Cosmic Reionization On Computers: Statistics, Physical Properties and Environment of Lyman Limit Systems at $z \sim 6$

JIAWEN FAN (樊稼问), HANJUE ZHU (朱涵珏), CAMILLE AVESTRUZ, and NICKOLAY Y. GNEDIN

1 Department of Physics; University of Michigan, Ann Arbor, MI 48109, USA
2 Department of Astronomy & Astrophysics; The University of Chicago; Chicago, IL 60637, USA
3 Leinweber Center for Theoretical Physics; Department of Physics; University of Michigan, Ann Arbor, MI 48109, USA
4 Theoretical Physics Division; Fermi National Accelerator Laboratory; Batavia, IL 60510, USA
5 Kavli Institute for Cosmological Physics; The University of Chicago; Chicago, IL 60637, USA

Submitted to ApJ

ABSTRACT

Lyman Limit Systems (LLSs) are dense hydrogen clouds with high enough HI column densities ($N_{HI}$) to absorb Lyman continuum photons emitted from distant quasars. Their high column densities imply an origin in dense environments; however, the statistics and distribution of LLSs at high redshifts still remain uncertain. In this paper, we use self-consistent radiative transfer cosmological simulations from the “Cosmic Reionization On Computers” (CROC) project to study the physical properties of LLSs at the tail end of cosmic reionization at $z \sim 6$. We generate 3000 synthetic quasar sightlines to obtain a large number of LLS samples in the simulations. In addition, with the high physical fidelity and resolution of CROC, we are able to quantify the association between these LLS samples and nearby galaxies. Our results show that a higher fraction of higher column density LLSs are spatially associated with nearby galaxies. Moreover, we find that LLSs that are not near any galaxy typically reside in filamentary structures connecting neighboring galaxies in the intergalactic medium (IGM). This quantification of the distribution and associations of LLSs to large scale structures informs our understanding of the IGM-galaxy connection during the Epoch of Reionization, and provides a theoretical basis for interpreting future observations.

Keywords: galaxies — methods, numerical — cosmology

1. INTRODUCTION

The recent launch of JWST opens up a new window in studying galaxy formation during the Epoch of Reionization. Recent and forthcoming data from JWST’s near-infrared and mid-infrared bands enable observers to study the earliest galaxies and to probe galaxy formation to fainter magnitudes. As the observations push beyond $z \sim 6$ (Crighton et al. 2019; Adams et al. 2022), however, they lose an important window into the galaxy formation - the Intergalactic Medium (IGM) as probed by the Ly-α forest absorption in the spectra of distant quasars. The class of Ly-α absorbers that is the hardest to study at the highest redshifts is Lyman limit systems (LLSs). The nature of these systems is still not fully understood, and the observational challenges in observing them at the highest redshifts places an undue burden on theoretical modeling.

As the first stars and galaxies form, photons emitted from these luminous sources travel through the IGM. Emitted photons encounter hydrogen atoms residing in the diffuse IGM and in LLSs. These photons rapidly ionize the neutral hydrogen in the diffuse IGM, but LLSs, which presumably trace high-density regions, shield a significant fraction of enclosed neutral hydrogen. As the ionized regions grow and merge during the process of reionization, the enclosed LLSs alter the propagation of ionizing photons inside ionized regions, thereby changing the Mean Free Path (MFP) of the Lyman continuum (LyC) photons. LLSs therefore affect reionization processes and better understanding of LLSs provide further insights in the physics of reionization.
LLSs are usually defined by the value of the hydrogen column density along the line-of-sight. By a formal definition LLSs have column densities greater than \( N_{\text{HI}} > 1.6 \times 10^{17} \text{cm}^{-2} \) (this corresponds to an optical depth of 1) and less than \( N_{\text{HI}} < 2 \times 10^{20} \text{cm}^{-2} \) (with higher column density systems called damped Ly\(\alpha\) systems), although of course this definition is rather arbitrary and does not correspond to any physical characterization of absorbers. LLSs contribute more to the Ly\(\alpha\) forest (systems with \( N_{\text{HI}} \lesssim 1.6 \times 10^{17} \text{cm}^{-2} \)) and they are much more frequently seen than the denser but more rare damped Ly\(\alpha\) systems. While many low-z observations of the LLSs exist in the literature (Prochaska et al. 2010; Ribaud et al. 2011; O’Meara et al. 2013), far fewer observations exist at high-z. Observational measurements of the statistics and properties of LLSs at high-z require more quasar spectra, particularly with complete spectral characterization of absorbers. LLSs contribute more to the Ly\(\alpha\) forest (systems with \( N_{\text{HI}} \lesssim 1.6 \times 10^{17} \text{cm}^{-2} \)) and they are much more frequently seen than the denser but more rare damped Ly\(\alpha\) systems. While many low-z observations of the LLSs exist in the literature (Prochaska et al. 2010; Ribaud et al. 2011; O’Meara et al. 2013), far fewer observations exist at high-z.

We use \( X_{\text{HI}} = 0.001 \) as our fiducial threshold for highly neutral regions. This choice is somewhat arbitrary; we tested other \( X_{\text{HI}} \) values \( (X_{\text{HI}} = \{0.0001, 0.001, 0.003\}) \), which all lead to similar distributions of column densities in the LLS range, though such a method for determining the HI column density would not work reliably in the Ly\(\alpha\) forest. We illustrate this comparison in Figure 2 and further describe the distribution of column densities in comparisons across different CROC boxes and also with observations in Section 3.1. The primary take-away from Figure 2 is that the number of LLSs with \( N_{\text{HI}} > 3 \times 10^{17} \text{cm}^{-2} \) does not significantly vary with threshold choice.

2. METHODOLOGY

2.1. “Cosmic Reionization on Computers”: CROC Simulation

CROC simulations use the Adaptive Refinement Tree code (Kravtsov et al. 1997, 2002; Rudd et al. 2008). To account for physics necessary to self-consistently model cosmic reionization, CROC simulations include gravity, gas dynamics, fully coupled (to gas dynamics) radiative transfer, atomic cooling and heating processes, molecular hydrogen formation, star formation, and stellar feedback. Full details of the simulation are described in the CROC method paper (Gnedin 2014).

For our analysis, we use three simulation boxes with box length \( L_{\text{box}} = 40 h^{-1} \text{Mpc} \). These simulation boxes have a maximum spatial resolution of 100 pc in proper units. We extract 1000 line-of-sight samples in each of the boxes that correspond to independent random realizations. The line-of-sights are sampled non-uniformly, fully tracing the underlying high resolution of the adaptively refined grid. Each line-of-sight sample is 100 \( h^{-1} \text{Mpc} \) in length, which allows us to find even rare absorbers.

2.2. Identifying Lyman Limit Systems

The observational signature of an LLS is a break at the Lyman limit (a drop in transmitted flux at 912 Å rest frame). In numerical simulations, we can simply look for regions with high enough neutral fraction \( (X_{\text{HI}}) \). Figure 1 shows the neutral fraction along one line-of-sight in our simulation. Along each line of sight we search for regions with \( X_{\text{HI}} \) values larger than a given threshold, e.g. the region indicated by the red rectangle in the lower panel of Figure 1. We define the edges of these highly neutral regions to be at the intersection points between the threshold and the line-of-sight, e.g. the left and right sides of the red rectangle, which also corresponds to the “piece” of the LLS identified.

We use \( X_{\text{HI}} = 0.001 \) as our fiducial threshold for highly neutral regions. This choice is somewhat arbitrary; we tested other \( X_{\text{HI}} \) values \( (X_{\text{HI}} = \{0.0001, 0.001, 0.003\}) \), which all lead to similar distributions of column densities in the LLS range, though such a method for determining the HI column density would not work reliably in the Ly\(\alpha\) forest. We illustrate this comparison in Figure 2 and further describe the distribution of column densities in comparisons across different CROC boxes and also with observations in Section 3.1. The primary take-away from Figure 2 is that the number of LLSs with \( N_{\text{HI}} > 3 \times 10^{17} \text{cm}^{-2} \) does not significantly vary with threshold choice.

2.3. Galaxy Associations and Visualization of Lyman Limit Systems

A primary goal of this paper is to have a clearer understanding of where LLSs are located with respect to surrounding galaxies. We use the Rockstar halo finder (Behroozi et al. 2013) to identify dark matter halos in our simulations. We use the identified center of the particles’ mass distribution inside the dark matter halos as a proxy for the location of each hosted galaxy, enabling spatial associations with LLSs.

We define the “association” between LLSs and nearby galaxies using both the spatial locations and virial radii \( (R_{\text{vir}}) \) of the dark matter halos hosting these galaxies. We quantify the strength of associations using the quantity \( f_n(R_{\text{vir}}) \), where \( n \) is an integer. With \( n = 1 \),
3. RESULTS

3.1. Statistics of LLSs

An important quantity that observers have studied extensively at low-z is the column density distribution of LLSs. Figure 3 shows the column density distribution of LLSs. We compute the y-axis quantity by taking the LLSs number count in each column density bin ($N_{\text{HI}}$) normalized by the total path length sampled in the simulation ($\Delta l$) and the bin size ($\Delta N_{\text{HI}}$). We then multiply this quantity by the column density value in each bin to reduce the dynamic range for comparison, which is equivalent to measuring the column density distribution per log of neutral hydrogen column density.

$$dX = \frac{H_0}{H(z)}(1 + z)^2dz,$$

$$X_{\text{HI}}$$ is the fraction of LLSs whose neutral fraction peaks are inside the $R_{\text{vir}}$ of any galaxy. With $n = 2$, $f_2(R_{\text{vir}})$ is the fraction of LLSs with neutral fraction peaks within $2 \times R_{\text{vir}}$ of some galaxy. Higher $n$ values indicate associations at larger distances with respect to $R_{\text{vir}}$.

We also generate gas density projection plots around every LLS in the simulation box snapshots using the “yt” simulation analysis package (Turk et al. 2011). In these projection plots, one can visually identify structures surrounding the LLS including galaxies and connective filaments.

$X_{\text{HI}}$ is the fraction of LLSs whose neutral fraction peaks are inside the $R_{\text{vir}}$ of any galaxy. With $n = 2$, $f_2(R_{\text{vir}})$ is the fraction of LLSs with neutral fraction peaks within $2 \times R_{\text{vir}}$ of some galaxy. Higher $n$ values indicate associations at larger distances with respect to $R_{\text{vir}}$.

An important quantity that observers have studied extensively at low-z is the column density distribution of LLSs. Figure 3 shows the column density distribution of LLSs. We compute the y-axis quantity by taking the LLSs number count in each column density bin ($\Delta N_{\text{HI}}$) normalized by the total path length sampled in the simulation ($\Delta l$) and the bin size ($\Delta N_{\text{HI}}$). We then multiply this quantity by the column density value in each bin to reduce the dynamic range for comparison, which is equivalent to measuring the column density distribution per log of neutral hydrogen column density.

The left panel shows the individual distributions from each of the three different realizations of the CROC simulations to account for the cosmic variance. Here, we see that the column density distributions of LLSs in CROC exhibit an overall trend where the LLS number density decreases with increasing column density. Overall, the three simulation boxes produce similar column density distributions.

The right hand panel shows the same column density distribution of LLSs from the combined three CROC boxes, alongside fits to observational data from Crighton et al. (2019) at $z = 2.4$ and $z = 4.4$. We see that our simulation result and the observational fits have a relatively similar shape, with CROC having more LLSs with $16 < \log_{10}(N_{\text{HI}}) < 19$. These differences are acceptable, as simulations are not expected to be at the level of fidelity to match all the data perfectly. It is sufficient for our purpose that they capture the dominant trends in the data, such as a depression at $N_{\text{HI}} \sim 10^{18}\text{cm}^{-2}$.

Historically, the observational data has been reported per unit “absorption distance” $\Delta X$; it is mostly a historical artifact and is defined as follows:

$$dX = \frac{H_0}{H(z)}(1 + z)^2dz,$$
whereas we measure the path length $\Delta l$ in the simulation in the comoving space. The conversion between $\Delta l$ and $\Delta X$ is:

$$\frac{dl}{dX} = \frac{(1+z)^2H_0}{c}. \tag{2}$$

We apply this conversion to the Crighton et al. (2019) observational fits to enable comparison between observations and simulations.

### 3.2. Association of LLSs with Galaxies

The association between LLSs and dense gas structures surrounding galaxies is well established neither at low nor at high redshift. A primary goal of this paper is to provide theoretical predictions of how LLSs at $z \sim 6$, near the end of cosmic reionization, are associated with galaxies and nearby structures.

In order to quantify the association of Lyman Limit Systems and galaxies in our simulation, we show in Figure 4 the fraction of LLSs whose neutral fraction peaks are within $n \times R_{\text{vir}}$ of any galaxy, i.e. the $f_n(R_{\text{vir}})$ quantity we discussed in Section 2.3 with $n \in \{1, 2, 3, 4\}$. The fraction is for all LLSs from each of our 40 $h^{-1}$cMpc simulation boxes. The overall trend with the neutral hydrogen column density is very clear: the damped Ly$\alpha$ systems of $\log_{10} N_{\text{HI}} > 20$ are predominantly located inside galactic halos. At lower column densities the association of LLSs with galaxies weakens, with a larger fraction of these LLSs located further than $2 \times R_{\text{vir}}$ from any galaxy. As we go down to the Ly$\alpha$ forest range, $N_{\text{HI}} < 10^{17}$cm$^{-2}$, only 10% of these systems are within $1 \times R_{\text{vir}}$ of a galaxy center and only 40% are within $2 \times R_{\text{vir}}$.

Of the population of highest column density LLSs (i.e. those with $10^{19}$cm$^{-2} < N_{\text{HI}} < 10^{20}$cm$^{-2}$), 80% are within $2 \times R_{\text{vir}}$ of some galaxy. Figure 5 shows the projected neutral hydrogen density around a LLS with the column density of $N_{\text{HI}} = 10^{19.4}$cm$^{-2}$, located within $1 \times R_{\text{vir}}$ of the virial radius of a galaxy. The left panel shows the projection in the y-z plane, and the right in the x-z plane. The depth of projection corresponds to the estimated size of the LLS ($\sim$10 kpc).

We indicate the LLS location with a green star, and surrounding dark matter halo centers with their virial radii respectively in black crosses and black circles. The galaxy associated with the LLS has a dark matter halo mass and virial radius of $M = 2.8 \times 10^9 M_\odot$ and $R_{\text{vir}} = 6.8$ kpc respectively; we show it in red.

The remaining 20% of the highest column density LLSs are outside of $2 \times R_{\text{vir}}$ of any galaxy. While these are not closely associated with any galaxy, we find that they are associated with overdense filamentary structures in the large-scale surrounding environments. Figure 6 shows the projected neutral hydrogen map surrounding such a system, with similar projections and color mapping as Figure 5. Here, we see that the LLS sits in a filamentary bridge connecting galaxies, but lies well outside twice the virial radius of nearby galaxies.

#### 3.3. Physical Properties of LLSs

In this section, we investigate the relationship between the optical depth of LLSs and their various physical properties in CROC simulations. The LLS physical properties of interest here are: optical depth, characteristic size, neutral fraction, and density.

Optical depth describes opacity of the LLSs and is directly proportional to the column density. Optical depth is defined as:

$$\tau = \sigma_{\text{LL}} N_{\text{HI}}, \tag{3}$$

where $\sigma_{\text{LL}} = 6.3 \times 10^{-18}$ cm$^2$ is the hydrogen photoionization cross section at the Lyman limit.

We define the characteristic size of a LLS as:

$$L_{\text{eff}} = N_{\text{HI}}/\sigma_{\text{LL}}, \tag{4}$$

where $n_{\text{HI}}$ is the peak HI number density of the LLS. If the peak had a Gaussian shape, that definition would closely correspond to the FWHM.

The top panel of Figure 7 illustrates the distribution of simulated Lyman Limit Systems in these four quantities. We use hexbin plots to clearly represent over-dense data points without overlapping them. The color bar on top of each plot represents the number of data points inside each hexagon. The horizontal dashed lines show the $\tau$ range corresponding to the LLS column density range, although we note that the numbers are arbitrarily defined.

We see from the middle and right panels that $\tau$ increases with the peak neutral fraction and the peak gas density of the LLSs. This means that higher column density LLSs are more neutral and denser at the center. However, there is not much of a trend between $\tau$ and $L_{\text{eff}}$ - i.e., at a fixed column density (or, equivalently, $\tau$), some LLSs are compact dense gas clumps while others are more spatially extended but less dense. Most LLSs, however, have sizes between 1 and $\sim$5 physical kpc. Note, that for LLSs with $\tau \geq 100$, their neutral fractions are close to unity. We also know that these high column density LLSs are strongly associated with nearby galaxies from Figure 4. This means that despite their proximity to ionizing sources, high column density LLSs still have highly neutral centers that are not affected by the ionizing photons.

It is also instructive to separate LLSs based on the strength of their association to galaxies, and plot the
physical quantities in the bottom panel of Figure 7. The colors we use here are the same as in Figure 4. It is clear that the strength of association with galaxies does not significantly affect the physical properties of LLSs at $z \sim 6$.

4. SUMMARY AND DISCUSSION

We use the state-of-the-art reionization simulation, CROC, to investigate the statistics, environment and physical properties of Lyman limit systems at $z \sim 6$. We cast a total of 3000 lines-of-sight to simulate synthetic quasar sight lines. Along these sight lines, we are able to identify Lyman Limit systems in the simulated IGM.

Our main results are:

1. The highest column density self-shielded absorbers, such as damped Ly$\alpha$ systems, have the strongest association with galaxies. These are typically within the virial radius of a galaxy.

2. Galaxy distribution largely dictates where we will find most of the LLSs; most LLSs are within 1-2 virial radii of a galaxy.

3. LLSs that are not located close to any galaxy are preferentially located along the filament connecting more distant galaxies.

4. The column density of LLSs strongly correlates with gas density and the hydrogen neutral fraction. Their characteristic sizes, however, show little correlation with other physical properties.

In future work, we plan to examine the main free path of LyC photons and relate it to the distributions of LLSs. With such future studies, we can develop a better understanding of what role the LLSs play in shaping the ionization history during the epoch of reionization.

JF was supported by the Summer Undergraduate Research Experience fund from the University of Michigan's Physics Department. CA acknowledges support from the Leinweber Center for Theoretical Physics. Fermilab is operated by Fermi Research Alliance, LLC, under Contract No. DE-AC02-07CH11359 with the United States Department of Energy. This work used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. An award of computer time was provided by the Innovative and Novel
Figure 5. Projected neutral hydrogen maps centered around a Lyman Limit system with $N_{HI} = 10^{19.4}$ cm$^{-2}$. The projection depth of the graph is $\sim 10$ kpc, corresponding to the characteristic size of the Lyman Limit system. Left: zy-projection plane. Right: xz-projection plane. The green star indicates the position of the Lyman Limit system; the black circles center on galaxies and indicate their corresponding dark matter halo virial radii; the red circle corresponds to the galaxy that contains the Lyman Limit system within its virial radius. Most of the high column density Lyman limit systems reside within one virial radius of a galaxy.

Figure 6. Projected neutral hydrogen maps centered around an example Lyman Limit system weakly associated with any galaxy. The projection depth is $\sim 10$ kpc. Markers indicate the same properties as in Figure 5. This Lyman Limit system is well outside of 2 times the virial radius of any galaxy. For illustration purposes, we annotate galaxies up to 5 virial radii away from the Lyman Limit system.

Computational Impact on Theory and Experiment (IN-CITE) program. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

REFERENCES


Figure 7. Lyman Limit system optical depth vs various physical properties of the Lyman Limit systems in CROC. From left to right: characteristic size, neutral fraction, and gas overdensity. The top panels are hexbin plots to more clearly visualize the distribution of data points with the color bar indicating the number of data points in each hexbin. The bottom panels show individual data points colored by their distance to the nearby galaxies at fixed multiples of the galaxy virial radius. The color coding is the same as in Figure 4: blue, orange, green, and red correspond to systems within 1, 2, 3, and 4 times the virial radius of any galaxy.