronomy and Computing*, 41:100663, 2022.
- [243] K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, and T. Ertl. Interactive visual exploration of a trillion particles. In *Proceedings of the IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 56–64, 2016.
- [244] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, et al. The EAGLE project: Simulating the evolution and assembly of galaxies and their environments. *Monthly Notices of the Royal Astronomical Society*, 446(1):521–554, 2015.
- [245] A. Scherzinger, T. Brix, D. Drees, A. Völker, K. Radkov, N. Santalidis, A. Fieguth, and K. H. Hinrichs. Interactive exploration of cosmological dark-matter simulation data. *IEEE Computer Graphics and Applications*, 37(2):80–89, 2017.
- [246] M. Schmidt. Redshifts of fourteen quasi-stellar radio sources. *The Astrophysical Journal*, 144:443, 1966.
- [247] W. Schroeder, K. M. Martin, and W. E. Lorenson. *The visualization toolkit an object-oriented approach to 3D graphics*. Prentice-Hall, Inc., 1998.
- [248] D. Segaller. The Development of the Scientific Film in Great Britain. *Sight and Sound*, 16(64):140–143, 1947. Num Pages: 4 Place: London, United Kingdom Publisher: British Film Institute.
- [249] M. C. Segers, R. A. Crain, J. Schaye, R. G. Bower, M. Furlong, M. Schaller, and T. Theuns. Recycled stellar ejecta as fuel for star formation and implications for the origin of the galaxy mass–metallicity re-

- lation. *Monthly Notices of the Royal Astronomical Society*, 456(2):1235–1258, 2016.
- [250] J. Seifert and F. Lemke. The sumerian k8538 tablet. the great meteor impact devastating mesopotamia-a 2019 translation addendum, 2019, hwp, 2019.
- [251] G. Shan, M. Xie, F. Li, Y. Gao, and X. Chi. Interactive visual exploration of halos in large-scale cosmology simulation. *Journal of Visualization*, 17(3):145–156, 2014.
- [252] D. Sijacki, M. Vogelsberger, S. Genel, V. Springel, P. Torrey, G. F. Snyder, D. Nelson, and L. Hernquist. The illustris simulation: the evolving population of black holes across cosmic time. *Monthly Notices of the Royal Astronomical Society*, 452(1):575–596, 2015.
- [253] S. W. Skillman, M. S. Warren, M. J. Turk, R. H. Wechsler, D. E. Holz, and P. Sutter. Dark sky simulations: Early data release. *arXiv preprint arXiv:1407.2600*, 2014.
- [254] N. N. Smit, B. K. Haneveld, M. Staring, E. Eisemann, C. P. Botha, and A. Vilanova. RegistrationShop: An interactive 3d medical volume registration system. In *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*, pp. 145–153, 2014.
- [255] B. D. Smith and M. Lang. ytree: A python package for analyzing merger trees. *Journal of Open Source Software*, 4(44):1881, 2019.
- [256] R. A. Society. 3-D holograms bringing astronomy to life.
- [257] N. Sofroniew, T. Lambert, K. Evans, J. Nunez-Iglesias, G. Bokota, P. Winston, et al. napari: Multi-dimensional image viewer for python, 2021. <https://napari.org/>, Accessed September 18, 2021.
- [258] A. Songaila. The minimum universal metal density between redshifts of 1.5 and 5.5. *The Astrophysical Journal Letters*, 561:L153–L156, 2001.
- [259] I. Spence and R. F. Garrison. A Remarkable Scatterplot. *The American Statistician*, 47(1):12–19,

Feb. 1993. Publisher: Taylor & Francis _eprint:
<https://www.tandfonline.com/doi/pdf/10.1080/00031305.1993.10475925>.
doi: 10.1080/00031305.1993.10475925

- [260] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha, and G. E. Robinson. Big data: Astronomical or genetical? *PLoS Biology*, 13(7):e1002195, 2015.
- [261] J. T. Stocke, B. A. Keeney, C. W. Danforth, J. M. Shull, C. S. Froning, J. C. Green, S. V. Penton, and B. D. Savage. Characterizing the circumgalactic medium of nearby galaxies with HST/COS and HST/STIS absorption-line spectroscopy. *The Astrophysical Journal*, 763:148, 2013.
- [262] M. Streit and N. Gehlenborg. Temporal data: use inherent properties of time to create effective visualizations. *Nature Methods*, 12(2):97–98, 2015.
- [263] N. Sukumar. Voronoi cell finite difference method for the diffusion operator on arbitrary unstructured grids. *International Journal for Numerical Methods in Engineering*, 57(1):1–34, 2003.
- [264] X. Sun. Water-Powered Astronomical Clock Tower. In C. L. Ruggles, ed., *Handbook of Archaeoastronomy and Ethnoastronomy*, pp. 2133–2139. Springer, New York, NY, 2015. doi: 10.1007/978-1-4614-6141-8_221
- [265] M. B. Taylor. TOPCAT & STIL: Starlink table/VOTable processing software. In *Astronomical Data Analysis Software and Systems XIV*, vol. 347, p. 29, 2005.
- [266] R. Taylor. Visualizing three-dimensional volumetric data with an arbitrary coordinate system. *Publications of the Astronomical Society of the Pacific*, 129(972):028002, 2017.
- [267] N. Tejos, S. L. Morris, N. H. M. Crighton, T. Theuns, G. Altay, and C. W. Finn. Large-scale structure in absorption: gas within and around galaxy voids. *Monthly Notices of the Royal Astronomical Society*, 425:245–260, 2012.

- [268] N. Tejos, J. X. Prochaska, N. H. M. Crighton, S. L. Morris, J. K. Werk, T. Theuns, N. Padilla, R. M. Bielby, and C. W. Finn. Towards the statistical detection of the warm-hot intergalactic medium in intercluster filaments of the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 455:2662–2697, 2016.
- [269] C. Thom, J. Tumlinson, J. K. Werk, J. X. Prochaska, B. D. Oppenheimer, M. S. Peeples, T. M. Tripp, N. S. Katz, J. M. O’Meara, A. B. Ford, and et al. Not dead yet: Cool circumgalactic gas in the halos of early-type galaxies. *The Astrophysical Journal Letters*, 758:L41, 2012.
- [270] V. E. Thoren. New Light on Tycho’s Instruments. *Journal for the History of Astronomy*, 4(1):25–45, Feb. 1973. Publisher: SAGE Publications Ltd. doi: 10.1177/002182867300400105
- [271] B. M. Tinsley. Constraints on models for chemical evolution in the solar neighborhood. *The Astrophysical Journal*, 192:629–641, 1974.
- [272] A. Toomre and J. Toomre. Galactic Bridges and Tails. *The Astrophysical Journal*, 178:623, Dec. 1972. doi: 10.1086/151823
- [273] M. Tory and T. Moller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, 2004.
- [274] T. M. Tripp, L. Lu, and B. D. Savage. The relationship between galaxies and low-redshift weak Ly α absorbers in the directions of H1821+643 and PG 1116+215. *The Astrophysical Journal*, 508:200–231, 1998.
- [275] T. M. Tripp, B. D. Savage, and E. B. Jenkins. Intervening O VI quasar absorption systems at low redshift: A significant baryon reservoir. *The Astrophysical Journal Letters*, 534:L1–L5, 2000.
- [276] T. M. Tripp and L. Song. The 21 cm "Outer Arm" and the outer-galaxy high-velocity clouds: Connected by kinematics, metallicity, and distance. *The Astrophysical Journal*, 746:173, 2012.
- [277] E. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, 1997.

- [278] J. Tumlinson, M. S. Peeples, and J. K. Werk. The circumgalactic medium. *Annual Review of Astronomy and Astrophysics*, 55:389–432, 2017.
- [279] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, N. Katz, R. Davé, B. D. Oppenheimer, J. D. Meiring, A. B. Ford, et al. The COS-halos survey: Rationale, design, and a census of circumgalactic neutral hydrogen. *The Astrophysical Journal*, 777(1):59, 2013.
- [280] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, N. Katz, R. Davé, B. D. Oppenheimer, J. D. Meiring, A. B. Ford, J. M. O’Meara, M. S. Peeples, K. R. Sembach, and D. H. Weinberg. The COS-Halos survey: Rationale, design, and a census of circumgalactic neutral hydrogen. *The Astrophysical Journal*, 777:59, 2013.
- [281] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, D. H. Weinberg, M. S. Peeples, J. M. O’Meara, B. D. Oppenheimer, J. D. Meiring, N. S. Katz, R. Davé, A. B. Ford, and K. R. Sembach. The large, oxygen-rich halos of star-forming galaxies are a major reservoir of galactic metals. *Science*, 334:948, 2011.
- [282] M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman. yt: A multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series*, 192(9):1–16, 2010.
- [283] G. L. Turner. The History of Optical Instruments: A Brief Survey of Sources and Modern Studies. *History of Science*, 8(1):53–93, Mar. 1969. Publisher: SAGE Publications Ltd. doi: 10.1177/007327536900800103
- [284] J. Van De Sande, C. D. Lagos, C. Welker, J. Bland-Hawthorn, F. Schulze, R.-S. Remus, et al. The SAMI galaxy survey: Comparing 3d spectroscopic observations with galaxies from cosmological hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 484(1):869–891, 2019.

- [285] A. Van Helden. The Invention of the Telescope. *Transactions of the American Philosophical Society*, 67(4):1–67, 1977. Publisher: American Philosophical Society. doi: 10.2307/1006276
- [286] A. Van Helden. The beginnings, from Lipperhey to Huygens and Cassini. *Experimental Astronomy*, 25(1):3, Apr. 2009. doi: 10.1007/s10686-009-9160-y
- [287] S. Veilleux, G. Cecil, and J. Bland-Hawthorn. Galactic winds. *Annual Review of Astronomy and Astrophysics*, 43:769–826, 2005.
- [288] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509(7499):177–182, 2014.
- [289] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the illustris project: simulating the coevolution of dark and visible matter in the universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547, 2014.
- [290] M. Vogelsberger, D. Nelson, A. Pillepich, X. Shen, F. Marinacci, V. Springel, R. Pakmor, S. Tacchella, R. Weinberger, P. Torrey, et al. High-redshift jwst predictions from illustrisng: dust modelling and galaxy luminosity functions. *Monthly Notices of the Royal Astronomical Society*, 492(4):5167–5201, 2020.
- [291] B. P. Wakker, A. K. Hernandez, D. M. French, T.-S. Kim, B. D. Oppenheimer, and B. D. Savage. Nearby galaxy filaments and the Ly α forest: Confronting simulations and the UV background with observations. *The Astrophysical Journal*, 814(1):40, 2015.
- [292] J. Walcher, B. Groves, T. Budavári, and D. Dale. Fitting the integrated spectral energy distributions of galaxies. *Astrophysics and Space Science*, 331:1–52, 2011.

- [293] S. Walter. The Historical Origins of Spacetime. In A. Ashtekar and V. Petkov, eds., *Springer Handbook of Spacetime*, pp. 27–38. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. doi: 10.1007/978-3-642-41992-8_2
- [294] D. Weiskopf. A Survey of Visualization Methods for Special Relativity. p. 14 pages, 2010. Artwork Size: 14 pages Medium: application/pdf Publisher: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany. doi: 10.4230/DFU.SCIVIZ.2010.289
- [295] D. Weiskopf, M. Borchers, T. Ertl, M. Falk, O. Fechtig, R. Frank, F. Grave, A. King, U. Kraus, T. Muller, H.-P. Nollert, I. Mendez, H. Ruder, T. Schafhitzel, S. Schar, C. Zahn, and M. Zatloukal. Visualization in the Einstein Year 2005: a case study on explanatory and illustrative visualization of relativity and astrophysics. In *VIS 05. IEEE Visualization, 2005.*, pp. 583–590, Oct. 2005. doi: 10.1109/VISUAL.2005.1532845
- [296] D. C. Wells and E. W. Greisen. FITS: A Flexible Image Transport System. *International Workshop on Image Processing in Astronomy*, 1979.
- [297] J. K. Werk, J. X. Prochaska, C. Thom, J. Tumlinson, T. M. Tripp, J. M. O’Meara, and M. S. Peeples. The COS-Halos survey: An empirical description of metal-line absorption in the low-redshift circumgalactic medium. *The Astrophysical Journal Supplement Series*, 204:17, 2013.
- [298] W. G. Wesley. The Accuracy of Tycho Brahe’s Instruments. *Journal for the History of Astronomy*, 9(1):42–53, Feb. 1978. Publisher: SAGE Publications Ltd. doi: 10.1177/002182867800900102
- [299] J. Wild, I. Hook, C. Arridge, S. Badman, N. Case, D. A. Constable, A. Grocott, J. Kinrade, L. C. Ray, N. Rogers, B. Simmons, D. Sobral, J. Stott, M. Walach, J. Wardlow, and S. Williams. *National Astronomy Meeting 2019 Abstract Book*. Royal Astronomical Society, July 2019.

- [300] D. J. Wilman and P. Erwin. The relation between galaxy morphology and environment in the local universe: An rc3-sdss picture. *The Astrophysical Journal*, 746(2):160, 2012.
- [301] A. M. Wolfe, D. A. Turnshek, H. E. Smith, and R. D. Cohen. Damped Lyman-alpha absorption by disk galaxies with large redshifts. I - The Lick survey. *The Astrophysical Journal Supplement Series*, 61:249–304, 1986.
- [302] J. Woodman. The Women 'Computers' Who Revolutionized Astronomy, Dec. 2016. Section: Science.
- [303] J. Woodring, K. Heitmann, J. Ahrens, P. Fasel, C.-H. Hsu, S. Habib, and A. Pope. Analyzing and visualizing cosmological simulations with ParaView. *The Astrophysical Journal Supplement Series*, 195(1):11, 2011.
- [304] P. R. Woodward. Shock-driven implosion of interstellar gas clouds and star formation. *The Astrophysical Journal*, 207:484, July 1976. doi: 10.1086/154516
- [305] M. M. Woolfson. A New Tidal Theory for the Origin of the Solar System. *Quarterly Journal of the Royal Astronomical Society*, 19:167, June 1978. ADS Bibcode: 1978QJRAS..19..167W.
- [306] R. Xu, M. M. Thomas, A. Leow, O. A. Ajilore, and A. G. Forbes. Tempocave: Visualizing dynamic connectome datasets to support cognitive behavioral therapy. In *2019 IEEE Visualization Conference (VIS)*, pp. 186–190. IEEE, 2019.
- [307] J. H. Yoon and M. E. Putman. The influence of environment on the circumgalactic medium. *The Astrophysical Journal Letters*, 772:L29, 2013.
- [308] J. H. Yoon and M. E. Putman. Ly α absorbers and the coma Cluster. *The Astrophysical Journal*, 839:117, 2017.
- [309] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, and et al. The Sloan Digital Sky Survey: Technical summary. *The Astronomical Journal*, 120:1579–1587, 2000.

- [310] F. S. Zahedy, H.-W. Chen, M. Rauch, M. L. Wilson, and A. Zabludoff. Probing the cool interstellar and circumgalactic gas of three massive lensing galaxies at $z = 0.4-0.7$. *Monthly Notices of the Royal Astronomical Society*, 458:2423–2442, 2016.
- [311] S. Zaroubi. The epoch of reionization. *The first galaxies: theoretical predictions and observational clues*, pp. 45–101, 2012.
- [312] Y. B. Zeldovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and Astrophysics*, 5:84–89, 1970.
- [313] J. Zhang, Y.-T. Sung, H.-T. Hou, and K.-E. Chang. The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction. *Computers & Education*, 73:178–188, Apr. 2014. doi: 10.1016/j.compedu.2014.01.003
- [314] E. Zinger, A. Dekel, A. V. Kravtsov, and D. Nagai. Quenching of satellite galaxies at the outskirts of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 475(3):3654–3681, 2018.