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

David Abramov, Joseph N. Burchett, Oskar Elek, Cameron Hummels, J. Xavier Prochaska, and Angus G. Forbes

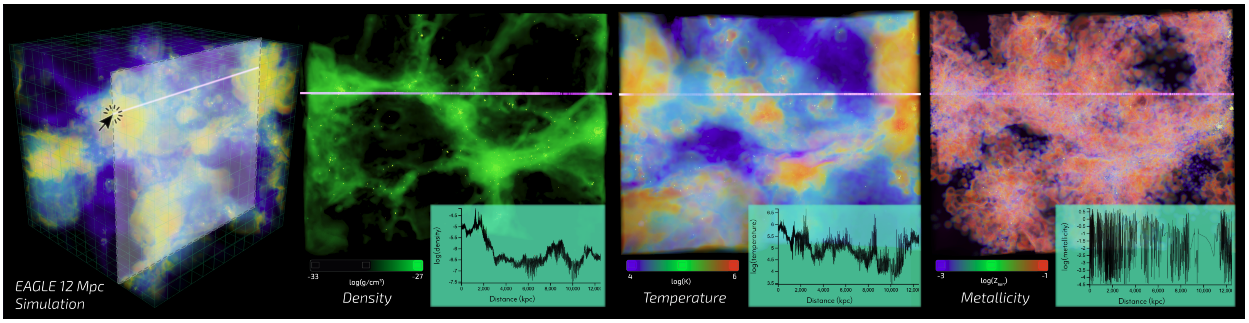


Fig. 1. Multiple volume renderings within the *CosmoVis* software illustrating the various types of gas attributes that can be retrieved from the cosmological simulations. Here, a user interactively places a "virtual skewer" within the *EAGLE* 12 Mpc simulation, as shown in the leftmost panel (temperature). 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.

**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 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.

**Index Terms**—Astrovis, astrographics, cosmological simulations, astronomy, astrophysics, virtual spectrography.

## 1 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*


- *David Abramov is with University of California, Santa Cruz. E-mail: dabramov@ucsc.edu.*
- *Joseph N. Burchett is with New Mexico State University. E-mail: jnb@nmsu.edu.*
- *Oskar Elek is with University of California, Santa Cruz. E-mail: oelek@ucsc.edu.*
- *Cameron Hummels is with California Institute of Technology. E-mail: chummels@gmail.com.*
- *J. Xavier Prochaska is with University of California, Santa Cruz. E-mail: xavier@ucolick.org.*
- *Angus G. Forbes is with University of California, Santa Cruz. E-mail: angus@ucsc.edu.*


*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx*

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 (Sect. 2 and Sect. 3).
- We delineate primary analysis tasks relevant for astrophysicists working with simulation datasets (Sect. 4);
- We provide a description of the feature set included in *CosmoVis*, along with a discussion of the iterative design process used in its development (Sect. 5);
- We describe the design and implementation of "virtual skewers", a novel tool that simulates the absorption line spectroscopy used in observational astronomy (Sect. 5.3);
- We provide an overview of the client-server architecture to interactive exploration and analysis of the simulation datasets via a web browser (Sect. 5.5);
- We introduce three detailed scientific use cases that document how *CosmoVis* has been used to gain new insight into contemporary research questions in astrophysics (Sect. 6).
- We summarize the response to *CosmoVis* from different communities of astrophysicists, and provide more detailed feedback from a pilot study (Sect. 7).

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.

## 2 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 (Mpcs) 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* [49, 66] and *IllustrisTNG* [58] are large-volume simulations, whereas *FIRE* [33], *Tempest* [35], and *FOGGIE* [57] 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. 2 shows an example of four 'big' boxes ranging from 12–100 Mpc along each edge.

Different simulations 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) [50] 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 [7] 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

### GLOSSARY

**Absorption line spectroscopy**— a widely-used observational technique for probing the spectral properties of gas in intragalactic environments and the intergalactic medium. Dozens of characteristic absorption signatures for specific element ions have been identified in astronomical spectra to date. 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 intervening material.

**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.

**Cosmic web**— the totality of knots, filaments, walls and voids, which together constitute the large-scale structure of the Universe. 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 [24, 37, 83] 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 [34]. The value $z = 0$ corresponds to the present day. Redshift results from the expansion of the space over which light emitted by a distant object at an earlier cosmic epoch travels towards an observer at present time [13]; 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/cm$^2$. 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)".

**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 $3 \times 10^{16}$ 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.

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

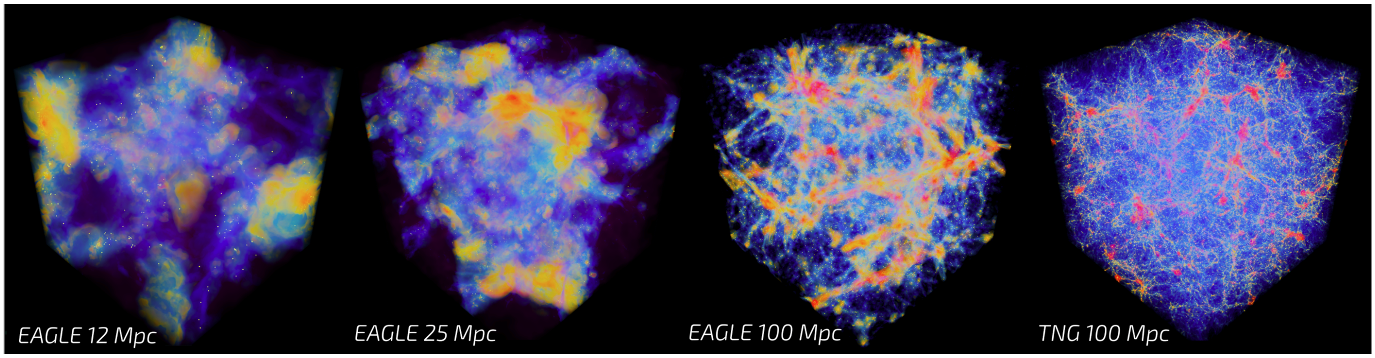


Fig. 2. Gas temperature volume renderings in simulations of various sizes. Snapshots from left to right: 12 Mpc *EAGLE* $z$=0.0; 25 Mpc *EAGLE* $z$=0.0; 100 Mpc *EAGLE* $z$=0.0; 100 Mpc *TNG100-1* $z$=2.3. While smaller volumes (left) can be easier to work with, larger simulations (right) provide a much greater sample size of various galaxy morphologies and environmental conditions.

grid cells at each timestep generated by a Voronoi tessellation [73] 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 $\sim$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 [34] 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 [13].

## 3 Related Work

A recent survey by Lan et al. [40] 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. 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 [2, 4] and Python libraries like *yt* [77] and *Napari* [72] support multiple visualization modalities, such as direct volume and particle rendering. Woodring et al. [82] introduces features for *ParaView* aimed at cosmological simulations, such as halo detection and analysis. The *AstroBlend* library [51] 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 [9] facilitates the interactive display of data from many various sources including simulation datasets, focusing on broader scientific communication. *Polyphorm* [25] 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. [31] 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. [67] 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 [58] includes its own interactive volume visualizer, and an interactive tool called *FIREFly* [27] supports the rendering of particles from *FIRE* datasets generated from the *GIZMO* code [32]. Work by Schatz et al. [65] 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 [70], which contains over a trillion particles. The *Hardware/Hybrid Accelerated Cosmology Code* (HACC) framework [30] 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. [54] and Fritschi et al. [26] 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* [15] 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"

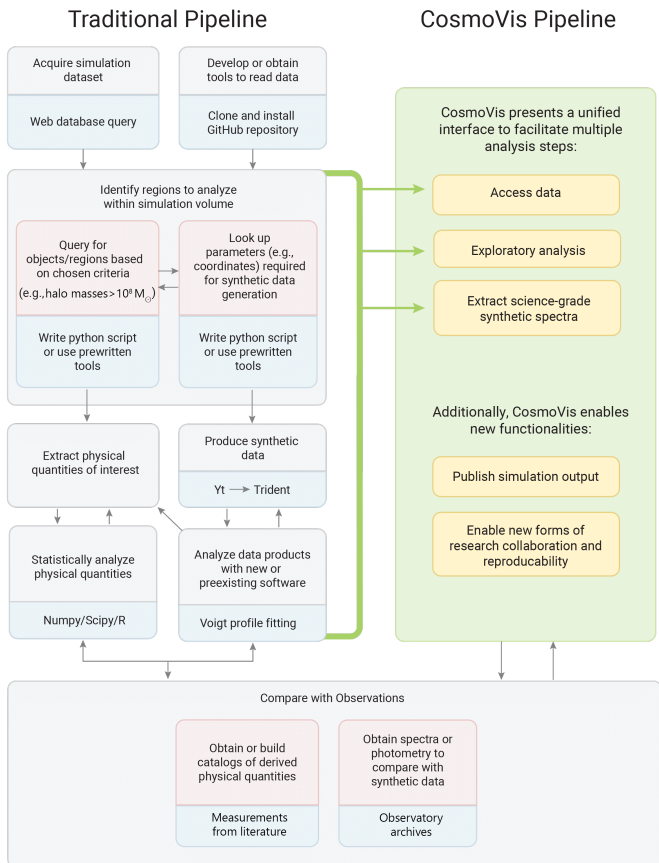


Fig. 3. Flowchart comparison between a traditional pipeline (left) versus the simplified *CosmoVis* pipeline (right) when working with these cosmological simulation datasets. In the traditional pipeline, in order to begin comparing a simulation with observations, a researcher would have to acquire the 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* [77] (used for data loading and processing) and *Trident* [36] (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 (Sect. 3), Analysis Task 2 (Sect. 4.2), Data Preprocessing (Sect. 5.1), Client-Server Architecture (Sect. 5.5), and Fig. 7.

(see Sect. 5.3) 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* [71] uses point and line elements in tomographic scans to annotate features of interest as well as to obtain tissue density profiles surrounding them.

## 4 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 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 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 counterintuitive 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?
- 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 **??**): 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**

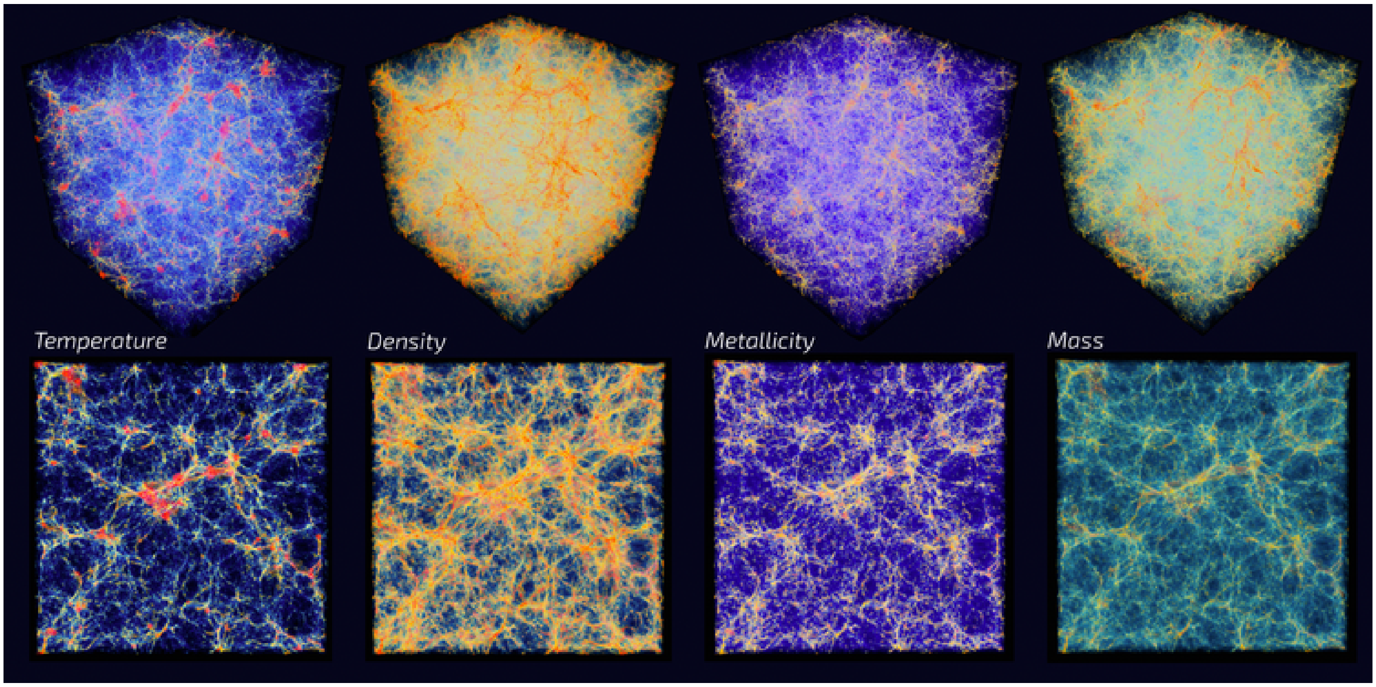


Fig. 4. This figure shows a series of interactive volume renders of gas fields created with the *CosmoVis* visualization tool. The data is extracted from the TNG100-1 z=2.32 cosmological simulation snapshot. The top row depicts the entire 100 Mpc volume from an angle, and the bottom row shows a head-on perspective of a 15 Mpc thick slice of the volume.

— *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.

### 4.1 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 [42, 47] 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 [45]. 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 [29]. 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 Sect. 6.1, provides an example of using *CosmoVis* to search for and identify properties of cosmic sheets.

### 4.2 Task 2: Simulate quasar absorption line spectroscopy

The circumgalactic environments of galaxies mediate the critical inflow and outflow processes that drive galaxy evolution [74]. 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 [17, 64]. 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 [5]. 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 [8, 18, 38, 75].

Unfortunately, individual galaxies that contain multiple QSO probes of their CGM are exceedingly rare [12, 43]. Due to this inherit limitation in observational data, cosmological simulations can play a pivotal role in testing hypotheses that would otherwise be impossible. The Python package *Trident* [36] 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 Sect. 6.2, provides an example of using *CosmoVis* to generate synthetic spectra to test the intrinsic variance of absorption properties within a galactic halo.

### 4.3 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 [56]. 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 [79]. 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 [10, 20]. 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 [60].

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 [1]. 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 [61]. 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 [62]. For example, studies examining five-times-ionized oxygen (O VI) in the *Illustris* [28, 69, 80, 81] and *EAGLE* [49] cosmological models found too low column densities as compared to observations, whereas the newer *IllustrisTNG* [58] simulations show more realistic distributions of highly ionized oxygen (O VI) [52].

Scientific Use Case 3 (Sect. 6.3) 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 (Sect. 6.1) and galaxy halos (Sect. 6.2).

### 4.4 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. [63] 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 [78]. A key consideration when comparing observational data and simulated data is taking into account observational biases. Donnari et al. [23] 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 galactice stellar mass content when compared to observations [22]. Therefore, star-formation driven feedback is crucial in producing galaxies with realistic masses [41]. 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* [33], 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. [68] 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 [41].

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 [6, 84]. 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 [44]. *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 exported 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.

## 5 Application Design of *CosmoVis*

The interdisciplinary *CosmoVis* development team is made up of visualization designers, graphics researchers, and astrophysicists. We

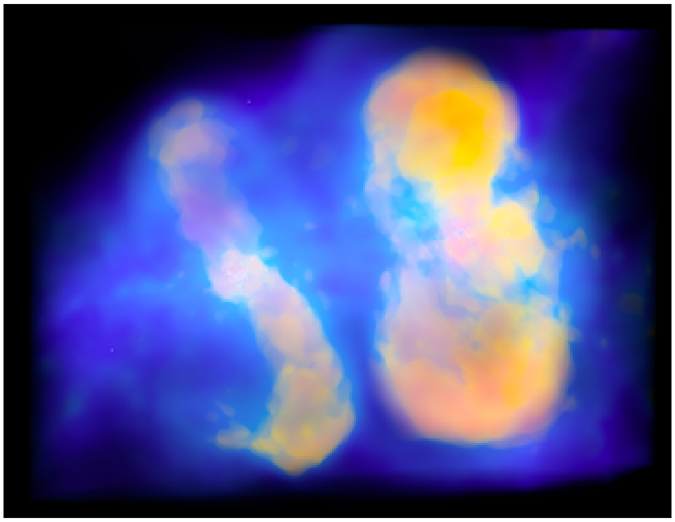

Fig. 5. Zoom in on a region containing two distinct circumgalactic environments in the 12 Mpc *EAGLE* simulation, with color representing temperature. Star particles are represented here as faint yellow disks at the center of these two adjacent systems. There appear to be visible galactic outflows of hot gas above and below the nuclei, indicated by the yellow-orange coloration.

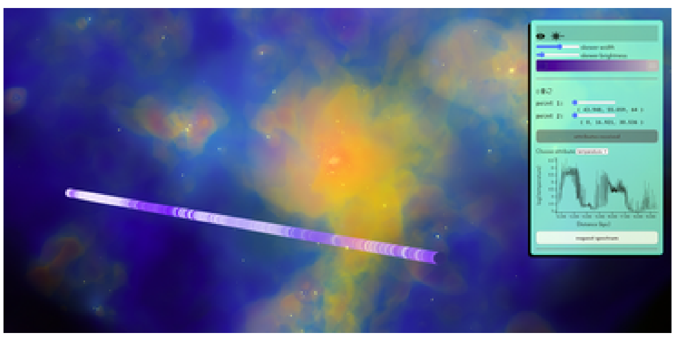

Fig. 6. *Temperature* profile along a virtual skewer. Using the default color map, blue indicates low temperatures ($10^4$ K) and red indicates high temperatures ($> 10^6$ K). These temperature values along the skewer are also displayed in the line plot in the *Observe* panel on the right, where the 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.

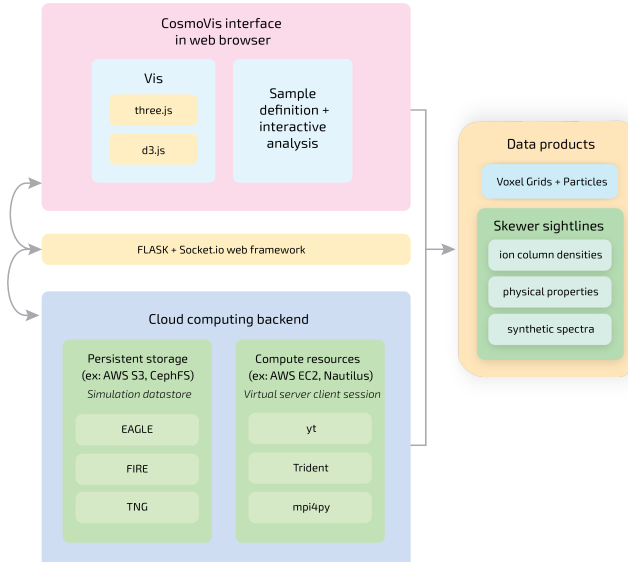


Fig. 7. Architecture of the *CosmoVis* software. The *CosmoVis* client interface is accessible via modern web browsers, where the visualization is powered by *D3.js* for loading and displaying 2D graphs and *Three.js* for 3D volume rendering. The interface allows for interactive exploration of the simulation 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.

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 Sect. 7) 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 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.

### 5.1 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 Sect. 5.3.)

### 5.2 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 [48]. 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 Fig. 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 Fig. 1 and Fig. 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 unhindered, 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 Fig. 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. 5.4). This additional functionality helps to better localize and probe regions of interest in the data (see Fig. 10).

### 5.3 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. 1, 6, and 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.

### 5.4 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. Fig. 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 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

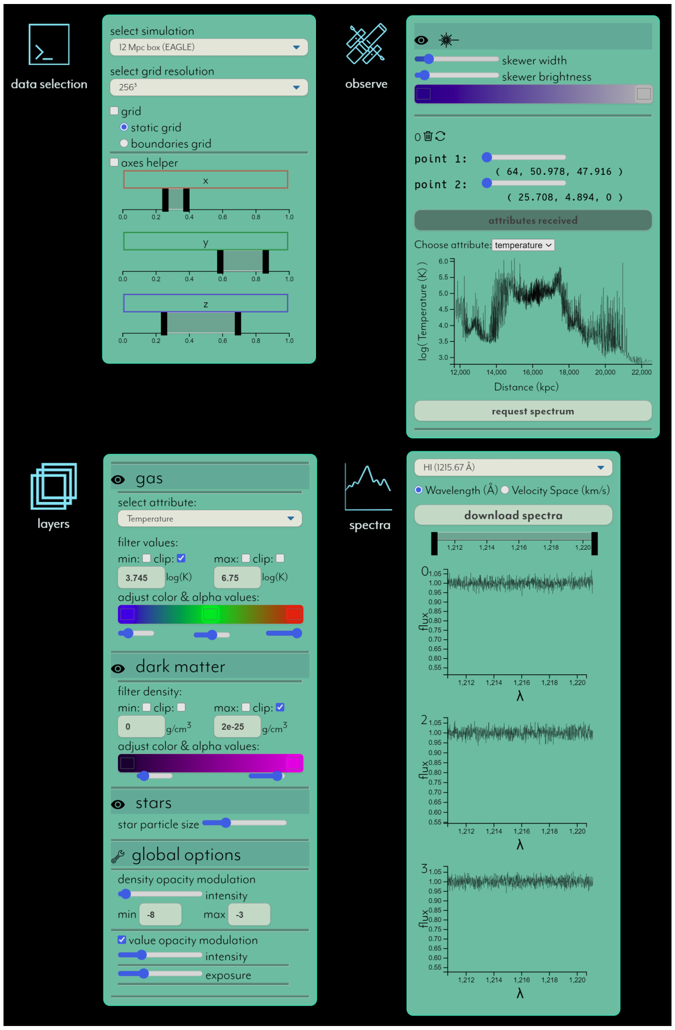


Fig. 8. Four user interface panels can be accessed on demand to load in and interact with simulation datasets, including retrieving spectra and CGM/IGM gas physical conditions: Data Selection, Layers, Observe, and Spectra.

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.

### 5.5 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 $\sim$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 Fig. 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 [3]. *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.

## 6 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).

### 6.1 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 [42]. 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. [45, 46] 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 [11, 16, 19]. 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 [21]. Furthermore, the initial visualization modality wherein the gas temperature is color coded reveals that the gas physical conditions can vary widely

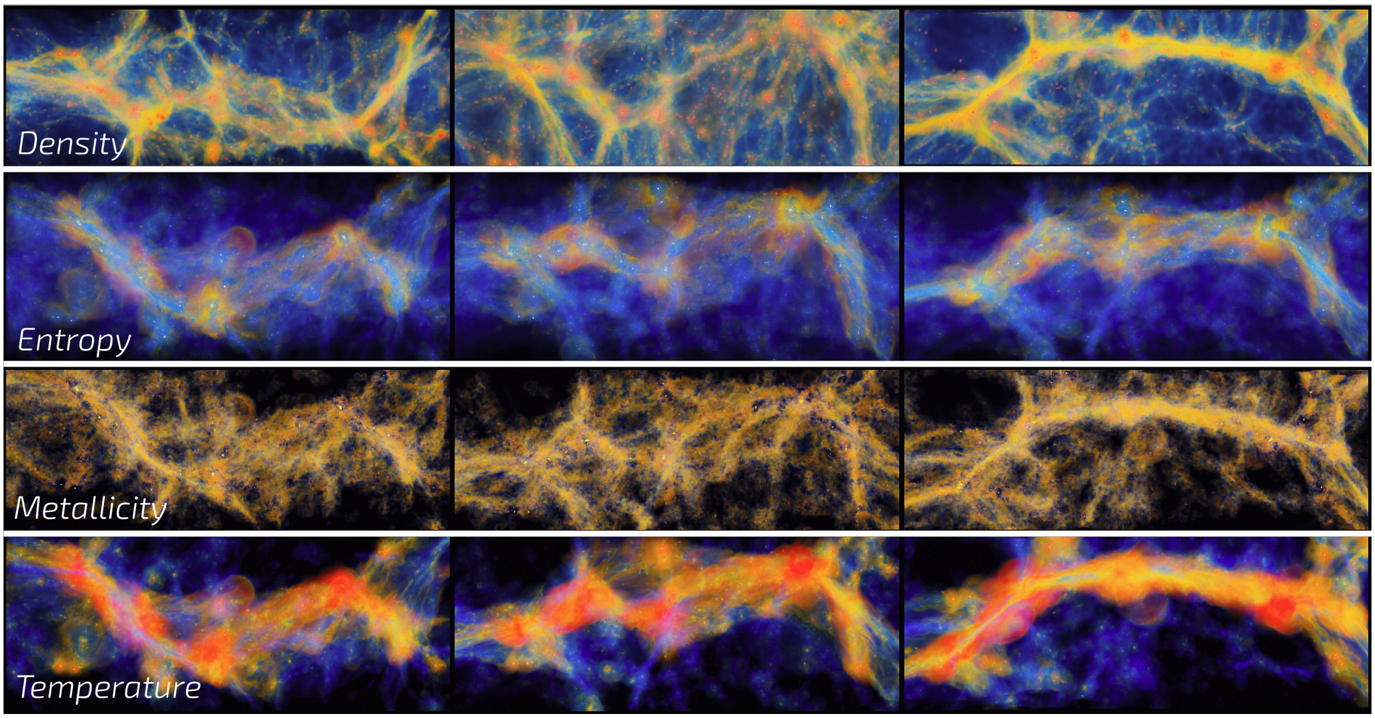


Fig. 9. Example of a cosmic sheet found in the 100 Mpc *EAGLE* simulation shown from three different angles rotated around the same axis, arranged left to right. Different gas fields are presented in each row, from top to bottom: density, entropy, metallicity, and temperature. See Use Case 1 in Sec. 6.1 for more information about these cosmic structures.

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. 9, and these regions have attracted heightened recent attention. As we discuss in **T1** (Sect. 4.1), 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 [66] 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 $\sim$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 (see Sec. 5.3). Using the 'request skewer attributes' functionality, we inspected the temperature, density, metallicity, and N(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(H I) ranged from log N(H I)/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/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.

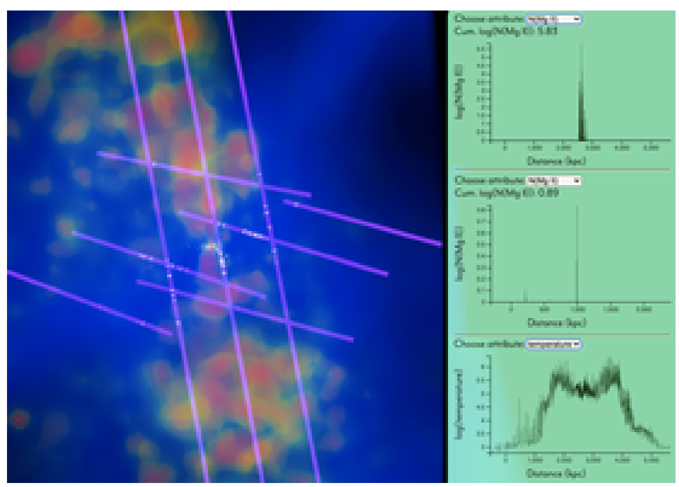

Fig. 10. The CGM environment analyzed in Use Case 2 (Sec. 6.2). Here we have placed ten skewers through an apparent biconical galactic superwind to analyze its thermodynamic structure.

### 6.2 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** (Sect. 4.2). 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 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 [53, 79]. Clearly, the CGM of this galaxy contains inhomogenities 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(H I)/cm$^{-1}$ = 11.8-13.3 and log N(O VI)/cm$^{-1}$ = 11.3-12.5. Typical values from the CGM literature are log N(H I)/cm$^{-1}$ $> 14.5$ and log N(O VI)/cm$^{-1}$ $> 14.0$ [39, 75, 76]. 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.

### 6.3 Case 3: Metal enrichment of the intergalactic medium

As discussed in **T3** (Sec. 4.3), 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 [10, 20], and nearly all of the metal line systems detected to date arise from circumgalactic environments [60]. 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.

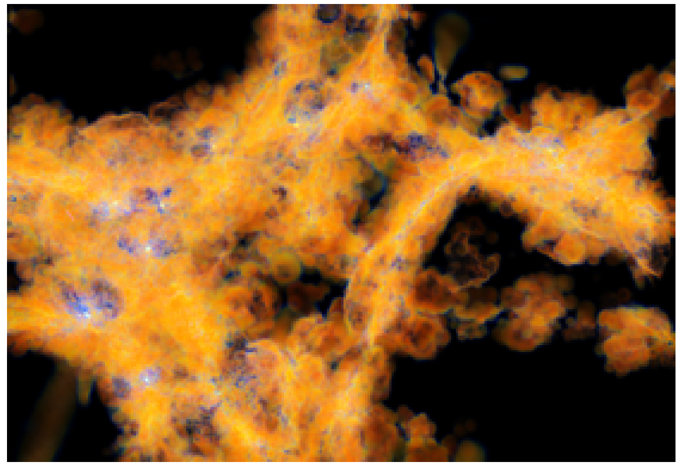

Fig. 11. The large-scale metallicity distribution in the *EAGLE* 25 Mpc volume. This is a $\sim 5$ Mpc slice of the cosmic web, revealing that much of the intergalactic medium in cosmic web filaments is highly enriched with heavy elements (low to high metallicity colored blue to red). Also seen in here are a number of regions in the immediate vicinities of galaxies likely evacuated of their metal-rich gas due to galactic winds [55].

Upon examining all simulation volumes, metal-enriched regions of intergalactic filaments are ubiquituous. 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 [55]. Fig. 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 the simulation, we should detect metal line absorbers in truly intergalactic space (i.e., not in the CGM).

## 7 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 [14]. 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 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* [59]. 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 *ComsoVis* 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.

## 8 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.

## References


[1] 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.

[2] 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.

[3] 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.

[4] U. Ayachit. *The ParaView Guide: A Parallel Visualization Application*. Kitware, Inc., 2015.

[5] J. N. Bahcall and L. Spitzer, Jr. Absorption lines produced by galactic halos. *The Astrophysical Journal Letters*, 156:L63, 1969.

[6] 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.

[7] M. J. Berger and J. Oliger. Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics*, 53(3):484–512, 1984.

[8] J. Bergeron and P. Boissé. A sample of galaxies giving rise to mg II quasar absorption systems. *A&A*, 243:344–366, Mar. 1991.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] E. F. Bunn and D. W. Hogg. The kinematic origin of the cosmological redshift. *American Journal of Physics*, 77(8):688–694, 2009.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] M. Donnari, A. Pillepich, D. Nelson, F. Marinacci, M. Vogelsberger, and L. Hernquist. Quenched fractions in the illustristng simulations: comparison with observations and other theoretical models. *Monthly Notices of the Royal Astronomical Society*, 506(4):4760–4780, 2021.

[24] A. G. Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6(4):320–330, 1970.

[25] 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.

[26] 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.

[27] A. Geller and A. Gurvich. Firefly: A WebGL tool to explore particle-based data, 2021. `https://agurvich.github.io/Firefly/`. Accessed March 30th, 2021.

[28] 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.

[29] J. Guedes, S. Callegari, P. Madau, and L. Mayer. Forming Realistic Late-type Spirals in a $\Lambda$CDM Universe: The Eris Simulation. *The Astrophysical Journal*, 742(2):76, 2011.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Science*, 15:168–173, 1929.

[35] 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.

[36] 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.

[37] V. Icke. Formation of galaxies inside clusters. *Astronomy and Astrophysics*, 27:1–21, 1973.

[38] 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.

[39] 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 $\Lambda$CDM simulations. *The Astrophysical Journal*, 870(2):137, Jan. 2019.

[40] 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.

[41] S. N. Leitner and A. V. Kravtsov. Fuel efficient galaxies: Sustaining star formation with stellar mass loss. *The Astrophysical Journal*, 734(1):48, 2011.

[42] 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.

[43] S. Lopez, N. Tejos, C. Ledoux, L. F. Barrientos, K. Sharon, J. R. Rigby, et al. A clumpy and anisotropic galaxy halo at redshift 1 from gravitational-arc tomography. *Nature*, 554:493–496, 2018.

[44] A. Man and S. Belli. Star formation quenching in massive galaxies. *Nature Astronomy*, 2(9):695–697, 2018.

[45] 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.

[46] 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.

[47] 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.

[48] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.

[49] 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.

[50] J. J. Monaghan. Smoothed particle hydrodynamics. *Annual Review of Astronomy and Astrophysics*, 30(1):543–574, 1992.

[51] J. Naiman. AstroBlend: An astrophysical visualization package for Blender. *Astronomy and Computing*, 15:50–60, 2016.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] P. J. E. Peebles. *Principles of Physical Cosmology*. Princeton University Press, 1993.

[57] 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.

[58] 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.

[59] 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.

[60] 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$\alpha$ and O VI absorption at $z < 0.2$. *The Astrophysical Journal*, 740:91, 2011.

[61] A. Rahmati, A. H. Pawlik, M. Raičevič, 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.

[62] 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.

[63] 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.

[64] 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.

[65] K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, and T. Ertl. Interactive visual exploration of a trillion particles. In *Proc. IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 56–64, 2016.

[66] 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.

[67] A. Scherzinger, T. Brix, D. Drees, A. Völker, K. Radkov, N. Santalidis, et al. Interactive exploration of cosmological dark-matter simulation data. *IEEE Computer Graphics and Applications*, 37(2):80–89, 2017.

[68] 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 relation. *Monthly Notices of the Royal Astronomical Society*, 456(2):1235–1258, 2016.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] J. Tumlinson, M. S. Peeples, and J. K. Werk. The circumgalactic medium. *Annual Review of Astronomy and Astrophysics*, 55:389–432, 2017.

[75] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, N. Katz, et al. The COS-Halos survey: Rationale, design, and a census of circumgalactic neutral hydrogen. *The Astrophysical Journal*, 777:59, 2013.

[76] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, D. H. Weinberg, et al. The large, oxygen-rich halos of star-forming galaxies are a major reservoir of galactic metals. *Science*, 334:948, 2011.

[77] 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.

[78] 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.

[79] S. Veilleux, G. Cecil, and J. Bland-Hawthorn. Galactic winds. *Annual Review of Astronomy and Astrophysics*, 43:769–826, 2005.

[80] 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.

[81] 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.

[82] 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.

[83] Y. B. Zeldovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and Astrophysics*, 5:84–89, 1970.

[84] 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.