Morphological Analysis of the IGM During the Cosmic Dawn and the Epoch of Reionization Using Large Cluster Statistics

Aman Kumar

IISER Kolkata

Summer Research Project under the supervision of Dr. Raghunath Ghara
IISER Kolkata

January 7, 2026

Abstract

We investigate the morphological evolution of the intergalactic medium (IGM) during the Cosmic Dawn (CD) and the Epoch of Reionization (EoR) using 21 cm brightness temperature (δT_b) fields generated with the `grizzly` radiative transfer code. By applying the `SURFGEN2` pipeline to simulated emission and absorption excursion sets, we extract three-dimensional structures and analyze their geometry and topology through Minkowski Functionals and Shapefinders.

We quantify structural evolution using cluster-wise shape parameters—Thickness, Breadth, Length, Planarity, Filamentarity, and Genus—and track the connectivity of regions through Largest Cluster Statistics (LCS). Our analysis reveals that the IGM evolves from isolated cold filaments to extended, filamentary heated regions, eventually forming complex, tunnel-rich networks. The percolation transitions, seen through sharp changes in LCS, highlight distinct phases in the connectivity of cold and hot regions.

This morphological approach offers insights beyond traditional power spectra, revealing hierarchical structure formation, the asymmetry between heating and ionization processes, and the emergence of universal shape trends across cosmic epochs.

1 Introduction

The 21 cm signal from neutral hydrogen offers one of the most powerful tools we have to study the early Universe—especially the Cosmic Dawn (CD) and the Epoch of Reionization (EoR). During this time, the first stars and galaxies formed and dramatically changed their surroundings by heating and ionizing the intergalactic medium (IGM). The fluctuations in the 21 cm brightness temperature δT_b carry detailed information about these changes.

In this project, we explore the evolving structures in simulated δT_b maps generated by the radiative transfer code `grizzly` , which includes X-ray and UV heating, reionization effects,

and redshift-space distortions. To analyze the complex morphology of these structures—both in absorption (cold regions) and emission (heated regions)—we use the `SURFGEN2` pipeline. This toolchain allows us to extract 3D surfaces, compute geometric descriptors (like Volume, Area, Curvature), and study how these structures grow and connect over time.

We focus on two main types of analysis. First, we look at the *Largest Cluster Statistics (LCS)* to see when disconnected regions begin to merge into large, network-like structures—a key signal of percolation. Second, we study the *shape evolution of individual clusters* using quantities like *Filamentarity*, *Planarity*, and *Genus*, which reveal whether structures are more like sheets, filaments, or blobs, and how "holey" or porous they are.

This approach helps us better understand how the cosmic web forms and how the morphology of structures reflects the underlying physical processes driving the evolution of the early Universe.

2 Procedure

2.1 Creating the 21 cm Brightness Temperature Map

The redshifted 21 cm signal from neutral hydrogen (H I) during the Cosmic Dawn (CD) and Epoch of Reionization (EoR) is a powerful probe of the thermal and ionization history of the early Universe. The key observable is the *differential brightness temperature* δT_b, defined as the contrast between the spin temperature of neutral hydrogen and the background radio temperature. For a region at angular position $\mathbf{x}$ and redshift z, it is given by:

$$\delta T_b(\mathbf{x}, z) = \frac{T_S(\mathbf{x}, z) - T_\gamma(z)}{1+z} \left[1 - \exp\left\{-\tau_b(\mathbf{x}, z)\right\}\right], \tag{1}$$

where:

- $T_S(\mathbf{x}, z)$: spin temperature of hydrogen atoms,
- $T_\gamma(z) = 2.73(1+z)\,\mathrm{K}$: radio background temperature (typically the CMB),
- $\tau_b(\mathbf{x}, z)$: optical depth of the 21 cm transition.

The optical depth is approximated as:

$$\tau_b(\mathbf{x}, z) = \frac{4\,\mathrm{mK}}{T_S(\mathbf{x}, z)}\, x_{\mathrm{HI}}(\mathbf{x}, z)\,[1 + \delta_B(\mathbf{x}, z)] \left(\frac{\Omega_B h^2}{0.02}\right) \left(\frac{0.7}{h}\right) \frac{H_0}{H(z)} (1+z)^3, \tag{2}$$

with:

- $x_{\mathrm{HI}}(\mathbf{x}, z)$: neutral hydrogen fraction,
- $\delta_B(\mathbf{x}, z)$: baryonic overdensity,
- Ω_B: baryon density parameter,
- $H(z)$: Hubble parameter at redshift z.

2.1.1 Simulation Pipeline and Map Generation

To construct δT_b maps, we use the `grizzly` [5] radiative transfer code. This code takes as input:

- N-body simulation outputs: dark matter halos, density and velocity fields,
- Source model: UV, X-ray, and Lyα emission from halos (emissivity tied to halo mass),
- Efficiency parameters such as star formation efficiency ($f_\star$), ionization efficiency (ζ), and X-ray heating efficiency (f_X).

The simulation proceeds as:

1. **Spin temperature field** T_S is computed using the Lyα and X-ray fluxes, assuming $\propto 1/R^2$ decay.
2. **Optical depth** τ_b is calculated using Eq. (2).
3. **Brightness temperature** δT_b is then evaluated throughout the 3D volume.
4. **Redshift-space distortions** are added using a cell-movement method to account for peculiar velocities.

2.1.2 Binary Emission Maps and Morphological Analysis

For morphological analysis, a binary field $\delta T_b^{\rm bin}(\mathbf{x}, z)$ is defined as [4]:

$$\delta T_b^{\rm bin}(\mathbf{x}, z) = \begin{cases} 1, & \delta T_b(\mathbf{x}, z) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

This binary map marks regions in **emission** and is input to the `surfgen2` algorithm, which:

- Implements periodic boundary conditions and friends-of-friends (FoF) region finding,
- Builds surface triangulations using Marching Cubes 33,
- Computes Minkowski functionals and Shapefinders for 3D morphology and topology analysis.

These δT_b maps are essential for extracting structural and statistical insights about the IGM during the early stages of cosmic evolution.

2.2 Method Overview

This framework provides a complete morphological characterization of a three-dimensional field via a multi-stage computational pipeline, implemented by code `SURFGEN2` [1]. The procedure begins by identifying all discrete, connected clusters within the data cube using a Friends-of-Friends (FoF) algorithm, which correctly handles periodic boundary conditions to properly identify structures that span the simulation volume .

Next, a topologically correct geometric representation of each cluster's surface is constructed using the Marching Cubes 33 (MC33) algorithm . This technique generates a continuous, triangulated isosurface for each cluster.

Finally, using this triangulated surface, we calculate key numbers that describe the cluster's shape and structure. This includes the four Minkowski Functionals - measuring volume, surface area, integrated mean curvature, and the Euler characteristic (from which the topological Genus is derived) and the dimensionless Shapefinders (Planarity and Filamentarity) that classify structure shape .Concurrently, Largest Cluster Statistics (LCS) are calculated to track global connectivity and identify the critical percolation transition.

2.3 The `SURFGEN2` Pipeline

The `SURFGEN2` code converts gridded δT_b data into geometric objects to precisely measure their shape, topology, and connectivity. This involves three key steps: defining regions of interest, identifying clusters, and constructing their surfaces.

2.3.1 Defining Regions of Interest

The first step in the analysis is to categorize the 21-cm signal data into physically distinct regions. We use an **"excursion set"** method, which applies a brightness temperature threshold, δT_{th}, to partition the data cube. This allows us to isolate and study the structures of the hot and cold parts of the intergalactic medium (IGM) separately.

- **Emission Regions**: These are hot spots where the gas has been heated by early X-ray sources, causing it to appear in emission against the radio background ($\delta T_b > 0$). We define this set as all locations where the brightness temperature meets or exceeds the threshold: $\{\mathbf{x}|\delta T_b(\mathbf{x}) \geq \delta T_{th}\}$.
- **Absorption Regions**: These are cold spots where neutral gas appears in absorption against the background ($\delta T_b < 0$). This set includes all locations where the brightness temperature is below the threshold: $\{\mathbf{x}|\delta T_b(\mathbf{x}) < \delta T_{th}\}$.

By varying the threshold δT_{th}, we can probe the morphology of structures at different temperature and density levels. A common baseline for analysis is $\delta T_{th} = 0$ mK, which provides a clean separation between the heated and cold components of the IGM.

2.3.2 Cluster Identification: Friends-of-Friends (FoF)

Once the simulation data is binarized, we use the Friends-of-Friends (FoF) algorithm to identify distinct, connected regions. It works by linking neighboring grid cells of the same type specifically, those that share a face into larger clusters. The algorithm also handles periodic boundary conditions, so if a structure wraps around the edges of the simulation box, it's still recognized as a single, continuous cluster.

2.3.3 Surface Reconstruction: Marching Cubes 33 (MC33)

A critical step in the analysis pipeline is the transformation of discrete, voxel-based clusters into continuous, triangulated surfaces suitable for geometric and topological analysis. This is achieved using the Marching Cubes 33 (MC33) algorithm, an improved and topologically consistent extension of the original Marching Cubes method [3]. Unlike the standard 15-case lookup, MC33 resolves face and interior ambiguities by incorporating a comprehensive set of 33 distinct configurations, ensuring that the generated surfaces preserve the true topology of the underlying field.

Within each cube, the scalar field is approximated by a trilinear interpolant:

$$g(x, y, z) = \sum_{i,j,k \in \{0,1\}} f(v_{ijk})\, x^i (1-x)^{1-i} y^j (1-y)^{1-j} z^k (1-z)^{1-k},$$

where $f(v_{ijk})$ are the scalar values at the cube's eight vertices. The isosurface is then defined as the set of points satisfying $g(x, y, z) = \delta T_{\rm th}$, which marks the boundary of the excursion set.

The MC33 algorithm proceeds in the following stages:

1. **Nodes Classification:** Each cube is scanned, and its eight nodes are marked as positive or negative depending on whether their field values lie above or below the threshold $\delta T_{\rm th}$.

2. **Topology Lookup:** Each cube in the grid has eight corner nodes, which are marked as positive or negative based on whether they lie above or below the threshold $\delta T_{\rm th}$. This creates $2^8 = 256$ possible sign combinations. However, many of these are equivalent under rotations, reflections, or sign inversion. After accounting for such symmetries and resolving geometric ambiguities, only 33 topologically distinct cases remain.

   Each of these 33 cases corresponds to a specific way the isosurface can pass through the cube and is stored in a lookup table. This table assigns a triangulation pattern consistent with the underlying trilinear scalar field. Configurations encountered during reconstruction are matched either directly or via symmetry to one of these canonical entries, ensuring that the surface mesh is both continuous and topologically accurate.

3. **Vertex Interpolation:** Triangle vertices are placed on the cube edges using linear interpolation between the scalar values at the edge's endpoints.

4. **Mesh Construction:** The resulting triangle patches from all cubes are combined to form a coherent and topologically faithful surface mesh.

The resulting mesh avoids holes, artifacts, and inconsistencies, and forms a reliable geometric object for further morphological analysis using Minkowski Functionals and Shapefinders.

2.4 Minkowski Functionals and Shapefinders

We have an accurate surface model of each cluster using the `MC33` algorithm and can begin to quantify its shape and structure. To do this, we use a set of geometric and topological measures known as Minkowski Functionals, along with derived quantities called Shapefinders, which provide deeper insight into the cluster's morphology [2].

2.4.1 Minkowski Functionals

For any closed two-dimensional surface in 3D space, its shape and topology can be completely described by four Minkowski Functionals [7]. Mathematically, these are defined by integrals over the surface, ∂K, involving its principal curvatures, κ_1 and κ_2:

1. **Volume (V):** $V = \int_K dV$
2. **Surface Area (S):** $S = \oint_{\partial K} dS$
3. **Integrated Mean Curvature (C):** $C = \frac{1}{2} \oint_{\partial K} (\kappa_1 + \kappa_2) dS$
4. **Euler Characteristic (χ):** $\chi = \frac{1}{2\pi} \oint_{\partial K} \kappa_1 \kappa_2 dS$

While these are the formal definitions, for a surface computationally constructed from a mesh of N_T triangles, these functionals are calculated using specific summation formulas:

- Volume (V): The volume enclosed by the surface. It is calculated by summing the signed volumes of the tetrahedra formed by each triangle and an arbitrary origin point:

$$V = \frac{1}{3} \sum_{i=1}^{N_T} \vec{r}_i \cdot \vec{A}_i$$

  where $\vec{r}_i$ is the position vector to a point on the i^{th} triangle and $\vec{A}_i$ is its area vector.

- Surface Area (S): The total surface area of the object, given by the sum of the areas of all constituent triangles:

$$S = \sum_{i=1}^{N_T} S_i$$

- Integrated Mean Curvature (C): A measure of the average curvature of the surface. For a triangulated mesh, it is calculated by summing over all shared edges:

$$C = \frac{1}{2} \sum_{i,j} \ell_{ij} \phi_{ij}$$

  where ℓ_{ij} is the length of the edge shared by adjacent triangles i and j, and ϕ_{ij} is the dihedral angle between their normals. This geometric construction is illustrated in Figure **??**, where a cylindrical surface is fitted along the shared edge to visualize the projected angle ϕ_{ij} and the contribution to the local curvature.

- Euler Characteristic(χ): A topological invariant that describes the surface's structure. For a triangulated surface, it is given exactly by Euler's polyhedral formula:

$$\chi = N_V - N_E + N_T$$

  where N_V, N_E, and N_T are the number of vertices, edges, and triangles, respectively.

To quantify objects's connectivity we derive genus(G) from Euler Characteristics. In simple terms, the genus is the number of "handles" an object has, minus the number of separate voids it encloses. This is related to the computationally derived Euler characteristic (χ) by:

$$G = (\text{Number of Handles}) - (\text{Number of voids}) = 1 - \frac{\chi}{2}$$

A simple sphere, having no handles or voids, has a genus of zero. A donut-shaped torus, with one handle and no voids, has a genus of one, while a pretzel has a genus of two.

2.4.2 Shapefinders

We define Shapefinders [7] as ratios of MFs to create physical descriptions of a cluster's shape. These were introduced to estimate the three physical extensions of an object in 3D and are defined as three characteristic length scales:

$$\text{Thickness: } T = \frac{3V}{S} \quad , \text{Breadth: } B = \frac{S}{C} \quad , \text{Length: } L = \frac{C}{4\pi}$$

These scales are spherically normalized such that $V = (4\pi/3)TBL$. A natural ordering of $L \geq B \geq T$ is expected; however, if this is not maintained in a calculation, the values are re-sorted by defining the smallest as T and the largest as L . In the rare cases where a cluster has a negative integrated mean curvature ($C < 0$), its absolute value is used to ensure B and L remain positive.

From these scales, we construct two dimensionless parameters, Planarity (P) and Filamentarity (F), which are used to classify the fundamental shape of the clusters :

$$\text{Planarity: } P = \frac{B - T}{B + T} \quad , \text{Filamentarity: } F = \frac{L - B}{L + B}$$

These two parameters map any object's shape onto a 2D plane. A perfect sphere has $P \approx 0$ and $F \approx 0$, a flattened "pancake-like" structure has high planarity ($P \gg F$), and an elongated "thread-like" structure has high filamentarity ($F \gg P$).

It is worth noting that alternative, "macroscopic" definitions for Length (L_1) and Filamentarity (F_1) can be defined by incorporating the genus, G :

$$L_1 = \frac{C}{4\pi(1 + |G|)} \quad \text{and} \quad F_1 = \frac{L_1 - B}{L_1 + B}$$

While these assess the overall macroscopic shape, this work focuses on the local or "microscopic" information provided by the standard definitions of L and F .

2.5 Largest Cluster Statistics

To analyze the global connectivity of clusters, we study percolation, the process by which isolated regions merge into a single, connected network that spans the entire volume. This transition marks a fundamental change in the topology of the field, from a collection of isolated "clumps" to a "sponge-like" structure. We quantitatively track this transition using a geometric approach centered on the Largest Cluster Statistics (LCS).

The analysis begins by partitioning the simulation volume into HI overdense (emission) and HI underdense (absorption) regions using a density threshold. We then compute a set of key statistics for each segment separately:

- **Filling Factor (FF):** The fraction of the total simulation box volume that a given segment (e.g., all emission regions combined) occupies.

$$FF = \frac{\text{Total volume of all clusters}}{\text{Volume of the simulation box}}$$

- **Largest Cluster Statistic (LCS):** The fraction of a segment's total volume that is contained within its single largest cluster. A sharp, sudden increase in the LCS value provides an unambiguous signal that percolation has occurred.

$$LCS = \frac{\text{Volume of the largest cluster}}{\text{Total volume of all clusters}}$$

- **Volume Fraction of the Largest Cluster (ff_{largest}):** The volume of the largest cluster as a fraction of the total simulation box volume. It is related to the other statistics by $ff_{\text{largest}} = FF \times LCS$.

- **Volume Fraction of Other Clusters (ff_{other}):** The combined volume of all clusters except for the largest one, as a fraction of the total simulation box volume. It is related to the other statistics by $ff_{\text{other}} = FF - ff_{\text{largest}}$.

3 Results and Discussion

In this section, we explore how the structure of the intergalactic medium (IGM) changes over time during the Cosmic Dawn (CD) and the Epoch of Reionization (EoR). We use the δT_b fields from our simulations and apply the SURFGEN2 pipeline to identify and analyze the shapes and connectivity of both heated (emission, $\delta T_b > 0$) and cold (absorption, $\delta T_b < 0$) regions. For this, we use geometric tools like **Thickness (T)**, **Breadth (B)**, **Length (L)**, and shape descriptors like **Planarity (P)**, **Filamentarity (F)**, and **Genus (G)**, tracking how these quantities evolve with the overall filling factor and individual cluster volume.

3.1 Analysis of Morphological Statistics

In this section, we analyze the statistical properties of the simulated 21-cm signal maps. We first examine the global connectivity of heated regions using the Largest Cluster Statistics (LCS). We then investigate the specific shapes of individual clusters using Shapefinders.

3.1.1 Largest Cluster Statistics and Percolation

To understand the global connectivity of the cosmic web, we use the **Largest Cluster Statistics (LCS)** to compare the different stories of the overdense (emission) and underdense (absorption) regions, as shown in Figure 1.

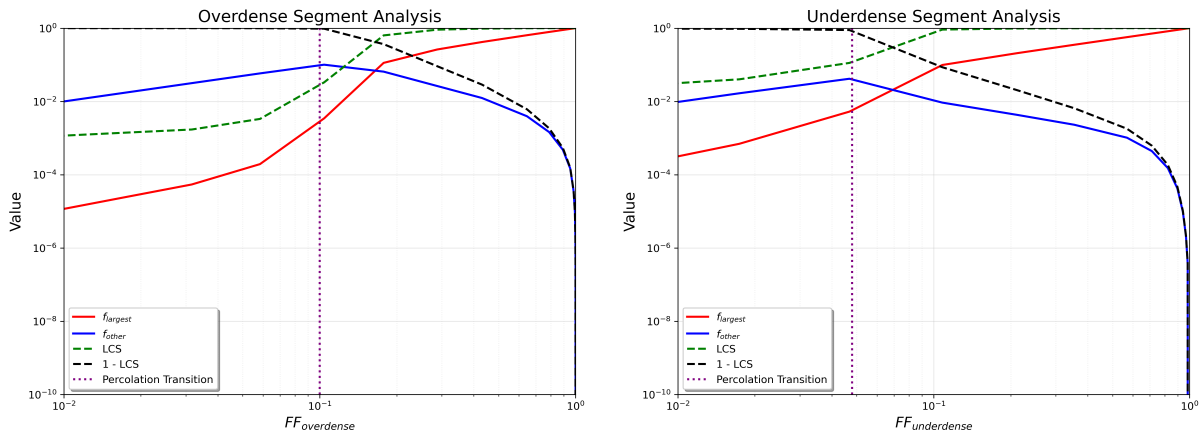

Figure 1: Largest Cluster Statistics for the overdense (emission) segment (left) and the underdense segment (right). The percolation transition for the overdense segment is marked by the sharp rise in its LCS. The transition for the underdense segment is marked by the sharp drop in its LCS as it breaks apart.

The two panels reveal the fundamentally different ways these two cosmic components evolve, with the transition point for each being clearly marked by the behavior of these four quantities.

The **left panel** tells a story of building up. The overdense segment starts as isolated "islands" of heated gas. The percolation transition [6] at $FF_{\text{overdense}} \approx 0.1$ is signaled by several sharp changes: the **LCS** and $f_{\mathbf{largest}}$ rise steeply, while $1 - \mathbf{LCS}$ plummets. This indicates that a single cluster has suddenly become dominant. The sharp drop in $f_{\mathbf{other}}$ is particularly significant, as it shows the exact moment when this single largest cluster begins to absorb the smaller, independent islands into one connected network.

The **right panel** tells the opposite story: one of breaking down. The underdense segment begins as one connected "sea" of cold gas, with its LCS near 1. The transition where this sea breaks apart, at $FF_{\text{underdense}} \approx 0.048$, is marked by the reverse behavior: the **LCS** and $f_{\mathbf{largest}}$ drop precipitously, while the $1 - \mathbf{LCS}$ rises sharply from near zero. This indicates the moment the single, connected sea shatters into multiple, disconnected "lakes," and the volume is no longer dominated by a single entity.

This stark difference in behavior is significant. In a perfectly random, symmetric (Gaussian) universe, the building-up and breaking-down processes would be mirror images of each other. The fact that they occur at different filling factors is a clear fingerprint of the **non-Gaussianity** of the cosmic web, a feature that arises from the complex influence of gravity on structure formation.

3.1.2 Cluster-wise Shape Analysis

To probe the geometry of cosmic structures, we analyze how their fundamental dimensions and resulting shapes evolve with cluster volume at a specific epoch, z=10.11, as shown in Figure 2.

The **left panel** reveals that as clusters grow, their fundamental dimensions evolve in a strikingly non-uniform way. The **Thickness** (T) and **Breadth** (B) remain nearly constant across all volumes(in units of 0.56 x Mpc). In sharp contrast, the **Length** (L) increases dramatically from ∼10 to over 1000 across the same volume range. This demonstrates that clusters grow

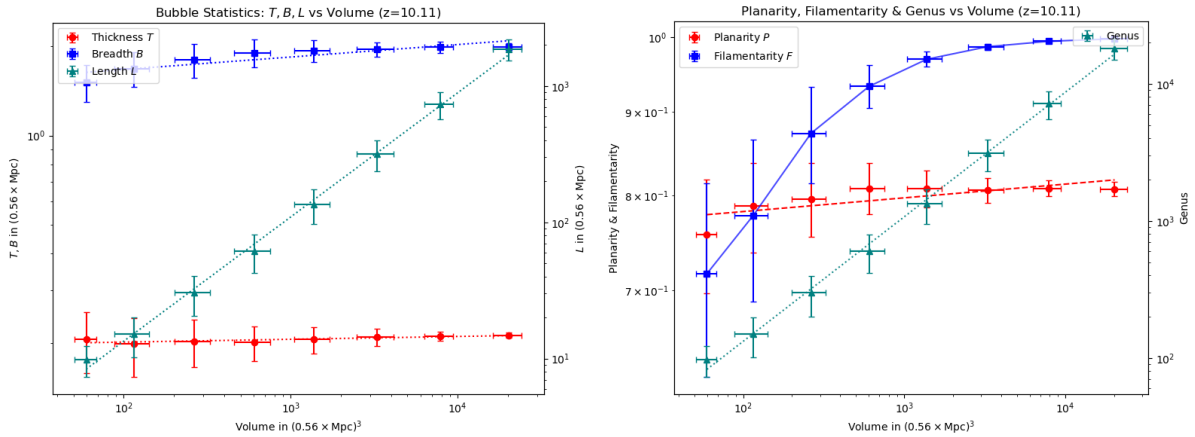

Figure 2: Shape analysis of heated clusters at z=10.11. **Left:** The three Shapefinders—T,B,L vs. volume. **Right:** The derived parameters P,G,F vs. volume.

primarily by elongating, a defining characteristic of filamentary structures.

This geometric behavior directly shapes the clusters' morphology, as quantified in the **right panel**. Because Length grows much faster than Breadth, the **Filamentarity** (F) rises sharply from ∼0.7 for the smallest clusters to nearly 1.0 for clusters with volumes greater than 10^3 units, confirming that larger structures are overwhelmingly filamentary. Conversely, since Thickness and Breadth are similar, **Planarity** (P) remains low and constant around ∼0.8, ruling out flat, sheet-like structures. Simultaneously, the **Genus** (G) increases from ∼100 to over 10,000, indicating that these large filaments are not simple strands but have a complex, porous topology with many internal tunnels.

The error bars represent the volume-weighted standard deviation of the shape measurements (y-axis) within each volume bin, quantifying the diversity of shapes.

For small-volume clusters (e.g., $V \sim 10^2$ units), the large error bars visible in both panels, and especially for Filamentarity (F) in the right panel, indicate a high degree of scatter. This signifies a wide diversity of shapes among the smaller "building-block" clusters.

As cluster volume increases ($V > 10^3$ units), the error bars shrink dramatically. This convergence signifies that as smaller, diverse clusters merge and grow, they adopt a much more uniform and highly filamentary morphology. The large clusters that will eventually form the single, percolating network therefore share a common, predictable shape, even before the network is fully connected across the entire cosmic volume.

3.2 Morphological Evolution during Cosmic Dawn

Figure 3 shows how the largest emission and absorption clusters evolve as more of the IGM becomes heated (measured by the emission filling factor $FF_{\rm emi}$).

At early stages ($FF_{\rm emi} \lesssim 0.1$), the cold absorption regions are much more prominent. They are larger and more connected, with higher values of Thickness, Breadth, and especially Length. This tells us that, initially, the IGM is dominated by long, coherent cold filaments, while hot spots are small and scattered.

As heating increases ($FF_{\rm emi} \sim 0.1$–0.3), the emission cluster grows quickly in Length. This

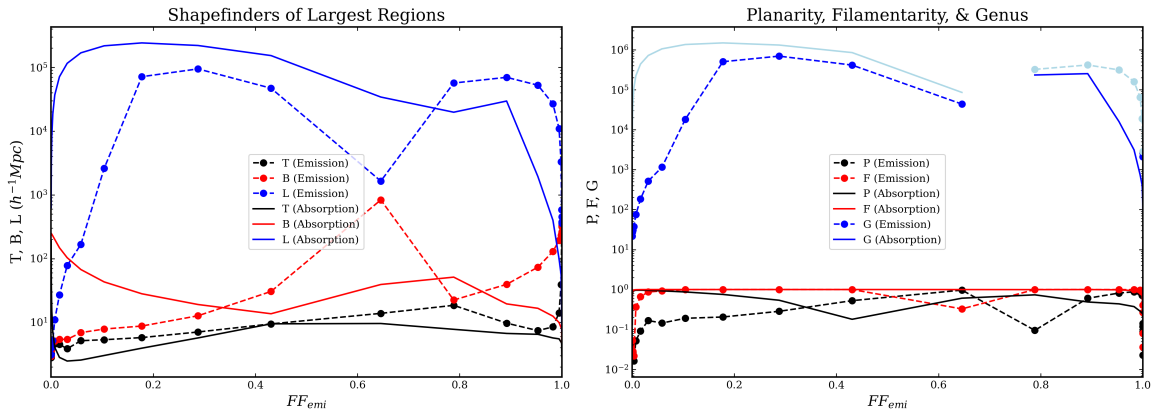

Figure 3: Evolution of Shapefinders (T, B, L), Planarity (P), Filamentarity (F), and Genus (G) for the largest emission (dashed) and absorption (solid) clusters with emission filling factor $FF_{\rm emi}$.

signals a shift from isolated hot regions to extended, filament-like heated structures as more X-ray sources contribute to heating. Meanwhile, the cold regions also grow, but more steadily, retaining their large-scale structure.

Planarity remains low throughout, suggesting that neither hot nor cold regions form flat, sheet-like structures. Filamentarity rises with $FF_{\rm emi}$ and peaks around 0.25, especially for absorption clusters — consistent with cold regions forming strong filamentary backbones in the early IGM. Genus increases notably for the emission cluster, indicating that the hot regions are developing more tunnels and internal connectivity — a sign of network-like structure forming among the heated zones.

In summary, cold filaments dominate the early morphology of the IGM, while heating causes hot regions to grow and connect, eventually forming extended, multiply connected structures.

3.3 Morphological Evolution during Reionization

Figure 4 shows a similar analysis during the EoR, where UV photons from galaxies ionize the neutral hydrogen.

The top-left panel shows the Largest Cluster Statistic (LCS), which jumps sharply around $FF_{\rm HII} \sim 0.15$. This is the percolation point — the moment when small, isolated ionized bubbles merge into one large, connected ionized structure.

Minkowski Functionals (top right) show that as reionization progresses, the volume and surface area of ionized regions increase steadily. The mean curvature peaks near percolation, reflecting complex bubble surfaces, and then drops, suggesting the surfaces become smoother as bubbles merge.

The Shapefinders (bottom left) show that ionized regions grow mainly by elongating — Length grows faster than Breadth and Thickness. This reflects the directional spread of ionization fronts, likely along lower-density paths in the IGM.

Filamentarity (bottom right) peaks just before percolation, then decreases. This means that ionized structures are most filamentary when they are still growing and merging. Once they

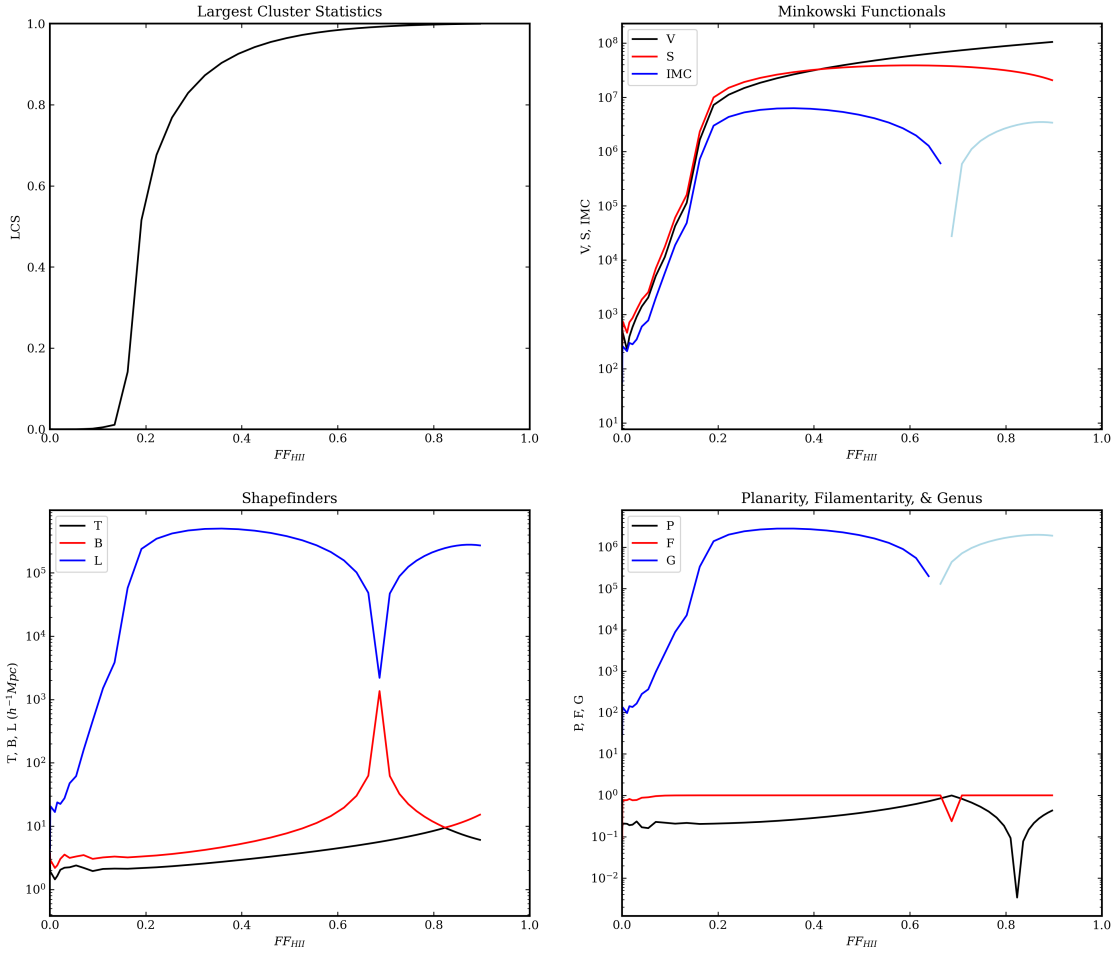

Figure 4: Morphological evolution of ionized (HII) regions during EoR. Top left: LCS. Top right: Minkowski Functionals. Bottom left: Shapefinders. Bottom right: P, F, and G. All vs. $FF_{\rm HII}$.

combine into large zones, their shape becomes more blob-like. Genus (G) also increases and then drops, showing a peak in internal complexity followed by a simpler, connected structure.

In essence, EoR transforms the IGM from a neutral, patchy medium into a large, connected ionized region, with filamentary and complex topology emerging midway through the process.

3.4 Cluster-wise Morphology and Variability

Figure 5 gives a closer look at individual clusters — how their geometry and topology change with volume at four different stages of the heating process. The error bars show the variation across many clusters at each volume.

As expected [2] , all Shapefinders (T, B, L) increase with cluster volume. Length grows fastest, especially for larger clusters, showing that the biggest structures tend to be long and stretched out. This effect is more prominent in absorption clusters, especially at early stages ($FF \approx 0.01$), reflecting the dominance of cold filaments early on.

The variation (spread) is large for small volumes and low FF, meaning early structures are

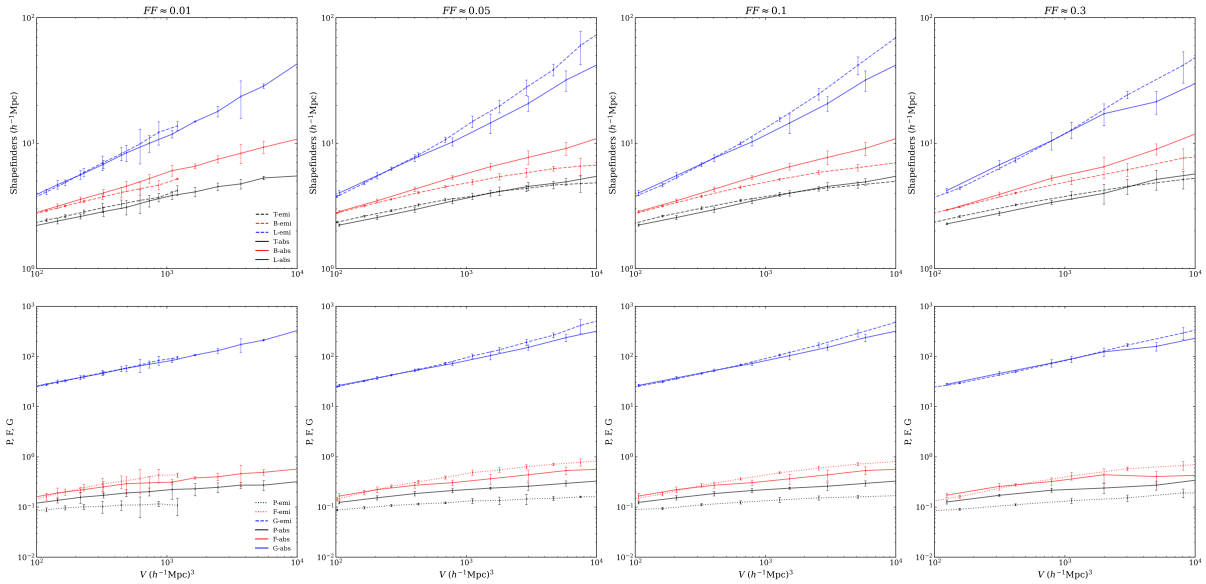

Figure 5: Cluster-wise Shapefinders and descriptors vs. volume V at fixed filling factors ($FF \approx 0.01, 0.05, 0.1, 0.3$) for emission (dashed) and absorption (solid). Error bars denote 1σ spread.

more diverse in shape. But as heating progresses ($FF \gtrsim 0.1$), the scatter narrows, and the shapes become more consistent, showing the emergence of a more uniform large-scale structure.

The shape descriptors (bottom row) show that Filamentarity increases with cluster volume and remains higher for absorption regions across all stages, reinforcing that cold clusters are more elongated. Planarity remains low, again showing that clusters are not sheet-like. Genus increases with volume too, meaning that larger clusters tend to have more tunnels and internal voids, i.e., more complex topology.

Altogether, this plot shows that structure formation is hierarchical: small clusters are compact and simple, while larger ones are stretched, connected, and topologically rich. Over time, as the IGM heats and reionizes, emission clusters become more similar to absorption clusters in structure, forming the connected web of the early universe.

4 Conclusion and Key Findings

Our analysis of the 21 cm signal reveals how the structures of the early Universe grow, connect, and change shape during the Cosmic Dawn and Reionization. Here's what we found:

- **Structures grow and merge in a very organized way.** As the Universe evolves, small regions of absorption or emission gradually connect into larger, coherent clusters. This process has a clear tipping point—called the *percolation threshold*—where the largest structure suddenly spans a large volume. We see this happening both in cold (absorption) and hot (emission) regions.

- **Filamentary structures dominate the early Universe.** Early on, especially during the Cosmic Dawn, we see long and thin structures forming. Our shape analysis shows that clusters grow mainly in length, while their width and thickness remain mostly the

same. This is a clear signature of *filament formation*, with little sheet-like or blob-like growth.

- **Cold filaments appear first and are highly connected.** In the early stages, cold regions form thin, thread-like filaments with high Filamentarity and Genus values—meaning they're not only long but also full of tunnels and loops. These early cold filaments form the skeleton of the cosmic web.

- **Heated and ionized regions later fill in the structure.** As heating progresses, hot emission regions grow around the cold filaments and start forming larger connected networks. The topology becomes more complex—filled with holes and tunnels—as captured by an increasing *Genus*. Later during EoR, ionized bubbles behave similarly, connecting and growing into large, complex networks before eventually smoothing out.

- **Shape statistics show a universal trend.** As structures get bigger and merge, their shapes start to look similar across different redshifts and types (cold, hot, ionized). This suggests some level of universality in how these structures evolve over time, driven by the underlying physics of heating and ionization.

In short, the 21 cm signal doesn't just tell us where things are in the early Universe—it tells us *how* structures grow, connect, and change shape. The combination of geometric and topological tools we used offers a new way to understand this evolution, going beyond traditional power spectrum analysis, and painting a detailed picture of how the cosmic web came to be.

Acknowledgements

I would like to express my sincere gratitude to Dr.Raghunath Ghara for their invaluable mentorship and for providing access to the simulation data used in this study. Their guidance on both the physical interpretation and the technical aspects of the project was instrumental throughout the research.

I also thank the developers of the `grizzly` radiative transfer code and the `SURFGEN2` pipeline, which formed the computational backbone of this work.

Finally, I acknowledge the authors of the referenced literature for their foundational contributions to the field.

References

[1] Satadru Bag, Rajesh Mondal, and Somnath Bharadwaj. Quantifying the morphology of reionization using minkowski functionals. *Monthly Notices of the Royal Astronomical Society*, 477(2):1984–1996, 2018.

[2] Satadru Bag, Rajesh Mondal, and Somnath Bharadwaj. Morphology of ionized regions during the epoch of reionization using large-scale structure statistics. *Monthly Notices of the Royal Astronomical Society*, 525(1):1042–1057, 2023.

[3] E. V. Chernyaev. Marching cubes 33: Construction of topologically correct isosurfaces. `https://cds.cern.ch/record/219282`, 1995. CERN Report CN/95-17.

[4] Raghunath Ghara, Satadru Bag, Saleem Zaroubi, and Suman Majumdar. The morphology of the redshifted 21-cm signal from the cosmic dawn. *Monthly Notices of the Royal Astronomical Society*, 2024. Preprint, arXiv:2308.00548.

[5] Raghunath Ghara, T. Roy Choudhury, and Kanan K. Datta. 21-cm signal from the cosmic dawn – i. insights into the first sources. *Monthly Notices of the Royal Astronomical Society*, 447(2):1806–1819, 2015.

[6] Anatoly Klypin and Sergei F. Shandarin. Three-dimensional topology of density perturbations and the morphology of the large-scale structure. *The Astrophysical Journal*, 413:48–62, 1993.

[7] Varun Sahni, B. S. Sathyaprakash, and Sergei F. Shandarin. Shapefinders: A new shape diagnostic for large-scale structure. *The Astrophysical Journal Letters*, 495(1):L5–L8, 1998.