

A 210 Megaparsec Ionized Structure at $z = 6.8$ Revealed by an Astrophysics Foundation Model

Minghao Yue

yuemh@arizona.edu

The University of Arizona

Yongda Zhu

University of Arizona <https://orcid.org/0000-0003-3307-7525>

Jiani Ding

University of Arizona <https://orcid.org/0000-0003-4651-8510>

Xiaohui Fan

University of Arizona

Yufeng Luo

University of Wyoming

Article

Keywords:

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11272446/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

A 210 Megaparsec Ionized Structure at $z = 6.8$ Revealed by an Astrophysics Foundation Model

Minghao Yue^{1*†}, Yongda Zhu^{1*†}, Jiani Ding^{1†}, Xiaohui Fan¹,
Yufeng Luo²

¹*Steward Observatory, University of Arizona, 933 N. Cherry Ave.,
Tucson, 85721, Arizona, USA.

²Department of Physics and Astronomy, University of Wyoming, 1000
E. University Ave., Laramie, 82071, Wyoming, USA.

*Corresponding author(s). E-mail(s): yuemh@arizona.edu;
yongdaz@arizona.edu;

Contributing authors: jianiding@arizona.edu; xfan@arizona.edu;
ylo4@uwyo.edu;

[†]These authors contributed equally to this work.

Abstract

Reionization is the last major transition in the Universe that transforms the intergalactic medium (IGM) from neutral to highly ionized, yet its spatial structure remains poorly constrained. Using an astrophysics foundation model trained for James Webb Space Telescope datasets, we identify a luminous galaxy at redshift $z = 7.567$ showing significant Ly α transmission at $z = 6.80 \pm 0.07$. This transmission, named “the First Ly α transmission in the Reionization Epoch at $z \sim 7$ ” (FL α RE-z7), indicates a cosmological-scale ionized structure extending over $210.5^{+55.7}_{-53.5}$ comoving Megaparsecs at cosmic dawn. The hydrogen neutral fraction of the ionized structure is $x_{\text{HI}} = (2.07^{+0.29}_{-0.30}) \times 10^{-6}$, about 10^5 times lower than the cosmic mean at similar redshifts ($x_{\text{HI}} \approx 0.4$). Cosmological simulations suggest a probability of $\approx 10^{-5}$ for a random sightline to reproduce the transmission. We find no significant galaxy overdensity associated with the transmission region, in tension with canonical inside-out reionization models. FL α RE-z7 severely challenges existing reionization models in the standard Λ CDM Universe, indicating extreme spatial fluctuation in ionizing radiation or IGM temperature. Our result marks the first example of a foundation model discovering a previously unpredicted astrophysical phenomenon, demonstrating the power of deep learning methods for mining large astronomical datasets and finding unexpected new astrophysics.

Keywords: High-redshift Galaxies, Reionization, Intergalactic Medium, Deep learning

1 Main

Deep learning and Artificial Intelligence (AI) are revolutionizing the way scientific discoveries are made across a wide variety of disciplines [1–4]. Among these advances, foundation models – a family of self-supervised learning networks that learn the structure of large datasets without class labels – are especially suitable for identifying previously unknown objects and phenomena (e.g., [5, 6]). In astronomy and astrophysics, several foundation models have been recently developed (e.g., [7–9]), yet these studies focused on surveying known source populations and refining property measurements of known objects. Whether deep learning and foundation models can reveal unknown or unexpected astrophysical phenomena remains poorly explored. Answering this question is critical for advancing our understanding of the Universe, as ongoing and upcoming sky surveys are expected to deliver large numbers of “unknown unknowns”, which might encode new astrophysics and cannot be identified using pre-defined source selection algorithms.

In this work, we report the first foundation-model-driven discovery of an unpredicted astrophysical phenomenon: an ionized cosmological-scale structure in the epoch of reionization (EoR) when the Universe is substantially neutral. This ionized superstructure, which we name “the First Ly α transmission in the Reionization Epoch at $z \sim 7$ ” (FL α RE-z7), is not predicted by cosmological simulations and challenges our understanding of the early Universe and reionization. Reionization is the last major phase transition of the Universe, where the hydrogen in the intergalactic medium (IGM) turns from neutral to ionized from redshift $z \gtrsim 10$ to $z \approx 6$ [10–12]. Reionization affects about 90% of the baryonic matter in the Universe and has profound influences on galaxy evolution through cosmic time, yet its spatial structure remains uncertain. Previous models and observations of reionization suggested an IGM neutral fraction of $x_{\text{HI}} \approx 0.4$ around redshift $z \approx 7$ [13, 14]. In contrast, FL α RE-z7 exhibits a neutral fraction of $x_{\text{HI}} \approx 10^{-6}$ at about 200 comoving megaparsecs (Mpc) scales, which cannot be explained by canonical models of reionization in the standard Λ CDM Universe. FL α RE-z7 implies a highly patchy spatial structure for reionization, and its discovery demonstrates the ability of foundation models to discover unknown and unpredicted astrophysical phenomena.

A Foundation-Model-Based Framework Leads to the Discovery of FL α RE-z7

We use FM-JADES-v1 [15], a foundation model for James Webb Space Telescope (JWST) extragalactic studies, to develop discovery frameworks for new astrophysical objects and phenomena. Here we briefly summarize the information about FM-JADES-v1 and refer readers to [15] for detailed information. FM-JADES-v1 is trained on the JWST Advanced Deep Extragalactic Survey (JADES) data release 5 (DR5) imaging and photometry dataset [16–19]. For each object in JADES, the image patches

and catalog columns are tokenized and enter a Transformer block [20] that learns the relationships between different tokens. No class labels are assigned to objects during training. Each object is assigned a CLS token, which receives information from image and catalog tokens via the Transformer block and thus represents the object’s features. We regard the CLS tokens of objects as their embedded vectors, which are used to investigate the distribution of objects in the representation space.

In this work, we introduce the *population-conditioned outlier search* framework for accelerated discovery of objects with unusual or unexpected properties. Specifically, we start from a pre-defined population with certain astrophysical properties, obtain the embedding vector for these objects, and search for outliers in the embedded space to identify objects with extraordinary features (Figure 1, left). As an application of this framework, this paper searches for high-redshift galaxies ($z > 7$) with unusual properties in the JADES survey. To this end, we use the frozen FM-JADES-v1 model to produce the embedded vector for 122,336 objects in JADES, which are uniformly selected to have coverage of at least seven wide JWST NIRCam filters (see [15, 21] for details). The right panel of Figure 1 shows the distribution of these objects in the projected UMAP space [22]. We set the population condition to be “spectroscopically confirmed galaxies at $z > 7$ ” and investigate their distribution in the representation space (Methods). Most of the $z > 7$ galaxies are concentrated on the left side of the UMAP, forming a high-redshift island in the representation space (see [15] for details). Meanwhile, a few $z > 7$ galaxies are scattered across other regions on the UMAP. We compute the distance between each object and the median of all $z > 7$ galaxies, and rank the objects by their distance to the sample median. Top-ranked objects are expected to have unusual astrophysical features compared to the main high-redshift galaxy population. Following the ranked list, for each object, we feed the observations to Large Language Models (LLMs) including Claude [23] and ChatGPT[24] to generate hypotheses for the origin of each anomaly. At the final stage of the framework, the authors investigate the top outliers and hypotheses to identify potentially new astrophysical phenomena.

In the final ranked object list, the top two are Little Red Dots (LRDs), a population of active galactic nuclei recently discovered by JWST [25]. LRDs have distinct morphology and spectral energy distribution compared to other galaxies, which explains why these two objects are outliers. The two LRDs have been reported by previous studies (e.g., [26]). The third one, JADES-GS-217704 ($z = 7.567$), exhibits an unusual feature: there is significant flux bluer than its Ly α wavelength, where we expect non-detection due to neutral IGM attenuation. In the rest of the study, we will focus on the properties of JADES-GS-217704. More discussion of other outliers can be found in Methods.

Figure 2 shows the Hubble Space Telescope (HST) image, JWST NIRCam images, and NIRSpec/MSA Prism spectra of JADES-GS-217704 (see Methods for details about the observations). JADES-GS-217704 is a luminous galaxy at $z_{\text{spec}} = 7.567$ with $m_{\text{F115W}} = 26.17 \pm 0.01$. The redshift is securely determined by H β and [O III] emission lines (Methods). Its Ly α wavelength is $\lambda_{\text{Ly}\alpha} = 1.042\mu\text{m}$, redder than the

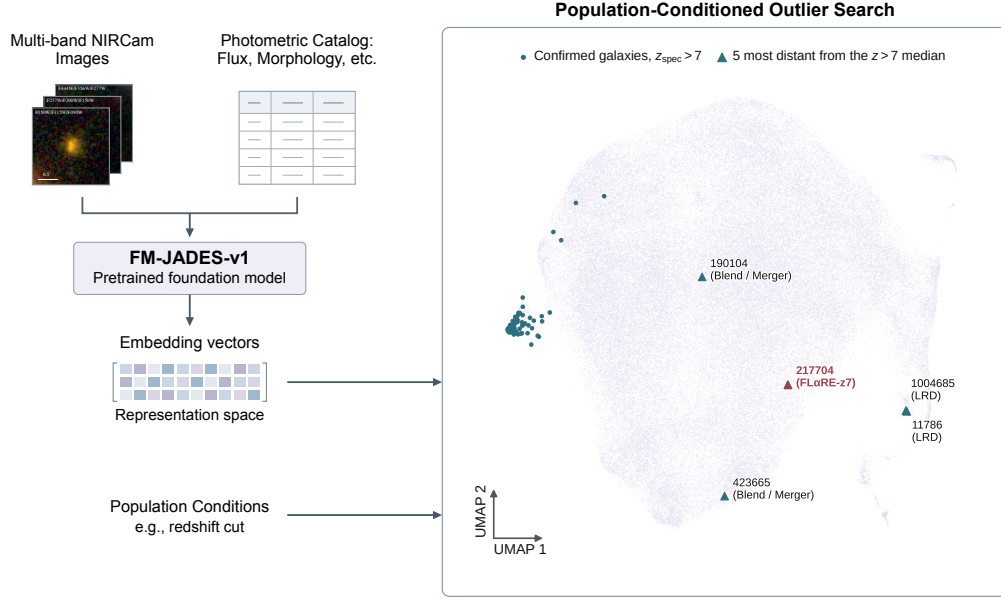


Fig. 1: Foundation-model-driven framework leads to the discovery of a high-redshift galaxy with extraordinary properties. We feed the multi-band NIRCcam images and photometric catalogs of 122,336 objects from the JADES DR5 dataset to the pre-trained astrophysics foundation model, FM-JADES-v1 [15]. For each object, FM-JADES-v1 encodes the observational data into an embedding vector. We then project the embedding vectors onto a two-dimensional plane using UMAP [22], and identify spectroscopically confirmed $z > 7$ galaxies among this sample. We then perform the Population-Conditioned Outlier Search by identifying outliers that are far from the center of the $z > 7$ galaxy distribution. These outliers are $z > 7$ galaxies that exhibit anomalous properties. The main target of this study, JADES-GS-217704 (red triangle), is among the top five outliers in the high-redshift galaxy sample. The small magenta dots mark the 122,336 JADES objects we investigate, the blue dots are $z > 7$ galaxies, and the triangles are the top five outliers.

coverage of the F090W filter ($0.795 - 1.005 \mu\text{m}$)¹. Meanwhile, JADES-GS-217704 has a significant F090W detection of $m_{\text{F090W}} = 28.20 \pm 0.08$, indicating positive flux bluer than its $\text{Ly}\alpha$ wavelength. The NIRSpect/MSA Prism spectrum also exhibits significant positive flux at $\lambda_{\text{obs}} \approx 0.95 \mu\text{m}$; when integrated between $0.9 \mu\text{m} < \lambda_{\text{obs}} < 1.0 \mu\text{m}$, this feature has a significance of 5.66σ . The significant flux bluer than $\lambda_{\text{Ly}\alpha}$ contradicts the standard picture of cosmic reionization, which suggests that the IGM at $z \approx 7$ has high neutral fractions of $x_{\text{HI}} \approx 0.4$ and should absorb all rest-frame ultraviolet

¹<https://jwst-docs.stsci.edu/jwst-near-infrared-camera/nircam-instrumentation/nircam-filters#gsc.tab=0>

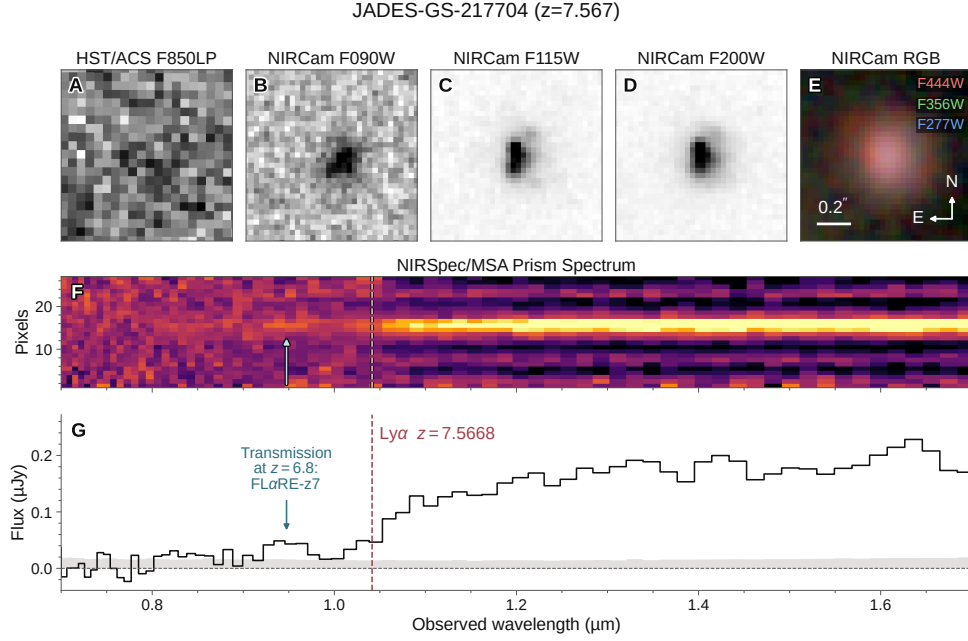


Fig. 2: HST and JWST observations for JADES-GS-217704. (A) The HST F850LP image. This object is undetected in deep optical imaging by HST. (B) NIRcam F090W imaging presents a 13σ detection, indicating integrated flux at wavelengths blueward of $\text{Ly}\alpha$ emission. (C-D) NIRcam F115W and F200W images. (E) NIRCam F277W, F356W, and F444W images, shown in the RGB format. (F and G) JWST NIRSpec/MSA Prism 2D spectrum and 1D spectrum. The red dashed line denotes the position of the $\text{Ly}\alpha$ break, and the teal arrow marks the position of the transmission feature (FLARE-z7). The gray-shaded region in the 1D spectrum indicates the 1σ error. The 1D spectrum indicates a 5.66σ detection between $0.9\mu\text{m}$ and $1\mu\text{m}$.

(UV) photons bluer than $\lambda_{\text{Ly}\alpha}$ for $z \gtrsim 7$ galaxies. To produce this feature, there must be an ionized structure on the sightline of JADES-GS-217704 at $z \approx 6.8$ with a very low hydrogen neutral fraction ($x_{\text{HI}} \lesssim 10^{-6}$). We refer to this ionized structure as “the First $\text{Ly}\alpha$ transmission in the Reionization Epoch at $z \sim 7$ ” (FLARE-z7). We note that the flux cannot come from a foreground interloper, given its HST photometry and the spectral shape at $\lambda_{\text{obs}} < 1\mu\text{m}$ (see Methods for details).

The discovery of FLARE-z7 challenges canonical models of the early Universe. Below we determine the properties of FLARE-z7 and discuss its implications for how reionization happens.

A Cosmological-Scale Ionized Superstructure in the Neutral Early Universe

To determine the redshift and size of the ionized structure, we fit the transmission signal between $0.9\mu\text{m}$ and $1\mu\text{m}$ as a Gaussian profile. We adopt the NIRSpec/MSA Prism Line Spread Function (LSF) from [27]. Figure 3 shows the best-fit model, which indicates a central wavelength of $\lambda_c = 0.948 \pm 0.008\mu\text{m}$ and a FWHM of $0.07 \pm 0.02\mu\text{m}$. These results translate to a redshift of $z = 6.80 \pm 0.07$ and a size of $210.5^{+55.7}_{-53.5}$ comoving Megaparsec (cMpc) for the ionized structure (Methods).

We also evaluate the effective Ly α optical depth for two wavelength windows centered at λ_c , with widths of FWHM and 50 cMpc/h, respectively. We use the 50 cMpc/h window following previous studies of the reionization-era IGMs [11, 28, 29]. We reconstruct the intrinsic continuum of the galaxy by fitting a spectral energy distribution (SED) using *Prospector* [30] (see details in Methods). Our analysis yields an effective optical depth of $\tau_{\text{eff}} = 1.28 \pm 0.18$ for the FWHM window and 1.08 ± 0.24 for the 50 cMpc/h window. The corresponding hydrogen neutral fraction of $x_{\text{HI}} = (2.07^{+0.29}_{-0.30}) \times 10^{-6}$ and $(1.74^{+0.38}_{-0.39}) \times 10^{-6}$, respectively (Methods). FL α RE-z7 has an optical depth comparable $z \approx 5$ IGM, which is much lower than the τ_{eff} at $z > 6.5$ measured by previous studies (Figure 4).

At $z \approx 6.8$ (about 800 Myrs after the Big Bang), the Universe is highly neutral, with an IGM hydrogen neutral fraction of $x_{\text{HI}} \approx 0.4$ [11, 29, 31]. Such a high neutral fraction produces complete Gunn-Peterson absorption troughs [32] for $z > 7$ galaxies, leaving no flux bluer than their Ly α wavelengths. Although ionized bubbles at $z > 7$ have been discovered by identifying Ly α emission lines [33, 34], these ionized bubbles are small (with sizes of $\lesssim 10$ cMpc) and are localized around luminous galaxies or active galactic nuclei (AGN). In contrast, FL α RE-z7 has a size of $\gtrsim 200$ cMpc, much larger than ionized bubbles produced by galaxies and even by the most luminous quasars [35]. We also check objects in the JADES DR5 catalog around the FL α RE-z7 sightline, and found no luminous objects ($m_{\text{F115W}} < 24$) within 3 arcminutes with photometric redshift $|z_{\text{phot}} - z_c| < 0.2$. We therefore conclude that FL α RE-z7 is an ionized superstructure at cosmological scales, rather than an ionized bubble produced by a single luminous object.

FL α RE-z7 Challenges Existing Models for the Early Universe

The existence of a highly-ionized superstructure like FL α RE-z7 challenges current models of the early Universe, in particular for its reionization history. We demonstrate this point by comparing the optical depth of FL α RE-z7 with cosmological simulations in [36]. To consider the impact of different astrophysical parameters on reionization, [36] considered three models, namely the high cosmic wave background optical depth (τ_{CMB}) model, the low τ_{CMB} model, and the hot low τ_{CMB} model. A high τ_{CMB} indicates earlier reionization, and the hot model means higher energy for the ionizing photons. For each model, [36] provided the Ly α optical depths for 20,000 skewers, and we evaluate the τ_{eff} for these skewers using the same wavelength window as for FL α RE-z7.

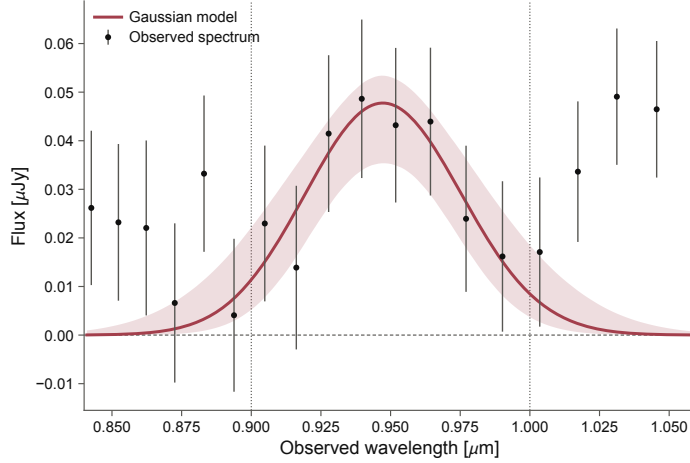


Fig. 3: Determining the redshift and the width of the transmission signal. We fit a Gaussian profile for the transmission signal, adopting the LSF from [27]. The dotted lines mark the fitting window, and the shaded region marks the 1σ posterior of the Gaussian model. The model suggests a central wavelength of $\lambda_c = 0.948 \pm 0.008 \mu\text{m}$, corresponding to a redshift of $z_c = 6.80 \pm 0.07$. The FWHM of the transmission is $0.07 \pm 0.02 \mu\text{m}$, corresponding to $210.5^{+55.7}_{-53.5}$ cMpc.

We evaluate the probability distribution of τ_{eff} for the simulated skewers. We use the continuum model from SED fitting as the background source spectrum at $z = 7.567$ and apply the IGM attenuation of each skewer on the model continuum. We then compute the transmission flux for the 50 cMpc/h centered at FLARE-z7’s redshift (Methods). All three models fail to produce sightlines with $\tau_{\text{eff}} < 4$. After including measurement errors, we estimate that the probability for a random sightline to have a measured flux higher than FLARE-z7 is $\approx 10^{-5}$ for all three models. Since FLARE-z7 is discovered among 100 spectroscopically confirmed $z > 7$ galaxies, the probability of finding such a transparent sightline in our entire experiment is about 10^{-3} .

This result indicates a tension between the discovery of FLARE-z7 and the existing cosmological simulations for reionization. This tension cannot be explained by an early reionization model, as the high- τ_{CMB} model fails to produce such a sightline. The most plausible solution is a highly patchy reionization model, where the ionizing radiation or IGM temperature has large fluctuations at $\gtrsim 50$ cMpc scales. This picture is consistent with the $\gtrsim 100$ cMpc neutral IGM structures at the end of reionization ($z \approx 5.5$), when the Universe has been largely ionized [37, 38].

No Evidence for Significant Galaxy Overdensities around FLARE-z7

In the canonical “inside-out” picture of reionization, large ionizing structures at $z \gtrsim 6$ should reside in galaxy overdensities or protoclusters [40, 41], as copious amounts of ionizing photons from galaxies are required to produce large ionized bubbles. Nevertheless, this canonical picture has been challenged by recent cosmological simulations

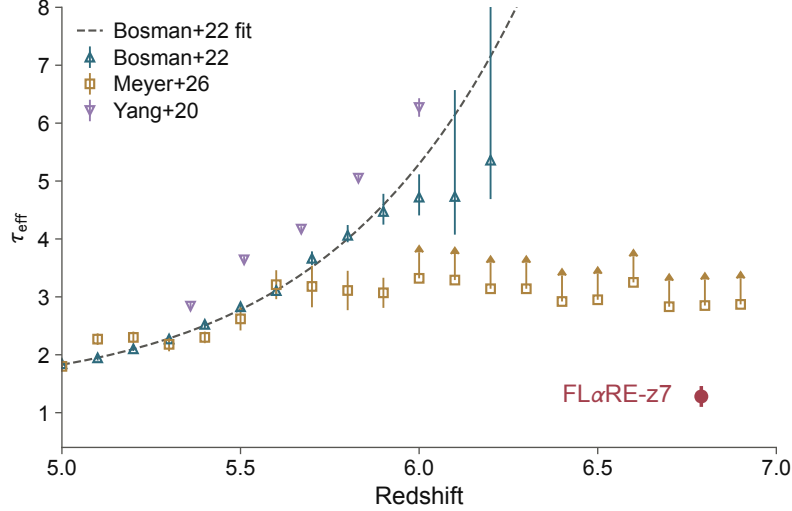


Fig. 4: The IGM $\text{Ly}\alpha$ optical depth as a function of redshift. We include IGM optical depth constraints from high-redshift quasar and galaxy observations [11, 12, 39]. The IGM at $z \approx 5.5$ has optical depth $\tau_{\text{eff}} \approx 2$, which quickly increase to $\tau_{\text{eff}} > 4$ at $z \gtrsim 6.5$. The optical depth FLARE-z7 ($\tau_{\text{eff}} = 1.28 \pm 0.18$) is similar to the cosmic mean value at $z \approx 5$, indicating near-complete reionization progress for FLARE-z7 despite its high redshift.

[42], which suggested that the IGM density in galaxy overdensities is too high to get ionized. Instead, the simulations support the “outside-in” reionization, where ionized structures emerge from underdense regions in the Universe.

FLARE-z7 provides a direct test of the two pictures about reionization. Figure 5 shows the distribution of objects with $|z_{\text{phot}} - z_c| < 0.2$ in the JADES GOODS-S field. We only count areas with at least seven NIRCам wide filters coverage and apply magnitude cuts to ensure a uniform coverage across the field (see Methods for details). The average object surface number density is $1.44 \pm 0.10 \text{ arcmin}^{-2}$ across the entire GOODS-S field, and is $2.28 \pm 0.44 \text{ arcmin}^{-2}$ within $r < 2 \text{ arcmin}$ around FLARE-z7’s sightline. In other words, there is no evidence for a strong overdensity around FLARE-z7. This result indicates that ionized superstructures in the EoR do not appear to coincide with obvious concentrations of ionizing sources. Similar behavior is seen near the end of reionization, where some of the clearest IGM $\text{Ly}\alpha$ transmission occurs in average-density or underdense regions [31, 43]. Recent JWST observations likewise found higher galaxy $\text{Ly}\alpha$ visibility in lower-density environments over $4.8 < z < 11$ [44].

The origin of such a large ionized structure without a strong galaxy overdensity remains unclear. Very metal-poor stellar populations can produce harder and more efficient ionizing spectra, and unusually high escape fractions or faint or obscured AGN could further increase the ionizing emissivity [45]. Anisotropic escape provides another possibility: ionizing photons need not trace the projected galaxy distribution, and

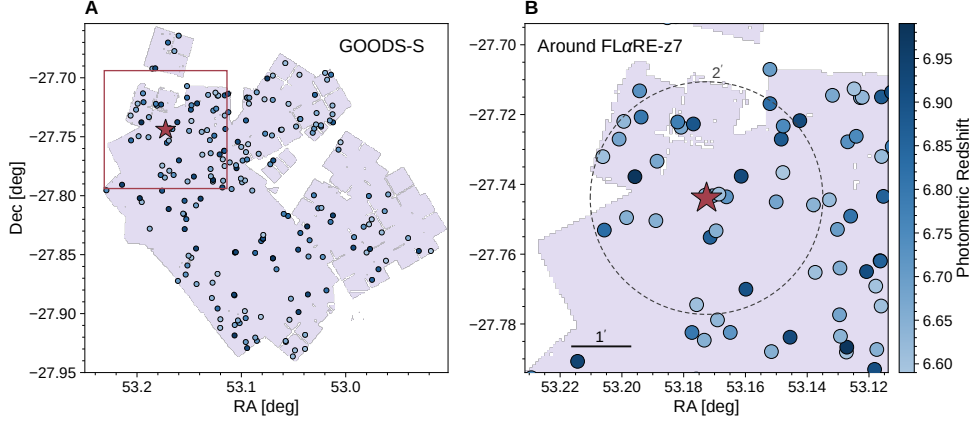


Fig. 5: The environment of FLaRE-z7. (A) the GOODS-S field; (B) the region around FLaRE-z7, corresponding to the red box in the left panel. The background shaded area shows the footprint that reaches a given set of depth criteria (defined in Methods); the dots are objects brighter than the depth criteria in the JADES DR5 catalog, colored by their photometric redshifts. We identify objects in this footprint with photometric redshifts $|z_{\text{phot}} - 6.8| < 0.2$. The entire field has an object surface number density of $\Sigma = 1.44 \pm 0.10 \text{ arcmin}^{-2}$, and the area within $2'$ around FLaRE-z7 has $\Sigma = 2.28 \pm 0.44 \text{ arcmin}^{-2}$. In other words, there is no evidence for an overdensity around FLaRE-z7.

both observations and simulations suggest that directional escape can substantially alter the geometry of reionization [31, 46, 47]. With a size of $\sim 200 \text{ cMpc}$, FLaRE-z7 may consist of multiple overlapping ionized regions rather than a single dominant structure. Deeper imaging and spectroscopy of galaxies throughout and around the transmission region will shed light on the origin of this structure.

Summary

We report the discovery of FLaRE-z7, an ionized superstructure at $z = 6.8$ with a line-of-sight size of $\sim 200 \text{ cMpc}$. FLaRE-z7 is located on the sightline of a luminous $z = 7.567$ galaxy, JADES-GS-217704, which shows significant positive flux bluer than its Ly α break in both NIRC*am* images and NIRS*pec*/MSA Prism spectrum. Using the transmission flux, we estimate a Ly α effective optical depth of $\tau_{\text{eff}} = 1.28 \pm 0.18$, corresponding to a neutral fraction of $x_{\text{HI}} = (2.07^{+0.29}_{-0.30}) \times 10^{-6}$. FLaRE-z7 provides the first observational evidence of near-complete reionization in such early cosmic times, when the Universe is overall substantially neutral. The environment of FLaRE-z7 suggests that reionization can happen early even without strong galaxy overdensities. FLaRE-z7 is not reproduced by cosmological simulations, challenging the canonical model of reionization in a ΛCDM Universe.

The discovery of FLaRE-z7 is driven by the foundation model FM-JADES-v1, which identifies the object as a $z > 7$ galaxy with unusual observational features.

We exploit the population-conditioned outlier search framework in our study, a novel approach of using foundation models for new phenomenon discovery. Notably, FM-JADES-v1 was not explicitly trained for IGM or reionization studies. This result emphasizes the potential of foundation models as general discovery machines for unknown or unexpected astrophysical phenomena. As ongoing wide-field sky surveys (like the Euclid survey, the Roman Space Telescope Surveys, and the Vera C. Rubin Legacy Survey for Space and Time) expand object samples by orders of magnitude, foundation models offer a way to identify rare and unexpected physical signatures across these much larger datasets, opening a new path for discovery in all fields of astrophysics.

2 Method

Cosmology and Magnitude System

We adopt a Flat Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$ and $\Omega_M = 0.3$. All magnitudes in this work are AB magnitudes [48].

HST and JWST Observations

The images and photometric information used in this work are from JADES DR5. JADES DR5 provides NIRCам mosaic images and photometric catalogs in GOODS-S and GOODS-N fields [17–19]. The total coverage is about 469 arcmin^2 depth of these imaging data varies from about 28 to 30 magnitudes. The JADES fields are also covered by HST observations, and the JADES DR5 catalog provides HST photometry when available. We obtain the HST Advanced Camera for Survey (ACS) images of objects from DAWN JWST Archive (DJA)², which were reduced using `grizli` [49, 50].

The spectra used in this work are obtained from the DJA, which provides publicly available JWST NIRSpec/MSA spectra uniformly reduced by the code `msaexp` [51]. For the spectrum of JADES-GS-217704, we multiply the uncertainty by a factor of 1.48. This factor is motivated by the `Prospector` SED fitting model. Specifically, we compute the difference between the best-fit SED model and the observed spectrum at $0.6\mu\text{m} < \lambda_{\text{obs}} < 1.5\mu\text{m}$, excluding the $0.8\mu\text{m} < \lambda_{\text{obs}} < 1.2\mu\text{m}$ window that is contaminated by the transmission signal and the Ly α damping wing. We then compute the noise scaling factor so that the reduced χ^2 in the selected window equals one. This procedure is consistent with previous studies, which found that the error of NIRSpec MSA spectra is usually underestimated.

JADES-GS-217704 is also observed by NIRSpec/MSA using the F070LP/G140M, F170LP/G235M, and F290LP/G395M setups. Figure 6 shows all the NIRSpec observations for JADES-GS-217704 available at DJA. The H β and [O III] emission lines are clearly detected.

We also note that JADES-GS-217704 is independently reported in [52].

²<https://dawn-cph.github.io/dja/index.html>

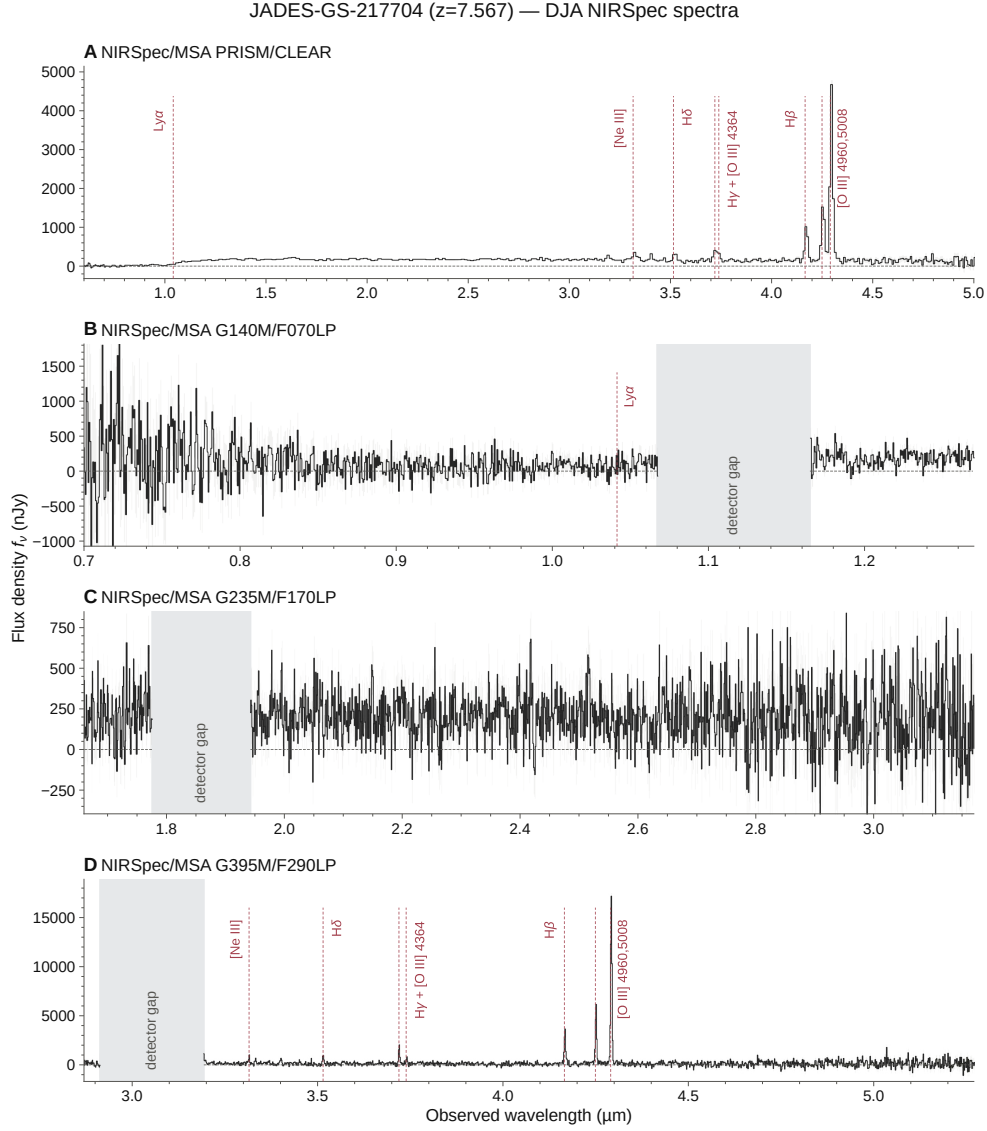


Fig. 6: The JWST NIRSpec/MSA spectroscopy for JADES-GS-217704. (A) the Prism spectrum. (B) the F070LP/G140M spectrum. (C) the F170LP/G235M spectrum. (D) the F290LP/G395M spectrum. The redshift of this galaxy is securely determined by multiple emission lines.

The FM-JADES-v1 Model and Population-Conditioned Outlier Search

We use the self-supervised foundation model, FM-JADES-v1 [15], to perform the Population-Conditioned Outlier Search for high-redshift galaxies. FM-JADES-v1 takes JADES DR5 NIRC*am* images and photometric catalogs as inputs, tokenizes these data, and feeds these tokens into a Transformer block to learn the relations between the tokens. We refer the readers to [15] for the architecture and training of the model.

In this work, we use the frozen FM-JADES-v1 model to compute the CLS embedding vector for JADES objects. Following [15], we focus on a subsample of JADES objects, which have (1) coverage by F090W, F115W, F150W, F200W, F277W, F356W and F444W imaging; (2) signal-to-noise ratio (SNR) higher than 10 in the F356W and F444W bands. We only use image and catalog tokens relevant to these seven broad bands as inputs to produce the CLS embedding vector, which avoids the embedding vector encoding survey coverage or depth information rather than astrophysical information. The corresponding sample contains 122,336 objects. We then project the CLS embedding vectors onto a two-dimensional UMAP. The structure of this UMAP projection has been discussed in [15]; this work focuses on its application in identifying high-redshift galaxies with unusual properties.

To perform the Population-Conditioned Outlier Search, we start from $z > 7$ galaxies in the JADES DR4 spectroscopic dataset [53, 54], where we require grade A or B for redshift determination. This criterion identifies 100 galaxies with $z > 7$ among the 122,336 objects. We obtain the embedding vector of this sample, compute the median position of the 100 galaxies on the UMAP, then compute the distance from this median to each galaxy and rank the galaxies using these distances. The top five outliers are highlighted in Figure 1. The main text discusses JADES-GS-217704 and the ionized superstructure FL*α*RE-*z*7 on its sightline. Figure 7 to Figure 10 show the other four outliers identified by this framework.

The probability of a foreground interloper

Before interpreting the transmission signal as an ionized structure, we carefully consider the possibility of a foreground interloper. In this scenario, JADES-GS-217704 does not have any flux bluer than its Ly*α* due to IGM attenuation, and a foreground object gives the signal we see at $\lambda_{\text{obs}} \approx 0.95\mu\text{m}$. The F090W signal is entirely contributed by the foreground interloper.

We evaluate the number density of possible interlopers as follows. Using the CIRC3 magnitudes in the JADES DR5 catalog, we find that JADES-GS-217704 has non-detections (below 2σ levels) in the HST F435W, F606W, F775W, and F850LP bands. and a signal at $\approx 3.5\sigma$ level in the F814W band. We note that F814W covers part of the positive signal at $\lambda_{\text{obs}} > 0.9\mu\text{m}$, and a marginal detection is plausible. Correspondingly, we select objects in the JADES catalog with $< 2\sigma$ signals in F435W, F606W, F775W, and F850LP, as well as $< 4\sigma$ signals in F814W. Since JADES-GS-217704 has CIRC3 $m_{\text{F090W}} = 28.20 \pm 0.08$, we further require the possible interloper to have $m_{\text{F090W}} < 28.3$. These criteria select $N = 1,547$ objects in the entire JADES

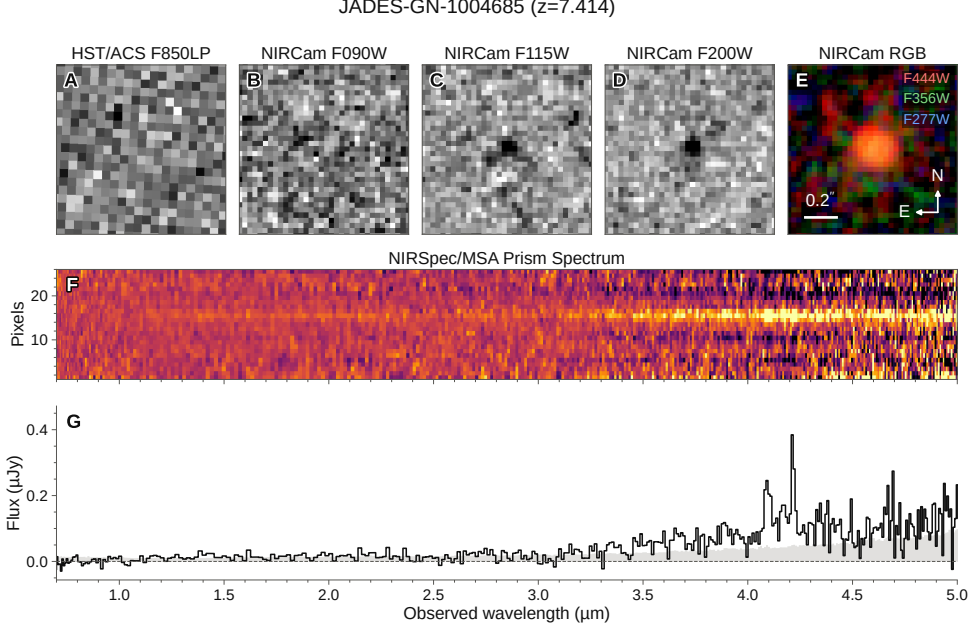


Fig. 7: HST and JWST observations for JADES-GN-1004685. This object is a previously reported LRD (e.g., [26]), explaining the reason for it to be an outlier. (A) the HST F850LP imaging. (B) NIRcam F090W imaging. Both of these bands show clear non-detections. (C-D) NIRcam F115W and NIRcam F200W images. (E) NIRcam F277W, F356W, and F444W color image. (F and G)) NIRSpect/MSA Prism Spectrum.

footprint. Assuming a maximum projected distance of $\Delta\theta = 0.1''$ between the interloper and the background galaxy, the chance for a background galaxy to hit such an interloper is

$$P(\text{interloper}) = N\pi\Delta\theta^2/A = 2.87 \times 10^{-5} \quad (1)$$

where $A = 469 \text{ arcmin}^2$ is the area of the JADES footprint.

Another argument against the interloper scenario is the spectral shape at $\lambda_{\text{obs}} < 1\mu\text{m}$. The positive flux peaks at $\lambda_{\text{obs}} \approx 0.95\mu\text{m}$ and drops towards both shorter and longer wavelengths, reaching about zero at $\lambda_{\text{obs}} \approx 0.9\mu\text{m}$ and $1\mu\text{m}$. If this feature is an emission line, the inferred line width is $2.3 \times 10^4 \text{ km s}^{-1}$ according to the Gaussian fit in Figure 3, which is too large even for broad-line AGNs. If the positive flux is from galaxy continuum, then the spectrum at $\lambda_{\text{obs}} < 1\mu\text{m}$ should be relatively flat, which cannot explain the gap at $\lambda_{\text{obs}} \sim 1\mu\text{m}$. We also search for foreground emission lines or absorption lines in the spectrum $\lambda_{\text{obs}} > 1\mu\text{m}$, and find no evidence for a foreground galaxy.

The spectral features, together with the low interloper probability, essentially rule out the foreground interloper scenario.

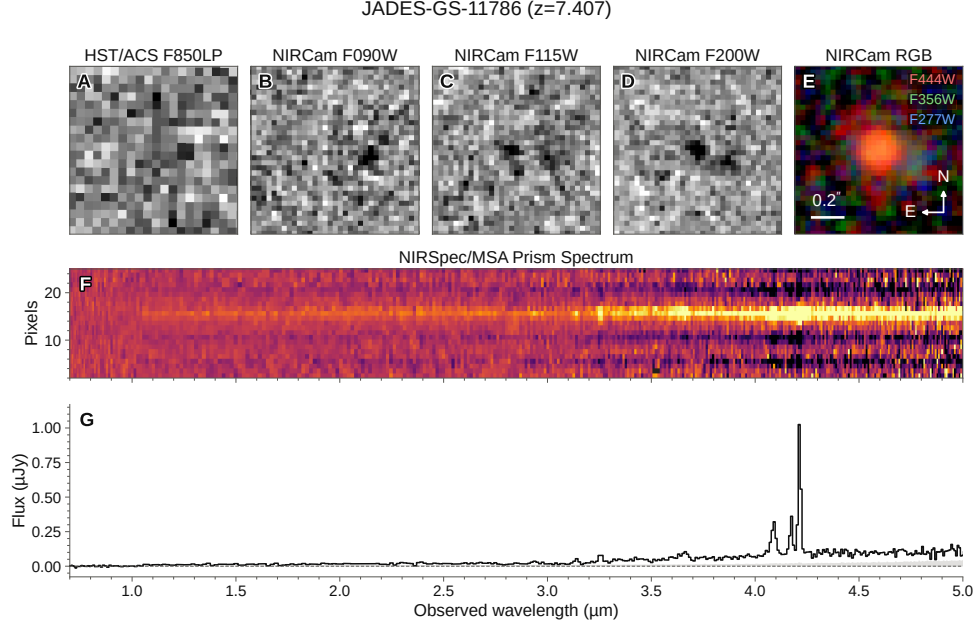


Fig. 8: Same as Figure 7, but for outlier JADES-GS-11786, which is a published LRD (e.g., [26]).

The wavelength-dependent morphology of JADES-GS-217704

JADES-GS-217704 exhibits clumpy structures, and its morphology varies with wavelengths. One noticeable feature is the northern clump visible in the F115W and F150W images; this feature is not visible in the F090W image. In this section, we show that the wavelength-dependent morphology of JADES-GS-217704 can be explained by differential extinction.

To this end, we perform pixel-by-pixel SED fitting for the galaxy using the PhotoIFU package PhotoIFU [55]. In short, PhotoIFU takes NIRCcam images as inputs, then performs SED fitting using Prospector for each pixel to get spatially resolved galaxy properties. All input NIRCcam images are matched to the point spread function (PSF) of the F444W band; the PSF-matched images are obtained from JADES DR5. We exclude the F090W image from the fitting as standard Prospector runs cannot model IGM transmission like FLaRE-z7. To model possible different extinction curves in different regions (e.g., [56]), we make the following modifications to PhotoIFU. We use `dust_type=4` in Prospector, corresponding to a power-law modification to the Calzetti dust attenuation curve [57], and allow the extinction curve power law index to vary across pixels. We also include a `dust1` component that only attenuates emission from young stars (with $t_{\text{age}} < 10$ Myr).

Figure 11 shows the overlay between the best-fit dust extinction map (A_{λ}) and the surface brightness contours in the F090W, F115W, and F150W images. The F090W transmission is located exactly at the position where the dust attenuation is the lowest

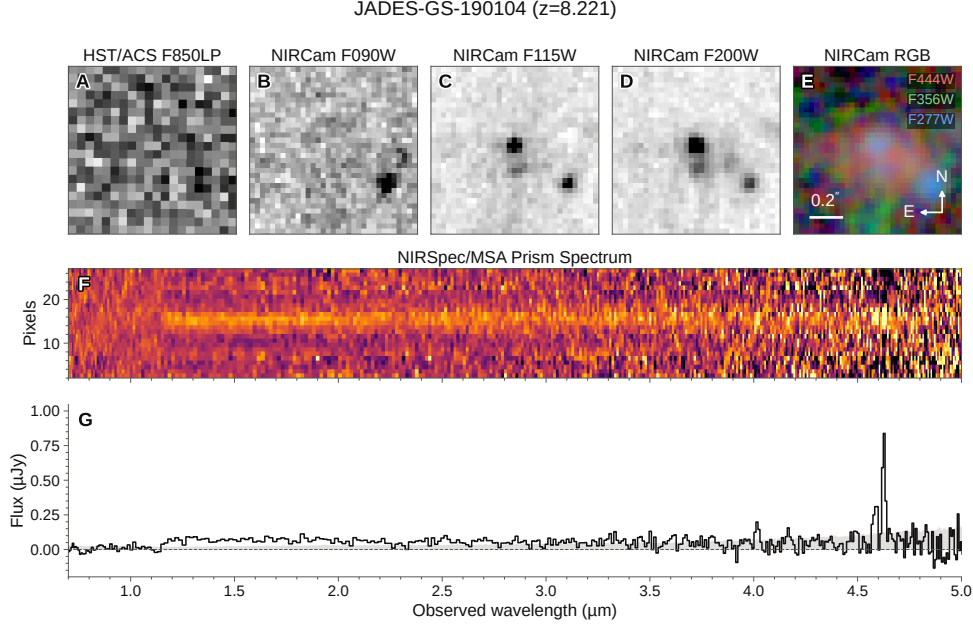


Fig. 9: Same as Figure 7, but for outlier JADES-GS-190104. This galaxy has clear companions and nearby foreground contaminant.

($A_\lambda \lesssim 1$ for the F090W filter). Meanwhile, the northern clump is located in the region with the largest attenuation, which has $A_\lambda \gtrsim 2.5$ in the F090W band. We note that the northern clump contributes to $\approx 20\%$ of the total flux in the F115W and the F150W images. A non-detection for the northern clump in the F090W band is consistent with its high dust attenuation and the integrated flux for the entire galaxy ($m_{\text{F090W}} = 28.2 \pm 0.08$, a 13σ detection) in the F090W band. Therefore, we suggest that JADES-GS-217704's wavelength-dependent morphology is a result of differential extinction [56].

Characterizing the transmission signal

We measure the redshift and size of the transmission signal by fitting a Gaussian profile for the one-dimensional spectrum between $0.9\mu\text{m}$ and $1\mu\text{m}$. We use the LSF from [27] in the fiducial fit shown in Figure 3. The model has three free parameters: the central wavelength λ_c , the amplitude A , and the FWHM of the transmission signal. We use the nested sampling algorithm Nautilus [58] to obtain the posterior of these parameters, where we adopt flat priors for all parameters. The fit yields $A = 0.0474^{+0.0094}_{-0.0095} \mu\text{Jy}$, $\lambda_c = 0.948 \pm 0.008 \mu\text{m}$, $\text{FWHM} = 0.07 \pm 0.02 \mu\text{m}$. The best-fit model is shown in Figure 3.

We also evaluate the optical depth of FLARE-z7. To this end, we fit the galaxy's SED using Prospector [30]. The input data includes the *HST* ACS, *JWST* NIRCam, and *JWST* MIRI CIRC3 photometry provided by the JADES DR5 photometric catalog,

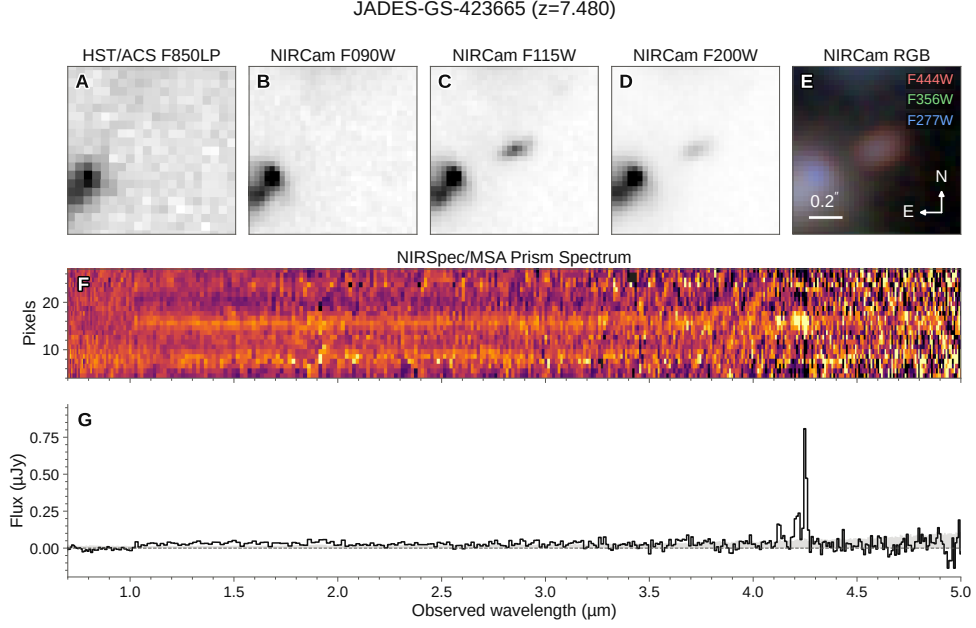


Fig. 10: Same as Figure 7, but for outlier JADES-GS-423665. This galaxy has bright foreground contaminants nearby.

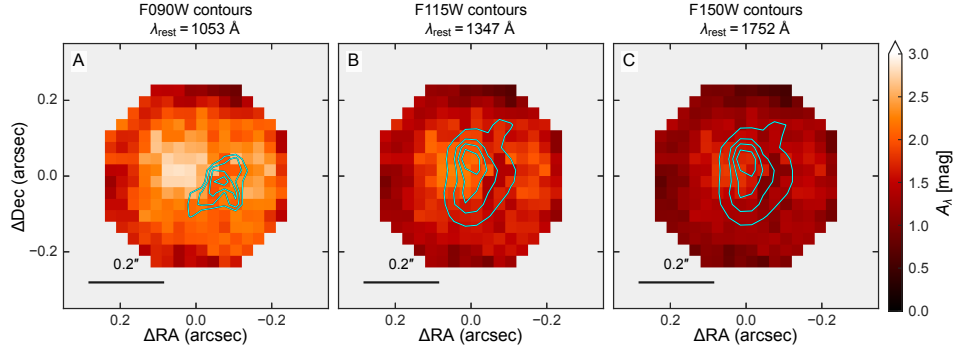


Fig. 11: The attenuation map of JADES-GS-217704 in three NIRCам short wavelength bands, derived from pixel-by-pixel SED fitting using PhotoIFU [55]. The map shows a high extinction region at the northeast corner and a low extinction “hole” near the center of the galaxy. The contours show the surface brightness in each band. The pixel-by-pixel SED fitting clearly indicates differential reddening across the galaxy. The F090W has the strongest extinction, with the flux dominated by the low-extinction region.

as well as the NIRSpect/MSA Prism spectrum. We exclude the F090W magnitude and the Prism spectrum at $0.9\mu\text{m} < \lambda_{\text{obs}} < 1.2\mu\text{m}$ to avoid the influence of the transmission signal and the Ly α damping wing. Figure 12 shows the best-fit SED and its posterior distribution. By turning off the IGM attenuation for the fitted SED, we can get the intrinsic continuum at $\lambda_{\text{obs}} < 1\mu\text{m}$.

Using the modeled intrinsic continuum, we measure the optical depth of FL α RE-z7 using the wavelength range $0.912\mu\text{m} < \lambda_{\text{obs}} < 0.983\mu\text{m}$, corresponding to $|z - z_c| < \text{FWHM}/2$. We compute the transmission flux and the modeled unattenuated continuum flux by summing the spectrum within the window. The effective optical depth is computed as

$$\tau_{\text{eff}} = -\ln\langle T \rangle = -\ln(F_{\text{transmission}}/F_{\text{cont}}), \quad (2)$$

We compute the posterior distribution of τ_{eff} using the posterior distributions of F_{cont} and $F_{\text{transmission}}$, yielding $\tau_{\text{eff}} = 1.28 \pm 0.18$. We convert the optical depths to neutral fractions using the Gunn-Peterson relation (e.g., [10, 32]):

$$\tau_{\text{eff}} = 1.8 \times 10^5 h^{-1} \Omega_m^{-1/2} \left(\frac{\Omega_b h^2}{0.02} \right) \left(\frac{1+z}{7} \right)^{3/2} x_{\text{HI}} \quad (3)$$

We use $\Omega_b h^2 = 0.0224$ from [59]. This gives $x_{\text{HI}} = (2.07^{+0.29}_{-0.30}) \times 10^{-6}$ for FL α RE-z7. We also perform the same measurements for the 50 cMpc/ h window around the central wavelength, yielding $\tau_{\text{eff}} = 1.08 \pm 0.24$ and $x_{\text{HI}} = (1.74^{+0.38}_{-0.39}) \times 10^{-6}$.

We note that the transmission at $z \approx 1\mu\text{m}$ is consistent with zero, indicating that FL α RE-z7 is not powered by the background luminous galaxy JADES-GS-217704 at $z = 7.567$. The distance from JADES-GS-217704 to the edge of FL α RE-z7 (i.e., from $z = 7.567$ to $z = 7$) is about 185 cMpc, much larger than the proximity zone sizes of the most luminous quasars [35]. The absence of the Ly α emission line and the existence of a Ly α damping wing for JADES-GS-217704 further reinforce this conclusion.

The tension between FL α RE-z7 and cosmological simulations

We compare the evaluated τ_{eff} of FL α RE-z7 with [36], who simulated the IGM optical depth for three reionization models with different ionizing photon energy and CMB optical depth: the low τ_{CMB} model with $\tau_{\text{CMB}} = 0.051$ and $E_\gamma = 17.1\text{eV}$, the high τ_{CMB} model with $\tau_{\text{CMB}} = 0.071$ and $E_\gamma = 17.1\text{eV}$, and the hot low τ_{CMB} model with $\tau_{\text{CMB}} = 0.051$ and $E_\gamma = 18.6\text{eV}$. We refer to [36] for a detailed setup of these models. For each model, [36] provides 20,000 skewers, and we obtain the optical depths for the simulated skewers in the 50 cMpc/ h window centered at z_c . None of these skewers have $\tau_{\text{eff}} < 4$. In other words, the simulations fail to produce sightlines like FL α RE-z7 (Figure 13).

We further evaluate the probability of finding a sightline like FL α RE-z7 in the simulated skewers given the measurement errors. To do this, we obtain simulated spectra by applying the simulated IGM transmission to the best-fit SED model, then add Gaussian random noise to the simulated spectrum using the observed spectrum error. We then compute the transmission signal in the 50 cMpc/ h window centered on z_c for the simulated spectra. We choose the size of 50 cMpc/ h to match previous

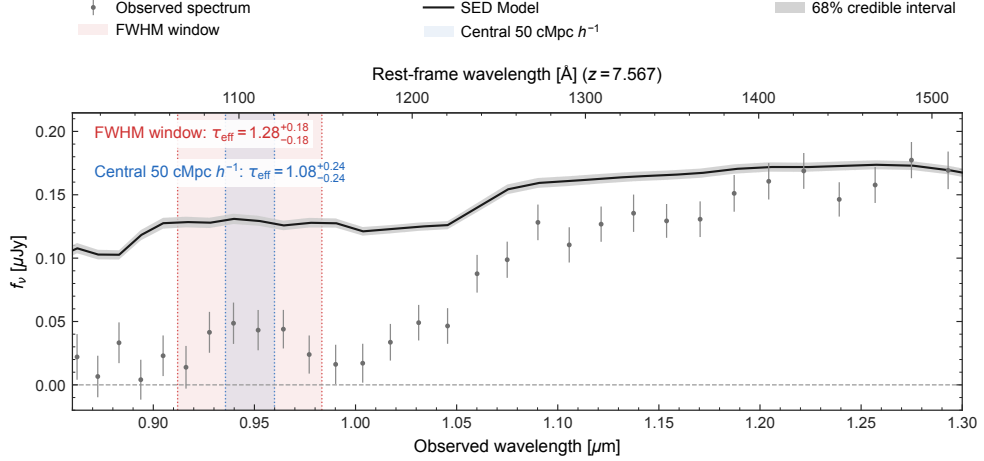


Fig. 12: Continuum fitting and optical depth measurement for FLaRE-z7. The black solid line shows the best-fit SED model from Prospector, with IGM attenuation turned off. The gray shaded area marks the 1σ error. Using this model, we evaluate the optical depth of FLaRE-z7 in the FWHM-width window and the central 50 cMpc/h window, yielding $\tau_{\text{eff}} = 1.28 \pm 0.18$ and $\tau_{\text{eff}} = 1.08 \pm 0.24$, respectively.

studies of reionization-era IGM (e.g., [28]). We estimate the chance that a random simulated sightline produces the observed spectrum of FLaRE-z7 to be 9.14×10^{-6} , 9.16×10^{-6} , 5.00×10^{-5} for the low τ_{CMB} model, the hot low τ_{CMB} model, and the high τ_{CMB} model, respectively. Since FLaRE-z7 is discovered from a sample of 100 spectroscopically confirmed $z > 7$ galaxies, the probability of finding such a system in our entire experiment is $\approx 5 \times 10^{-3}$. Therefore, FLaRE-z7 is in marginal tension with the reionization models in [36]. The small probabilities evaluated using the three models indicate that this tension cannot be solved by early-reionization models or a higher ionizing photon energy.

Our result suggests that reionization might have larger spatial variations at $\gtrsim 100$ cMpc scales compared to previous theoretical models. Recent observations of the “dark gaps” in high-redshift quasar spectra strengthen this argument, which are large neutral IGM structures at the end of EoR, when the Universe has been largely ionized [37, 38, 60].

The environment around FLaRE-z7

To investigate the environment around FLaRE-z7, especially whether it resides in a galaxy overdensity, we evaluate object number densities in the GOODS-S field and around the FLaRE-z7 sightline. To prevent the heterogeneous coverage of JWST observations in the JADES fields from biasing the measurements, we construct a depth-limited footprint as follows. We first compute the 5σ depth maps of GOODS-S by converting the ERR extensions of the mosaic using the following equation:

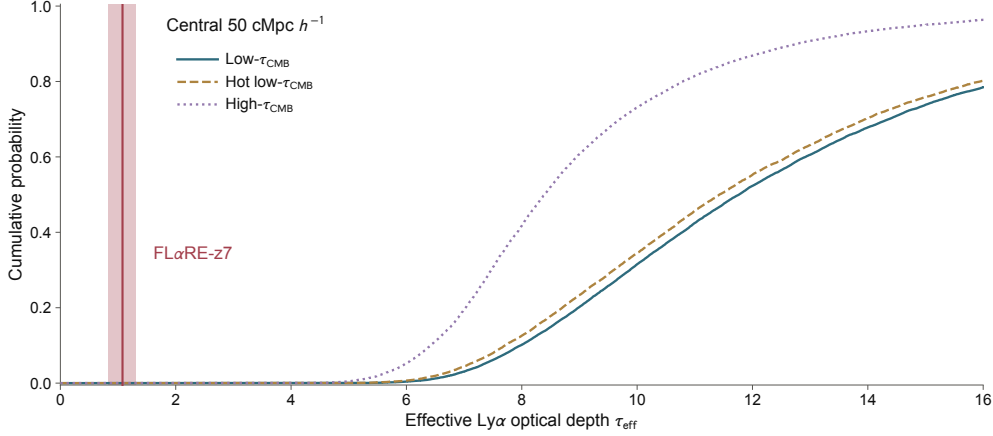


Fig. 13: The cumulative distribution of τ_{eff} from cosmological simulations. The models are from [36], corresponding to different reionization model parameters. The optical depths are measured for 50 cMpc/ h windows. None of the models produce a sightline with $\tau_{\text{eff}} < 4$, while FL α SH-z7 gives $\tau_{\text{eff}} = 1.08 \pm 0.24$. Therefore, FL α SH-z7 gives at least a marginal tension with current cosmological simulations and reionization models.

$$m^{\text{lim}} = 31.4 - 2.5 \log(5\sigma_{\text{aper}}) = 31.4 - 2.5 \log\left(5\sqrt{N_{\text{pix}}} \times \text{ERR}\right) \quad (4)$$

where 31.4 is the magnitude zeropoint when using nJy as flux units, and N_{pix} is the aperture for magnitude measurement. Here we use apertures with $r = 0.1''$, yielding $N_{\text{pix}} = \pi(r/0.03'')^2 = 34.9$. We neglect the noise correlation between pixels, as our primary goal is to select a depth-limited coverage footprint, rather than accurately compute the depth map. We then construct the depth-limited footprint by selecting pixels with the following magnitude limits: $m_{\text{F090W}}^{\text{lim}} > 28.887$, $m_{\text{F115W}}^{\text{lim}} > 28.902$, $m_{\text{F150W}}^{\text{lim}} > 28.790$, $m_{\text{F200W}}^{\text{lim}} > 29.011$, $m_{\text{F277W}}^{\text{lim}} > 30.169$, $m_{\text{F356W}}^{\text{lim}} > 30.200$, $m_{\text{F444W}}^{\text{lim}} > 29.941$. These depths are determined as the 1st percentile of all pixels that have $d < 2'$ away from the FL α RE-z7 sightline. This results in a footprint of 149 arcmin² in GOODS-S, which is shown as the gray area in Figure 5.

We then select objects with $|z_{\text{phot}} - z_c| < 0.2$ in the GOODS-S field, using the JADES DR5 photometric catalog. We only select objects that are brighter than the magnitude limits listed above, except for the F090W magnitudes, which is a dropout band for $z \gtrsim 7$ galaxies. This forms a flux-limited sample on the corresponding depth-limited footprint. We find 214 objects across the 149 arcmin² footprint, yielding an object number density of 1.44 ± 0.10 arcmin⁻² (assuming Poisson noise). There are 27 objects within 2' around FL α RE-z7, corresponding to 2.28 ± 0.44 arcmin⁻². In other words, we do not find significant overdensities around FL α RE-z7. We perform the measurement using different depths and aperture sizes around FL α RE-z7, and the conclusion remains unchanged.

We note several limitations for our analysis. First, we use photometric redshift in our analysis, as only a small subset of galaxies around FL α RE-z7's sightline have

875 spectroscopic redshifts. Second, the object sample we use to compute surface number
876 densities is flux-limited, and the abundance of faint galaxies ($M_{\text{UV}} \lesssim -17$) around
877 FLaRE-z7 is unknown. Those faint galaxies might be the primary contributor to
878 reionization (e.g., [61]), and we cannot rule out the scenario that FLaRE-z7 resides
879 in an overdensity of faint galaxies.

880

881 **Artificial Intelligence Disclosure**

882

883 We use AI to aid the coding in this study, including GPT-6 Astra [24], Claude Opus
884 5, Opus 5.5, Fable 5, and Fable 5.1 [23]. We also use these AI to improve the writing
885 of the manuscript.

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

Acknowledgements. This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST.

Some of the data products presented herein were retrieved from the Dawn JWST Archive (DJA). DJA is an initiative of the Cosmic Dawn Center (DAWN), which is funded by the Danish National Research Foundation under grant DNRF140.

The authors used generative AI to improve the writing of the manuscript. Specific models used include GPT-6 Astra [24], Claude Opus 5, Opus 5.5, Fable 5, and Fable 5.1 [23].

- Funding
M.Y. acknowledges support from the Steward Observatory, the University of Arizona through the Bart J. Bok Fellowship. Y.Z. acknowledges support from the NIRCам Science Team contract to the University of Arizona, NAS5-02105. Y.Z. is also supported by JWST Program #6434.
- Conflict of interest/Competing interests
The authors declare no competing interests.
- Ethics approval and consent to participate
- Consent for publication
- Data availability
The JADES DR5 dataset is available online at <https://slate.ucsc.edu/~brant/jades-dr5/>. The DJA is accessible at <https://dawn-cph.github.io/dja/>. Other data used in this work will be available online upon the acceptance of this paper.
- Materials availability
- Code availability
The code used in this work will be available online upon the acceptance of this paper.
- Author contribution
MY applied the deep learning model for the discovery of the object, performed measurements for the object, and came up with the astrophysical interpretation. YZ provided interpretations and modeling related to cosmic reionization in this work. JD provided the deep learning model FM-JADES-v1 and provided suggestions in the application of this model in astronomical discovery. All other authors provided suggestions in data analysis and manuscript preparation.

References

- [1] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat* **596**, 583–589 (2021).
- [2] Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nat* **624**, 80–85 (2023).
- [3] Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science*

967 **382**, 1416–1421 (2023).
968
969 [4] Degraeve, J. *et al.* Magnetic control of tokamak plasmas through deep reinforce-
970 ment learning. *Nat* **602**, 414–419 (2022).
971
972 [5] Bommasani, R. *et al.* On the Opportunities and Risks of Foundation Models.
973 *arXiv e-prints* arXiv:2108.07258 (2021).
974
975 [6] Hayat, M. A., Stein, G., Harrington, P., Lukić, Z. & Mustafa, M. Self-supervised
976 Representation Learning for Astronomical Images. *ApJ* **911**, L33 (2021).
977
978 [7] Parker, L. *et al.* AstroCLIP: a cross-modal foundation model for galaxies. *MNRAS*
979 **531**, 4990–5011 (2024).
980
981 [8] Parker, L. *et al.* AION-1: Omnimodal Foundation Model for Astronomical
982 Sciences. *arXiv e-prints* arXiv:2510.17960 (2025).
983
984 [9] Euclid Collaboration *et al.* Euclid Quick Data Release (Q1): XIII. Exploring
985 galaxy properties with a multi-modal foundation model. *A&A* **711**, A13 (2026).
986
987 [10] Fan, X. *et al.* Constraining the Evolution of the Ionizing Background and the
988 Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars. *AJ* **132**,
989 117–136 (2006).
990
991 [11] Yang, J. *et al.* Measurements of the $z \sim 6$ Intergalactic Medium Optical Depth
992 and Transmission Spikes Using a New $z \sim 6.3$ Quasar Sample. *ApJ* **904**, 26 (2020).
993
994 [12] Bosman, S. E. I. *et al.* Hydrogen reionization ends by $z = 5.3$: Lyman- α opti-
995 cal depth measured by the XQR-30 sample. *Monthly Notices of the Royal*
996 *Astronomical Society* **514**, 55–76 (2022).
997
998 [13] Jin, X. *et al.* (Nearly) Model-independent Constraints on the Neutral Hydrogen
999 Fraction in the Intergalactic Medium at $z \sim 5\text{--}7$ Using Dark Pixel Fractions in Ly α
1000 and Ly β Forests. *ApJ* **942**, 59 (2023).
1001
1002 [14] Āurovčíková, D. *et al.* Chronicling the Reionization History at $6 \lesssim z \lesssim 7$ with
1003 Emergent Quasar Damping Wings. *ApJ* **969**, 162 (2024).
1004
1005 [15] Ding, J., Yue, M., Zhu, Y., Fan, X. & Luo, Y. Learning JWST. I. A Founda-
1006 tion Model for New Population Discoveries and Morphology-Aware Photometric
1007 Redshift Measurements in the JADES Survey. *arXiv e-prints* arXiv:2609.11879
1008 (2026).
1009
1010 [16] Carreira, C. *et al.* JWST Advanced Deep Extragalactic Survey (JADES)
1011 Data Release 5: Catalogs of inferred morphological properties of galaxies
1012 from JWST/NIRCam imaging in GOODS-N and GOODS-S. *arXiv e-prints*
arXiv:2601.15957 (2026).

[17] Eisenstein, D. J. <i>et al.</i> Overview of the JWST Advanced Deep Extragalactic Survey (JADES). <i>ApJS</i> 283 , 6 (2026).	1013 1014 1015
[18] Johnson, B. D. <i>et al.</i> JWST Advanced Deep Extragalactic Survey (JADES) Data Release 5: NIRCам Imaging in GOODS-S and GOODS-N. <i>arXiv e-prints</i> arXiv:2601.15954 (2026).	1016 1017 1018 1019
[19] Robertson, B. E. <i>et al.</i> JWST Advanced Deep Extragalactic Survey (JADES) Data Release 5: Photometric Catalog. <i>arXiv e-prints</i> arXiv:2601.15956 (2026).	1020 1021 1022
[20] Vaswani, A. <i>et al.</i> Attention Is All You Need. <i>arXiv e-prints</i> arXiv:1706.03762 (2017).	1023 1024 1025
[21] Materials and methods are available as supplementary material.	1026
[22] McInnes, L., Healy, J., Saul, N. & Großberger, L. Umap: Uniform manifold approximation and projection. <i>Journal of Open Source Software</i> 3 , 861 (2018). URL https://doi.org/10.21105/joss.00861 .	1027 1028 1029 1030
[23] Anthropic. Claude Opus 5. https://claude.ai/ (2026).	1031 1032
[24] OpenAI. ChatGPT 5.6 Sol. https://chatgpt.com/ (2026).	1033 1034
[25] Matthee, J. <i>et al.</i> Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at $z \sim 5$ Revealed by the EIGER and FRESCO JWST Surveys. <i>ApJ</i> 963 , 129 (2024).	1035 1036 1037
[26] Rinaldi, P. <i>et al.</i> Not Just a Dot: The Complex UV Morphology and Underlying Properties of Little Red Dots. <i>ApJ</i> 992 , 71 (2025).	1038 1039 1040
[27] de Graaff, A. <i>et al.</i> Ionised gas kinematics and dynamical masses of $z \gtrsim 6$ galaxies from JADES/NIRSpec high-resolution spectroscopy. <i>A&A</i> 684 , A87 (2024).	1041 1042 1043
[28] Eilers, A.-C., Davies, F. B. & Hennawi, J. F. The Opacity of the Intergalactic Medium Measured along Quasar Sightlines at $z \sim 6$. <i>The Astrophysical Journal</i> 864 , 53 (2018).	1044 1045 1046 1047
[29] Jin, X. <i>et al.</i> Probing Direct Contributions of Galaxies and AGN to Cosmic Reionization in a Quasar Field J0226+0302 with JWST NIRCам and NIRSpec (2026).	1048 1049 1050 1051
[30] Johnson, B. D., Leja, J., Conroy, C. & Speagle, J. S. Stellar Population Inference with Prospector. <i>ApJS</i> 254 , 22 (2021).	1052 1053 1054
[31] Zhu, Y. <i>et al.</i> Galaxy Underdensities Host the Clearest Intergalactic Medium Ly α Transmission and Indicate Anisotropic Reionization. <i>ApJ</i> 1002 , 93 (2026).	1055 1056 1057 1058

1059 [32] Gunn, J. E. & Peterson, B. A. On the Density of Neutral Hydrogen in Intergalactic
1060 Space. *ApJ* **142**, 1633–1636 (1965).
1061
1062 [33] Willott, C. J. *et al.* In Search of the First Stars: An Ultra-compact and Very-
1063 low-metallicity Ly α Emitter Deep within the Epoch of Reionization. *ApJ* **988**,
1064 26 (2025).
1065
1066 [34] Witstok, J. *et al.* Witnessing the onset of reionization through Lyman- α emission
1067 at redshift 13. *Nat* **639**, 897–901 (2025).
1068
1069 [35] Eilers, A.-C. *et al.* Implications of $z \sim 6$ Quasar Proximity Zones for the Epoch
1070 of Reionization and Quasar Lifetimes. *ApJ* **840**, 24 (2017).
1071
1072 [36] Keating, L. C., Kulkarni, G., Haehnelt, M. G., Chardin, J. & Aubert, D. Con-
1073 straining the second half of reionization with the Ly β forest. *MNRAS* **497**,
1074 906–915 (2020).
1075
1076 [37] Davies, F. B., Becker, G. D. & Furlanetto, S. R. Determining the Nature of Late
1077 Gunn-Peterson Troughs with Galaxy Surveys. *ApJ* **860**, 155 (2018).
1078
1079 [38] Zhu, Y. *et al.* Long Dark Gaps in the Ly β Forest at $z \lesssim 6$: Evidence of Ultra-late
1080 Reionization from XQR-30 Spectra. *ApJ* **932**, 76 (2022).
1081
1082 [39] Meyer, R. A., Roberts-Borsani, G., Oesch, P. A. & Ellis, R. S. Probing patchy
1083 reionization with JWST: IGM opacity constraints from the Lyman α forest of
1084 galaxies in legacy extragalactic fields. *Monthly Notices of the Royal Astronomical
1085 Society* **542**, 1952–1968 (2025).
1086
1087 [40] Neyer, M. *et al.* The THESAN project: Connecting ionized bubble sizes to their
1088 local environments during the Epoch of Reionization. *Monthly Notices of the
1089 Royal Astronomical Society* **531**, 2943–2957 (2024).
1090
1091 [41] Lu, T.-Y. *et al.* The reionizing bubble size distribution around galaxies. *Monthly
1092 Notices of the Royal Astronomical Society* **528**, 4872–4890 (2024). URL <https://doi.org/10.1093/mnras/stae266>.
1093
1094 [42] Garaldi, E., Bellscheidt, V., Smith, A. & Kannan, R. The galaxy-IGM connec-
1095 tion in THESAN: the physics connecting the IGM Lyman- α opacity and galaxy
1096 density in the reionization epoch. *The Open Journal of Astrophysics* **8**, 116 (2025).
1097
1098 [43] Christenson, H. M. *et al.* The Relationship between IGM Ly α Opacity and Galaxy
1099 Density near the End of Reionization. *The Astrophysical Journal* **955**, 138 (2023).
1100
1101 [44] Zhu, Y. *et al.* Low Ly α Visibility in Galaxy Overdensities: Reionization
1102 Topology and Neutral-Fraction Ceilings from DIVER over $z > 4.8$ (2026).
1103
1104 [45] Schaerer, D. The transition from Population III to normal galaxies: Ly α and
He II emission and the ionising properties of high redshift starburst galaxies.

<i>A&A</i> 397 , 527–538 (2003).	1105
[46] Gangolli, N., D’Aloisio, A., Cain, C., Becker, G. D. & Christenson, H. On the correlation between Ly α forest opacity and galaxy density in late reionization models. <i>Journal of Cosmology and Astroparticle Physics</i> 2025 , 069 (2025).	1106
[47] Schwandt, T. P., Georgiev, I., Giri, S. K., Mellema, G. & Iliev, I. T. Impact of anisotropic photon emission from sources during the epoch of reionization. <i>MNRAS</i> 547 , stag129 (2026).	1107
[48] Oke, J. B. & Gunn, J. E. Secondary standard stars for absolute spectrophotometry. <i>ApJ</i> 266 , 713–717 (1983).	1108
[49] Brammer, G. Grizli: Grism redshift and line analysis software. Astrophysics Source Code Library, record ascl:1905.001 (2019). 1905.001 .	1109
[50] Brammer, G. grizli (2023).	1110
[51] Brammer, G. msaexp: NIRSpec analysis tools (2023).	1111
[52] Helton, J. M. <i>et al.</i> Discovery of the Lyman- α Forest in a Star-Forming Galaxy at $z \approx 7.57$: Implications for the Timing and Topology of Cosmic Reionization. <i>arXiv e-prints</i> arXiv:2609.38080 (2026).	1112
[53] Curtis-Lake, E. <i>et al.</i> JADES Data Release 4 – Paper I. Sample Selection, Observing Strategy and Redshifts of the Complete Spectroscopic Sample. <i>MNRAS</i> 549 , stag836 (2026).	1113
[54] Scholtz, J. <i>et al.</i> JADES Data Release 4 – Paper II. Data Reduction, Analysis, and Emission-Line Fluxes of the Complete Spectroscopic Sample. <i>MNRAS</i> 549 , stag939 (2026).	1114
[55] Zhu, Y. <i>et al.</i> PhotoIFU: NIRCcam as a Photometric Integral Field Unit for Mapping Feedback in Galaxies. <i>arXiv e-prints</i> arXiv:2607.23020 (2026).	1115
[56] Nakazato, Y. <i>et al.</i> Clump-Scale Dust Attenuation in Epoch of Reionization Galaxies: Spatially Resolved Properties from FirstLight Simulations (2026). 2602.07347 .	1116
[57] Kriek, M. & Conroy, C. The Dust Attenuation Law in Distant Galaxies: Evidence for Variation with Spectral Type. <i>ApJ</i> 775 , L16 (2013).	1117
[58] Lange, J. U. nautilus: boosting Bayesian importance nested sampling with deep learning. <i>Monthly Notices of the Royal Astronomical Society</i> 525 , 3181–3194 (2023). URL https://doi.org/10.1093/mnras/stad2441 .	1118
[59] Planck Collaboration <i>et al.</i> Planck 2018 results. VI. Cosmological parameters. <i>A&A</i> 641 , A6 (2020).	1119
	1120
	1121
	1122
	1123
	1124
	1125
	1126
	1127
	1128
	1129
	1130
	1131
	1132
	1133
	1134
	1135
	1136
	1137
	1138
	1139
	1140
	1141
	1142
	1143
	1144
	1145
	1146
	1147
	1148
	1149
	1150

1151 [60] Becker, G. D. *et al.* Evidence of patchy hydrogen reionization from an extreme
1152 Ly α trough below redshift six. *Monthly Notices of the Royal Astronomical Society*
1153 **447**, 3402 (2015).
1154
1155 [61] Atek, H. *et al.* Most of the photons that reionized the Universe came from dwarf
1156 galaxies. *Nat* **626**, 975–978 (2024).
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196