CLEAR

Survey Overview, Data Analysis, and Products

Simons, Raymond C.; Papovich, Casey; Momcheva, Ivelina G.; Brammer, Gabriel; Estrada-Carpenter, Vicente; Finkelstein, Steven L.; Gosmeyer, Catherine M.; Matharu, Jasleen; Trump, Jonathan R.; Backhaus, Bren E.; Cheng, Yingjie; Cleri, Nikko J.; Ferguson, Henry C.; Finlator, Kristian; Giavalisco, Mauro; Ji, Zhiyuan; Jung, Intae; Lotz, Jennifer M.; O’Brien, Rosalia; Skelton, Rosalind E.; Tilvi, Vithal; Weiner, Benjamin

Published in:
Astrophysical Journal

DOI:
10.3847/1538-4365/acc517

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
CLEAR: Survey Overview, Data Analysis, and Products

Raymond C. Simons1, Casey Papovich2,3,4, Ivelina G. Momcheva5, Gabriel Brammer6, Vicente Estrada-Carpenter7,8,9, Steven L. Finkelstein7,9, Catherine M. Gossmeyer6, Jasleen Matharu9, Jonathan R. Trump1,10, Bren E. Backhaus3,10, Yingjie Cheng10,11, Nikko J. Cler11,12,3, Henry C. Ferguson11,12,13, Kristian Finlator12,13,14, Mauro Giavalisco14,15, Intae Jung15,16, Jennifer M. Lotz15,17, Rosalind E. Skelton17,18, Vithal Tilvi16,18, and Benjamin Weiner16,18

1 Department of Physics, University of Connecticut, 196A Auditorium Road, Unit 3046, Storrs, CT 06269, USA; raymond.simons@uconn.edu
2 Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
3 George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
4 Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
5 Cosmic Dawn Centre, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
6 Department of Astronomy & Physics, Saint Mary’s University, 923 Robie Street, Halifax, NS B3H 3C3, Canada
7 Department of Astronomy, The University of Texas at Austin, Austin, TX 78759, USA
8 Kairo Aerospace, Sunnyvale, CA 94085, USA
9 Cosmic Dawn Center, Niels Bohr Institute, University of Copenhagen, Rådmandsgade 62, DK-2200 Copenhagen, Denmark
10 University of Massachusetts Amherst, Amherst, MA 01003, USA
11 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
12 Cosmic Dawn Center, Niels Bohr Institute, University of Copenhagen, Jagttej 128, DK-2200 Copenhagen N, Denmark
13 Department of Astronomy, New Mexico State University, Las Cruces, NM 88003, USA
14 Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA
15 Gemini Observatory/NSF’s National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tuscon, AZ 85719, USA
16 School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287-1404, USA
17 South African Astronomical Observatory, P.O. Box 9, Observatory, Cape Town 7935, South Africa
18 MMT/Steward Observatory, 933 N. Cherry Street, University of Arizona, Tucson, AZ 85721, USA

Received 2022 December 8; revised 2023 March 3; accepted 2023 March 14; published 2023 May 5

Abstract

We present an overview of the CANDELS Lyα Emission At Reionization (CLEAR) survey. CLEAR is a 130 orbit program of the Hubble Space Telescope using the Wide Field Camera 3 (WFC3) IR G102 grism. CLEAR targets 12 pointings divided between the GOODS-N and GOODS-S fields of the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS). Combined with existing spectroscopic data from other programs, the full CLEAR data set includes spectroscopic imaging of these fields over 0.8–1.7 μm. In this paper, we describe the CLEAR survey, the survey strategy, the data acquisition, reduction, processing, and science products and catalogs released alongside this paper. The catalogs include emission line fluxes and redshifts derived from the combination of the photometry and grism spectroscopy for 6048 galaxies, primarily ranging from 0.2 ≤ z ≤ 3. We also provide an overview of CLEAR’s science goals and results. In conjunction with this paper we provide links to electronic versions of the data products, including 1D+2D extracted spectra and emission line maps.

Unified Astronomy Thesaurus concepts: Emission line galaxies (459); Early-type galaxies (429); Galaxies (573); Galaxy evolution (594); High-redshift galaxies (734); Catalogs (205); Redshift surveys (1378)

1. Introduction

The spectroscopic capabilities of the Hubble Space Telescope (HST) provide a novel method to characterize and study the evolution of galaxies. Lying above the Earth’s atmosphere, HST is able to produce high-angular resolution images without the high sky backgrounds that plague ground-based observations. Slitless spectroscopy from HST therefore has two main advantages compared to terrestrial observations. First, it provides the spatial quality of HST (0.′′41 – 0.′′2 FWHM) with low backgrounds. Since the installation of the Wide Field Camera 3 (WFC3), we have seen a revolution in the slitless spectroscopy of distant galaxies. Primarily this has been provided by the grisms in the WFC3 IR camera, G102 and G141, which disperse light from 0.8 to 1.1 μm, and 1.1 to 1.7 μm, respectively, with low spectral resolution (R = δλ/λ ~ 200 and ~100, respectively). From initial work with the Early Release Science (ERS) programs (van Dokkum & Brammer 2010; Straughn et al. 2011), the community has carried out a series of programs including both targeted deep and wide-field surveys (e.g., FIGS, Pirzkal et al. 2017; 3D-HST, Momcheva et al. 2016; GLASS, Treu et al. 2015; AGHAST, Weiner 2012; MAMMOTH-Grim, Wang et al. 2022; 3D-DASH, Mowla et al. 2022; MUDF, Revalski et al. 2023), snapshot programs (e.g., WISPS, Atek et al. 2010), and targeted observations of transient sources (such as supernovae, e.g., Rodney et al. 2012).

Following in the legacy of these studies, we present here the data set from the CANDELS Lyα Emission At Reionization (CLEAR) survey. CLEAR is a HST Cycle 23 program that obtained deep (10–12 orbit depths) observations with the HST/WFC3 using the G102 grism in the IR camera. The observations (130 orbits total) cover 12 fields in the north and south Great Observatories Origins Deep Survey (GOODS-N and GOODS-S, respectively) fields overlapping the WFC3 imaging footprint of the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011). The primary goal of CLEAR was to characterize the evolution of the Lyα equivalent width...
distribution at $6 < z < 8$ and to interpret this in the context of reionization—as the intergalactic medium of the universe transitions from one that is mostly ionized at $z < 6$ to one that is mostly neutral at $z > 6$ (Robertson et al. 2013). This is important as $\text{Ly}_\alpha$ emission is sensitive to neutral HI fractions of $0.01–1.0$ (McQuinn et al. 2007), and there is a need to trace $\text{Ly}_\alpha$ from the ionized universe at $z = 6–6.5$ to the neutral universe at $z > 7$ with systematic, homogeneous surveys. In addition, the CLEAR pointings overlap with G102 and G141 observations from a number of previous programs (including the FIGS, AGHAST, and 3D-HST surveys). Together with CLEAR, this data set provides slitless spectroscopy at the spatial resolution of HST covering most of the $Y$, $J$, and $H$ bands, $0.8–1.7 \mu m$. This enables a wide range of science using strong emission lines and stellar continuum features in the rest-frame optical, that are redshifted into the near-infrared (NIR) and observable in the grism data. Furthermore, a major advantage of slitless spectroscopy is that it provides a spectrum for all galaxies in the field—target preselection is not required.

Here, we describe the CLEAR survey strategy, data acquisition, reduction, and science products. Along with this paper, we release the high-level 1D and 2D spectra, emission line maps, and redshift/line catalogs produced through this survey.

To date, the CLEAR data set has been used to study the evolution of the $\text{Ly}_\alpha$ equivalent width distribution into the epoch of reionization (Jung et al. 2022), galaxy stellar population properties including ages, star formation histories, and chemical enrichment histories (Estrada-Carpenter et al. 2019, 2020), emission line ratios, metallicities, and ionization properties in galaxies in both a spatially integrated (Backhaus et al. 2022; Papovich et al. 2022) and spatially resolved sense (Simons et al. 2021; Matharu et al. 2022; Backhaus et al. 2023), supermassive black holes (Yang et al. 2021), $\text{P}_{\alpha}/\beta$ as a star formation indicator (Cleri et al. 2022a), high-ionization [Ne V] emission in galaxies (Cleri et al. 2022b, 2023), and the mass–metallicity relation (Henry et al. 2021; Papovich et al. 2022). These studies demonstrate that the CLEAR data products provide a resource for identifying and characterizing the properties of galaxies over a wide range of redshift, including the peak of the cosmic star formation density (Madau & Dickinson 2014) and supermassive black hole accretion density (Brandt & Alexander 2015).

The outline for this paper is as follows. In Section 2 we describe the design of the survey, and provide the details of the CLEAR observing program. In Section 3 we describe the ancillary HST grism data sets that we include in our analysis of the CLEAR data set. In Section 4 we describe the multi-wavelength photometric catalog we employ for analysis of the CLEAR galaxies. In Section 5 we describe the process for data reduction, calibration, spectral extractions, and derived quantities including redshifts and emission line fluxes from the grism spectroscopy. In Section 6, we discuss the catalogs and data products released alongside this paper. In Section 7 we discuss the CLEAR science, and provide additional examples of using the data for science. Finally, in Section 8 we provide a brief summary. Throughout this paper, we use magnitudes on the Absolute Bolometric system (Oke & Gunn 1983) and a cosmology that assumes $\Omega_m,0 = 0.3$, $\Omega_{\Lambda},0 = 0.7$, and $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$. We use a Chabrier-like initial mass function (IMF) for any quantities such as stellar mass and star formation rate (SFR).

### 2. Survey Design and Data Acquisition

The CLEAR program was designed in area and depth to survey a sufficient number of high-redshift galaxies to the line flux sensitivities needed to achieve the primary science goals of the survey—constraints on the Ly$\alpha$ line emission in $6 < z < 8$ galaxies to limits of $\approx 10^{-12}$ erg s$^{-1}$ cm$^{-2}$. We targeted 12 new fields with WFC3, evenly divided between the GOODS-N and GOODS-S galaxy fields. Figures 1 and 2 show the locations of the CLEAR pointings.

#### 2.1. Target Field Selection

The primary goal of the CLEAR survey was to constrain the amount of Ly$\alpha$ emission from galaxies in the epoch of reionization. To that end, we selected fields in GOODS-N and -S which maximized the number of photometrically selected target galaxies over the redshift range $6 < z < 8$.

To select the fields for CLEAR, we used the Lyman-break (LBG) catalog of Finkelstein et al. (2015). This provided $>6$ potential pointings in GOODS-N and GOODS-S each. We then downscaled to six in each field. The CLEAR fields are illustrated in Figures 1 and 2. They are labeled “GN1–GN5, GN7” in GOODS-N (where they are nonsequential as we dropped a GN6 field) and “GS1–GS5” in GOODS-S. GS1 overlaps with the Hubble Ultra Deep Field/Advanced Camera for Surveys (HDF/ACS) parallel field (Beckwith et al. 2006) and the sixth field in GOODS-S coincides mostly with the WFC3/ERS field (Straughn et al. 2011), which we designate “ERSPRIME.” The coordinates of the fields, including the number of new orbits provided by CLEAR, are given in Table 1. The field area of CLEAR is significantly larger than the typical spatial extent of ionized structures during the epoch of reionization (e.g., Ocvirk et al. 2020). Moreover, cosmic variance is not an issue for CLEAR as the GOODS-N and -S fields are sufficiently separated on the sky, and the redshift range $6 < z < 8.2$ over which the G102 wavelength coverage is sensitive to redshifted Ly$\alpha$ provides sufficient volume for galaxy populations to be unrelated in redshift.

#### 2.2. Considerations for the HST Observations

We split each orbit of the HST/WFC3 observations into a direct image (F105W) and two G102 grism exposures of the same pointing. Each WFC3 exposure used the MULTIACCUM mode, with the sample sequencing (SAMP-SEQ) and number of samples (NSAMP) depending on the type of observation. Each WFC3/F105W direct image comprises a single iteration (exposure) with SPARS25 and NSAMP = 11. This produced 303 s observations. The G102 exposures used a single iteration with SPARS100 and either NSAMP = 12 or 13 samples—depending on the amount of usable time per orbit. This provided a total exposure time of 1103 or 1203 s per exposure. In all cases, we adopted the dither pattern employed by 3D-HST (Momcheva et al. 2016) to match the sampling of those data as closely as possible.

We observed each pointing in CLEAR using two orbits at a single position angle (ORIENT), repeating the pattern above. We required additional orbits to have a position angle offset by at least $20^\circ$. That requirement ensures that the spectral trace from each object falls on different portions of the detector and
that contamination from nearby sources occurs in only a single position angle (see, e.g., the discussion in Estrada-Carpenter et al. 2019). Table 1 lists the ORIENTs and number of orbits per pointing.

In addition, WFC3 Y-band exposures are known to suffer time-variable backgrounds during the HST orbit (Lotz et al. 2017). The origin of this background is due to He I 10830 Å emission from the Earth’s atmosphere when HST observes at low limb angles. This background is strongest when HST is not in the Earth’s shadow, which occurs at the start or end of each orbit. Following Lotz et al. (2017) we predicted the HST ephemeris for each of our orbits and scheduled the sequence of F105W direct images and two G102 grism exposures so that the latter were taken when HST was in the shadow of the Earth. In doing so, the grism observations were protected from the He I background. As a tradeoff, the F105W imaging suffers from higher backgrounds. This was acceptable as those images are used only for alignment while the grism spectroscopy is required for the primary science. Table 1 lists the observing sequence of F105W and G102 during the observation where either the direct image occurs first in the orbit (F105W, G102, G102) or last in the orbit (G102, G102, F105W).

3. Ancillary Observations

3.1. Imaging Data

The CLEAR pointings lie in the well-studied GOODS-S and GOODS-N galaxy fields. These fields have extensive UV to IR imaging. We refer the reader to Table 3 of Skelton et al. (2014) for full details, and briefly describe the relevant imaging data sets here.

HST/ACS + WFC3 imaging is available in 7 and 10 bands in GOODS-N and GOODS-S, respectively. The majority of this HST imaging is provided by three large programs: GOODS (Giavalisco et al. 2004), the CANDELS Multi-Cycle Treasury Project (Grogin et al. 2011; Koekemoer et al. 2011), and the 3D-HST Treasury Program (Brammer et al. 2012; Skelton et al. 2014; Momcheva et al. 2016).

In addition, UV to 8 μm imaging is available from a number of ground- and space-based observatories: KPNO 4 m/Mosaic (U; Capak et al. 2004), Very Large Telescope (VLT)/VIMOS (U–R; Nonino et al. 2009), WFI 2.2 m (U38BVRCI; Erben et al. 2005; Hildebrandt et al. 2006), Keck/LRIS (G–R; Steidel et al. 2003), Subaru/Suprime-Cam (BVRizci and 14 medium bands; Capak et al. 2004; Cardamone et al. 2010), Subaru/MOIIRCS (JHKs; Kajisawa et al. 2011), VLT/ISAAC (JHKs; Wuyts et al. 2008;
3.2. Grism Data

To supplement the CLEAR G102 grism spectroscopy, we queried the Barbara A. Mikulski Archive for Space Telescopes (MAST) for G102 (0.8–1.1 µm) and G141 (1.1–1.7 µm) observations that overlap the CLEAR footprint. We retrieved a total of 52 orbits of G102 and 76 orbits of G141 observations—taken through the programs listed in Table 2.

Of note, CLEAR ER includes ultradeep 40 orbit G102 spectra in the HUDF (the “GS4” pointing of CLEAR) taken as a part of the FIGS program (Pirzkal et al. 2017).

Combined, the G102 and G141 grisms cover a continuous wavelength range of 0.8–1.7 µm. The visibility windows of bright rest UV–NIR lines are shown for both grisms in Figure 3. With joint grism coverage, we are able to capture a more complete set of emission lines for the same galaxy. As an example, with both grisms employed, the full R23 complex (Hβ, [O III], [O II]) is visible in galaxies over the redshift range of 1.2 < z < 2.4. With only one of the grisms, this range is considerably smaller—1.2 < z < 1.3 for the G102 grism alone and 2.0 < z < 2.4 for the G141 grism alone.

4. Updated 3D-HST Photometric Catalogs

As a part of the 3D-HST survey, Skelton et al. (2014) carried out source detection and photometric analysis on the full set of imaging described in Section 3.1. The resulting photometric catalogs are available on the 3D-HST website19 ("v4.1" as of this publication). These are the root catalogs used for the CLEAR ER data set. As described above and in Table 1, we supplement this catalog with HST/WFC3 F105W photometry for the sources in the CLEAR footprint. The F105W fluxes and uncertainties are measured in a manner that is consistent with Skelton et al.

19 https://archive.stsci.edu/prepds/3d-hst/
(2014). We also incorporate new ground-based spectroscopic redshifts ("z_spec") from the KMOS-3D (Wisnioski et al. 2019) and MOSDEF surveys (Kriek et al. 2015) in GOODS-S and GOODS-N. The original compilation of spectroscopic redshifts in the 3D-HST catalog derives from the MOIRCS Deep Survey catalog in GOODS-N (Kajisawa et al. 2011) and the FIREWORKS catalog in GOODS-S (Wuyts et al. 2008)—see Skelton et al. (2014) for details. We supplant these redshifts with those from the KMOS-3D (quality flag = 1 in their catalog; N = 43) and MOSDEF (quality flag \(\geq 3\) in their catalog; \(N = 143\)) surveys, when the latter two are available.

With these updates to the catalog, we use the eazy-py code (a Python photometric analysis and redshift tool based on EAZY; Brammer et al. 2008) to derive new zero-point corrections, photometric redshifts, and rest-frame colors for the full 3D-HST sample in GOODS-S and GOODS-N. We also use eazy-py to derive new broadband-based estimates of stellar masses, SFRs, and dust attenuation \(A_V\).

We adopt the set of "fsps_QSF_12_v3" Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) continuum templates available in the eazy-Py library. The FSPS templates assume a Chabrier (2003) IMF and were constructed to span a range of galaxy types (following the methodology of Blanton & Roweis 2007; Brammer et al. 2008). The updated version of the 3D-HST photometric catalog ("v4.6"\(^21\)) is released alongside this paper. The full eazy-Py code is available at https://github.com/gbrammer/eazy-py.

\(^21\) The public photometric catalog jumps from version v4.1 to v4.6, skipping over intermediate internal CLEAR team releases.
The columns of the catalog are described in Table 10 of Skelton et al. (2014), with two new columns of F105W flux and flux uncertainties provided by CLEAR. In addition to the photometric catalog, we also release a catalog of eazy-Py-derived galaxy properties. The contents of this catalog are described in Table 3.

5. Data Reduction and Processing

We process the complete data set of grism and imaging observations described in Sections 2, 3 and Tables 1 and 2 using the grism redshift and line analysis software Grizli22 (Brammer 2019). As described below, Grizli performs end-to-end processing of HST imaging and slitless spectroscopy data sets. This includes retrieving and preprocessing the raw observations, performing astrometric alignment, modeling contamination from overlapping spectra, extracting the 1D and 2D spectra of individual sources, fitting continuum + emission line models, and generating emission line maps.

5.1. Preprocessing

We use Grizli to retrieve the observations described in Tables 1 and 2 from the MAST archive. Then, the raw observations are reprocessed with the calwf3 pipeline and corrections for variable sky backgrounds (Brammer 2016) are applied. Cosmic rays and hot pixels are identified with the AstroDrizzle software (Gonzaga et al. 2012). Flat-field corrections are applied to the G102 (G141) grism exposures using the F105W (F140W) calibration images. We use the “Master Sky” constructed in Brammer et al. (2015) to carry out sky subtraction. Using the deeper 3D-HST HST/WFC3 F140W galaxy catalog of these fields (Skelton et al. 2014) as reference, a relative astrometric correction is applied to the data.

5.2. Full-field Contamination Models

For each pointing, a contamination model is created to account for spectral overlap of adjacent sources on the WFC3 detector. The contamination model is generated from an iterative forward model of the full-field HST Y-band mosaic. A first pass model is constructed for all objects in the Y-band mosaic brighter than mF105W = 25. For each object, a spectrum is constructed that is flat in fλ flux density and normalized to the F105W flux of the source. A second pass “refined” continuum model is then created for objects brighter than mF105W = 24. These objects are assigned spectra by fitting second-order polynomials to the spectrum of each source after subtracting the first pass models of suspected contaminating sources. This process is repeated for each visit. The end-to-end reduction and

---

Table 2
Ancillary WFC3 Grism Observations Overlapping the CLEAR Pointings

<table>
<thead>
<tr>
<th>Field</th>
<th>Number of Orbits</th>
<th>HST Cycle</th>
<th>Proposal ID</th>
<th>Principal Investigator</th>
<th>Survey or Pointing Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOODS-N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G102</td>
<td>21</td>
<td>17</td>
<td>11600</td>
<td>Weiner</td>
<td>AGHAST</td>
</tr>
<tr>
<td>G141</td>
<td>3</td>
<td>19</td>
<td>12461</td>
<td>Riess</td>
<td>SN COLFAX</td>
</tr>
<tr>
<td>G102</td>
<td>9</td>
<td>21</td>
<td>13420</td>
<td>Barro</td>
<td></td>
</tr>
<tr>
<td>G141</td>
<td>25</td>
<td>25</td>
<td>13871</td>
<td>Oesch</td>
<td></td>
</tr>
<tr>
<td>GOODS-S</td>
<td>1</td>
<td>17</td>
<td>11359</td>
<td>O’Connell</td>
<td>WFC3/ERS</td>
</tr>
<tr>
<td>G102</td>
<td>13</td>
<td>18</td>
<td>12099</td>
<td>Riess</td>
<td>GEORGE, PRIMO</td>
</tr>
<tr>
<td>G141</td>
<td>29</td>
<td>18</td>
<td>12177</td>
<td>van Dokkum</td>
<td>3D-HST</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The number of orbits listed is of the subset of pointings overlapping the CLEAR field, and does not reflect the total number of orbits of the respective programs.

---

parameter file that is used in the run is also provided in the release. The columns of the catalog are described in Table 10 of Skelton et al. (2014), with two new columns of F105W flux and flux uncertainties provided by CLEAR. In addition to the photometric catalog, we also release a catalog of eazy-Py-derived galaxy properties. The contents of this catalog are described in Table 3.

Figure 3. The redshift ranges over which different emission lines are observable with the HST/WFC3 G102 and G141 grisms are shown. Each row indicates the redshifts of the G102 (blue) and G141 (red) grism spectroscopic coverage. The emission lines are labeled along the ordinate.

22 https://github.com/gbrammer/grizli
continuum modeling of a single G102 grism exposure is shown in Figure 4. While the continuum model generally performs remarkably well for the majority of the sources and detector area (see the residual image in the lower right panel of Figure 4), for pointlike sources there can be a residual signal due to the imperfect point-spread function (PSF) reconstruction in the blotting procedure. To make the grism model, we must blot the more finely sampled drizzled reference image to the coarser pixels of the detector — where the PSF is undersampled. That transformation is not perfect/lossless, and will not preserve the exact pixel phase sampling. This is most apparent in the residual continuum of bright PSF-sized sources (e.g., stars and active galactic nuclei (AGN)).

We measure the fraction of the extracted source spectra (see the next subsections) that are contaminated as a function of the contamination level \( F_{\lambda,\text{contamination}}/F_{\lambda,\text{source}} \). We find that \( \sim25\% \) (\( \sim65\% \)) of the spectra are contaminated at a level of \( F_{\lambda,\text{contamination}}/F_{\lambda,\text{source}} \geq 1 \) and 0.1, respectively. In all cases, this continuum contamination is modeled and subtracted.

### 5.3. Extraction of Spectra

We use Grizli to extract the 2D grism spectra of all objects brighter than \( m_{F105W} = 25 \). Each 2D spectrum is known as a “beam.” One beam is extracted for each grism observational visit of each object. Therefore, each object normally has multiple beams—one for each position angle of each grism instrument. The beam files carry along the local contamination model relevant for the 2D spectrum and a full description of the WFC3 detector. In total, 6048 objects have at least one 2D spectrum extracted from the CLEARIR data set. Of these, 4707 were observed with both grisms, 533 were observed with only the G102 grism, and 808 were observed with only the G141 grism.

The grism exposure times of the extracted objects are shown in Figure 5 and range from 0.5 to 28 hr in G102 and 0.5 to 12 hr in G141. There are several distinct peaks in the distribution of exposure times, which correspond to different programs in the CLEARIR observational set. The notable peaks indicated in Figure 5 are associated with programs of depths of \( \sim2 \) orbits (from Barro/G102, AGHAST/G141, 3D-HST/G141), 12 orbits (CLEAR/G102), and 40 orbits (FIGS/G102).

### 5.4. Redshifts

Redshift and emission line fits are carried out in Grizli using both the grism spectra and the available multiwavelength photometry. The spectra are scaled to the photometry using a simple wavelength-independent scaling factor. The continuum is modeled using a basis set of template FSPS models (FSPS; Conroy et al. 2009; Conroy & Gunn 2010). The FSPS templates reflect a range of galaxy types and star formation...
The distribution of HST/WFC3 G102 and G141 exposure times for all objects extracted (m$_{105W} < 25$) as a part of the CLEAR$^{ER}$ data set. The peaks in the distribution correspond to observing programs of ~2 orbit (3D-HST, BARRO, AGHAST), 12 orbit (CLEAR), and 40 orbit (FIGS) depths, as indicated.

![Figure 5](https://archive.stsci.edu/hlsp/clear/)

Figure 5. The distribution of HST/WFC3 G102 and G141 exposure times for all objects extracted (m$_{105W} < 25$) as a part of the CLEAR$^{ER}$ data set. The peaks in the distribution correspond to observing programs of ~2 orbit (3D-HST, BARRO, AGHAST), 12 orbit (CLEAR), and 40 orbit (FIGS) depths, as indicated.

5.5. Emission Line Fluxes and Maps

Emission line fluxes are measured at the best-fit redshift using the basis FSPS templates including emission lines, and following the forward-modeling technique described above. However, now the emission lines and complexes are considered as separate components without fixing their line ratios. The [S II] $\lambda\lambda6718$ and 6732 and [O III] $\lambda\lambda4960$ and 5007 doublets are fit as single components with line ratios that are fixed at 1:1 and 1:2.98, respectively (Osterbrock & Ferland 2006). The [S II] ratio is appropriate for interstellar medium (ISM) electron densities of $\sim10^{-10} \text{ cm}^{-3}$ (Kewley et al. 2019). The H$\alpha$+[N II] complex is blended at the resolution of the G141 and G102 grisms. We therefore fit these lines with a single component at the wavelength of H$\alpha$.

The 2D+1D spectra (G102 and G141) of a single galaxy in the CLEAR$^{ER}$ data set is shown in Figure 7, along with its full FSPS+emission lines fit.

Emission line maps are created by drizzling the continuum- and contamination-subtracted 2D spectral beams to the wave-length of the redshifted line center. This is carried out using the astrometry of the spectral trace. The line maps have a pixel scale of 0.61. The uncertainties on the line maps are computed using the weights of the constituent pixels in the drizzling procedure. Emission line maps are generated automatically for H$\alpha$+[N II], [O III] $\lambda\lambda4960$, 5008, H$\beta$, and [O II] $\lambda\lambda3727, 3730$. They are created for the remaining lines and line complexes listed in Figure 3 if they are detected with an S/N greater than 4 in the 1D spectrum. Example line maps created from the CLEAR data set can be found in Simons et al. (2021), Matharu et al. (2022), and Backhaus et al. (2022, 2023).

6. Data Products and Catalogs

This section provides a description and validation of the science products and redshift+line flux catalogs that are produced from the CLEAR survey. The products described here are released alongside this paper at https://archive.stsci.edu/hlsp/clear/. An interactive map and the “biographical” information for each galaxy in our sample is available at https://clear.physics.tamu.edu.
6.1. Data Products

As described in Section 5, we use the Grizli grism analysis software to extract spectra and emission line maps for 6048 sources. Each source is associated with a set of four Grizli products, following the naming scheme `FIELD_<ID>_<PRODUCT>.fits`. `FIELD` is the CLEAR field name (e.g., “GN1”; see Table 1), `<ID>` is the identification number from the 3D-HST catalog (Skelton et al. 2014), and `<PRODUCT>` is the product type. The product types are “full,” “beams,” “stack,” and “1D.” These are multiextension fits (MEF) files, and are described here:

1. The “full.fits” product stores the (i) results of the FSPS + emission line model including the redshift likelihood information (`ZFIT_STACK`), best-fit line fluxes and equivalent widths and covariance elements (`COVAR`), (ii) the model templates (`TEMPL`), (iii) the source segmentation map (extension SED), (iv) the F105W or F141W direct image (extension DSCI) and associated weight map (extension DWHT), and (iv) a map of the PSF in each WFC3 direct imaging bandpass (extension DPSF). In addition, emission line maps are included for those galaxies studied in Simons et al. (2021), Matharu et al. (2022), and Backhaus et al. (2023). As described above, the emission line maps are generated for the following line/line complexes if they are in the observed wavelength window of the object: Hα+[N II], [O III], [S II], [S III], Mg II, or Lyα. The following figure shows the redshift distribution of sources detected with S/N \(\geq 3\) in a single emission line (as labeled). The G102+G141 are sensitive to emission from these lines over different ranges in redshift.

![Redshift distribution of galaxies with emission lines measured by CLEAR](image)

**Figure 6.** Redshift distribution of galaxies with emission lines measured by CLEAR (in the combined G102+G141 data set). The top panel shows the distribution of all sources with at least one emission line detected with S/N \(\geq 3\) in one of Hα, Hβ, [O II], [O III], [S II], [S III], Mg II, or Lyα. The lower panels show the redshift distribution of sources detected with S/N \(\geq 3\) in a single emission line (as labeled). The G102+G141 are sensitive to emission from these lines over different ranges in redshift.
produced for the remaining lines/line complexes listed in Figure 3 only if they are detected in the 1D spectrum with an S/N greater than four. The maps are 160 pixels × 160 pixels, which correspond to 16″ × 16″ at our pixel scale of 0″1 × 0″1. For each emission line that is fit, the MEF contains extensions for the emission line map (LINE), an associated weight map (WHT), continuum map (CONTINUUM), and contamination map (CONTAM).

2. The "_beams.fits" product stores the full set of G102 and/or G141 grism 2D spectra along with postage stamps of the associated direct reference images. For a G102 spectrum, the corresponding direct image is from the
WFC3/F105W filter. For a G141 spectrum, the direct image is from the WFC3/F140W filter. As defined above, an individual 2D spectrum is referred to here as a beam. This product serves as the main input to Grizli’s spectral fitting and emission line map-making tools. The MEF extensions in this product have the same definitions as for those in the "stack.fits" products, but the "beams.fits" contain information for each individual beam.

3. The "stack.fits" product stores a stacked 2D spectrum of the beams, including the science extension (SCI), a weight extension (WHT), a contamination model extension (CONTAM), a best-fit continuum model extension (MODEL), and an estimate of the PSF (KERNEL).

4. The "1D.fits" product stores the optimally extracted 1D grism spectrum of the source. There is one MEF extension for each of the G102 and G141 spectra. Each of the fits extensions of this product includes the columns of the wavelength (“wave”), (unnormalized) flux density (“flux”), flux density error (“err”), number of grism spectral pixels per wavelength bin (“npx”), flat (“flat,” used for normalization of the flux), contamination model (“contam”), and a decomposition of the spectrum into its line (“line”) and continuum (“cont”) components. To convert the unnormalized spectrum to flux density (in units of erg s\(^{-1}\) cm\(^{-2}\) Å\(^{-1}\)) one divides the “flux” column by the “flat” column.

6.2. Line Fluxes and Redshifts

Here we describe the line flux and redshift catalogs that are released alongside this paper. We also carry out a relative validation of the redshifts and line fluxes by comparing them against a compilation of high spectral resolution redshifts from ground-based spectroscopic surveys and previous grism-based measurements from the 3D-HST team (Momcheva et al. 2016).

6.2.1. Catalogs

The redshifts and line fluxes that are measured from the CLEAR\(^\text{ER}\) data set are released in two spectroscopic catalogs: one for GOODS-S (“GDS_v4.1_CLEAR.fits”) and one for GOODS-N (“GDS_v4.1_CLEAR.fits”). The catalog version released alongside this paper is v4.1. The columns of these catalogs are listed in Table 4.

The catalogs include the proper properties of the galaxies and the grism observations: the source ID (identical to those of 3D-HST; Skelton et al. 2014), the J2000 International Celestial Reference System (ICRS) R.A. and decl., the number of emission lines/line complexes observed by the grisms, the source G102 and G141 exposure time, and the diagnostics of the template fit, including the minimum \(\chi^2\) and the Bayesian information criterion BIC\_TEMP.

The catalogs also include the redshift and emission line measurements from Grizli: the confidence intervals of the redshift probability distribution, the maximum likelihood\(^23\) and minimum “risk” redshifts (Tanaka et al. 2018), the line flux and line flux uncertainties for the full suite of lines listed in Figure 3, and the confidence intervals of the rest-frame equivalent widths of these lines.

\(^{23}\)Grizli adopts the maximum likelihood redshift to create emission line maps from the grism spectra.
of the redshifts will be similar to those of the “photometric redshifts.” On the other hand, for galaxies with emission lines detected in the grism spectra, the uncertainty of the derived redshifts is generally much smaller. With an emission line detection, the redshift precision will only be limited by the ability of the grism data to centroid the emission feature(s)—which is set by the spectral resolution and pixel sampling.

**CLEARER versus ground-based redshifts:** in the top panel of Figure 8, we compare the ground-based spectroscopic redshifts against those measurements from the joint CLEARER grism + photometry data set. The ground-based redshifts \( z_{\text{spec, ground}} \) are sourced from three matched catalogs: the compilation provided in the original 3D-HST catalog in GOODS-N and GOODS-S (Skelton et al. 2014; \( N = 1206 \)), the KMOS-3D survey in GOODS-N (Wisnioski et al. 2019; \( N = 43 \)), and the MOSDEF survey in GOODS-N and GOODS-S (Kriek et al. 2015; \( N = 143 \)). For the KMOS-3D and MOSDEF surveys, we include redshifts with quality flags of \( \geq 1 \) (described as a secure redshift) and \( \geq 3 \) (described as a redshift-based emission line(s) detected at an \( S/N \geq 2 \) or better), respectively.

We only consider sources with at least one secure line detected in the CLEARER data set \( (S/N > 3) \). In both fields, we find excellent systematic agreement between the two redshifts, with a distribution of differences that is centered on \( \Delta z/(1+z) \sim 0 \) (shown in the subpanel). The median of \( \Delta z/(1+z) \) is \( 0.0002 \pm 0.0001 \) and \( 0.0003 \pm 0.0002 \) in GOODS-N and GOODS-S, respectively—i.e., the median is statistically consistent with 0.

Given the high spectral resolution of the ground-based data \( (\times 10 \) higher than the WFC3 grisms), the width of this distribution is effectively an exclusive measure of the redshift precision of the grism data. To quantify the level of that precision, we measure the normalized median absolute deviation (\( \sigma_{\text{NMAD}} \)) of the redshift differences (following Brammer et al. 2008; Momcheva et al. 2016):

\[
\sigma_{\text{NMAD}} = 1.48 \times \text{median}|x - \text{median}(x)|, \quad (1)
\]

where \( x = \Delta z/(1+z_{\text{spec, grism}}) \). The quantity \( \sigma_{\text{NMAD}} \) is the median absolute deviation multiplied by a factor of 1.48. For a normal distribution, \( \sigma_{\text{NMAD}} \) is equal to the standard deviation. However, it is less impacted by outliers than the standard deviation.

The \( \sigma_{\text{NMAD}} \) of the differences between the ground-based redshifts \( z_{\text{spec, ground}} \) and the grism redshifts \( z_{\text{spec, grism}} \) is \( \sim 0.0014 \) for both fields. The implied FWHM of the redshift accuracy is \( 0.0033 \times (1+z) \). This corresponds to 33 and 46 \( \AA \) at 1.02 \( \mu m \) and 1.41 \( \mu m \), respectively—the characteristic wavelengths of the WFC3 grisms. These are roughly equal to the spectral size of one WFC3 pixel for both the G102 grism (24.5 \( \AA \)) and the G141 grism (46 \( \AA \)). To conclude, the FWHM redshift accuracy of the WFC3 grisms is roughly equal to their spectral sampling size. This is generally consistent with that found for the G141 grism in the 3D-HST survey (Momcheva et al. 2016).

Next, we measure the outlier fraction of the distribution of differences. We define an outlier as a galaxy with a difference in the redshift measurements that is larger than \( 5 \times \sigma_{\text{NMAD}} \). We measure an outlier fraction of \( \sim 13\% \) in both fields. Roughly half of the outliers have a discrepancy that is statistically consistent (i.e., within \( 5 \times \sigma_{\text{NMAD}} \)) with that expected from simple line confusion between the surveys—e.g., a redshift based on a line identified as H\( \alpha \) in one data set and [O \( \text{III} \)] in another. For the remaining outliers (24 of 378 total, or 6\% of the full sample), the \( >5\sigma \) discrepancies are unexplained. A potential explanation is that the grism spectra of these galaxies could be contaminated by one or more emission lines from another source—which could lead to a spurious spectroscopic redshift based on those impostor lines. Such line contamination is not accounted for in the Grizli models. Of the 24 unexplained outliers, 13 have redshifts constrained by multiple secure line detections (5\( \sigma \)) in the grism spectra. These objects (13 of 378, or 3\% of the full sample) cannot be easily explained by grism line contamination.

**CLEARER versus photometric redshifts:** in the bottom panel of Figure 8, we compare the CLEARER redshifts with those estimated using only the photometry \( (z_{\text{phot}}) \). As above, we find excellent systematic agreement between the two—in both fields the medians of the distributions of redshift differences are statistically consistent with 0. The \( \sigma_{\text{NMAD}} \) of the differences is \( 0.0320 \) and 0.0189 for GOODS-N and GOODS-S, respectively. Those are a factor of \( \sim 10-20 \) larger than that measured above between the grism and ground-based spectroscopic redshifts.
We conclude that the redshifts based on photometry alone are $\sim 10^{-20}$ less precise than those based on the grism + photometry.

**CLEAR** versus 3D-HST redshifts and line fluxes: in Figure 9, we compare the CLEAR redshifts and emission line fluxes with those measured by the 3D-HST survey for the same galaxies (Brammer et al. 2014; Skelton et al. 2014; Momcheva et al. 2016). The 3D-HST G141 observations are also included in the CLEAR data set, and so this comparison is effectively a test of the differences between the 3D-HST and Grizli reduction pipelines.

In the left panel of Figure 9, we compare redshifts for two types of sources: (1) those with a secure line detected in both surveys (black) and (2) those with no line detected in either survey (gray). The distribution of differences is shown in the subpanel. The systematic agreement between the two survey measurements is excellent, peaking at $\Delta z/(1+z) \sim 0$ (with a median that is statistically consistent with 0). As expected, we
find that the distribution of differences for those galaxies with a secure line detection is much more narrow ($\sigma_{\text{NMAD}} \sim 0.002$) than for those without one ($\sigma_{\text{NMAD}} \sim 0.036$).

In the right panel of Figure 9, we compare the measured fluxes for the bright Balmer and oxygen rest-frame optical emission lines that are securely detected ($S/N > 3$) in both surveys. In the subpanel, we show the distribution of the differences normalized by the uncertainty of the differences ($\Delta F / \sigma_{\text{fl}}$). The quantity $\sigma_{\text{fl}}$ is calculated as $(\sigma^2_{\text{flux}} + \sigma^2_{\text{phot}})^{1/2}$. If the quoted uncertainties of the individual measurements reflect the true uncertainty and there is no systematic offset between the measures, the quantity $\Delta F / \sigma_{\text{fl}}$ should be distributed as a standard normal ($\sigma_{\text{NMAD}} \sim 1$, centered on 0). For the full sample of line detections shown in Figure 9, we find that the peak of the distribution of $\Delta F / \sigma_{\text{fl}}$ is $\sim 0$, indicating systematic agreement, but that $\sigma_{\text{NMAD}} \sim 1.4$. At face value, the fact that $\sigma_{\text{NMAD}}$ is larger than 1 could indicate that the individual uncertainties are generally underestimated.

However, we note that at the brighter end of the sample the Grizli-derived fluxes $F_{\text{CLEAR}}$ from the CLEAR$^{\text{ER}}$ data set are generally larger than those measured using the 3D-HST pipeline $F_{\text{DIST}}$. We explore this further by dividing the full sample into bright and faint lines, defined as $F_{\text{CLEAR}} < 10^{-16}$ and $>10^{-16}$ erg s$^{-1}$ cm$^{-2}$, respectively. For the brighter lines, we find that the CLEAR$^{\text{ER}}$ fluxes are systematically higher than the 3D-HST fluxes by $\sim 0.05$ dex. For the fainter lines, we find general systematic agreement with no offset on average. Splitting by line flux, the $\sigma_{\text{NMAD}}$ of the fainter lines ($<10^{-16}$ erg s$^{-1}$ cm$^{-2}$) is 0.96 while for the brighter lines ($>10^{-16}$ erg s$^{-1}$ cm$^{-2}$) it is 1.86. This indicates that the larger $\sigma_{\text{NMAD}}$ for the full population is fully driven by the discrepancy at the brighter end.

In summary, the Grizli-derived redshifts measured from the CLEAR$^{\text{ER}}$ data set are generally consistent with earlier measurements from the ground and from the 3D-HST survey (Momcheva et al. 2016). We find an overall redshift precision of the grism of $\sigma_{\text{NMAD}} = 0.0014$ in $\Delta z/(1+z)$ for galaxies with a secure emission line detected. For bright emission lines ($>10^{-16}$ erg s$^{-1}$ cm$^{-2}$), we find that the Grizli-derived line fluxes derived as a part of the CLEAR$^{\text{ER}}$ processing are $\sim 0.05$ dex higher than those measured from the 3D-HST pipeline—using the same G141 grism data. However, for faint lines ($<10^{-16}$ erg s$^{-1}$ cm$^{-2}$), which comprise the slight majority (70%) of the CLEAR$^{\text{ER}}$ line detections, the Grizli-derived line fluxes are in excellent systematic agreement with those from 3D-HST.

6.2.3. Line Flux Limits

In this subsection, we describe the emission line flux density limits of the CLEAR$^{\text{ER}}$ spectra. The line sensitivity of the grisms depend on three factors: (1) the on-source observing time, (2) the spatial extent of the emission, and (3) the observed wavelength of the emission line. The sensitivity is lower for galaxies with more extended emission, which distribute over a larger number of pixels on the WFC3 detector and collect more spatially integrated noise per observing time.24 The line sensitivity also depends on the wavelength-dependent throughput of the grisms (see the inset of Figure 10). For both the WFC3/G102 and G141 grisms, the throughput is generally higher at the redder end. The G141 grism is roughly twice as sensitive as the G102 primarly because the spectral resolution is twice as low (Kuntschner et al. 2011). With the G141 grism spectra taken through the 3D-HST survey, Momcheva et al. (2016) show that the line uncertainty scales linearly with the size of the observed galaxy and as the squared inverse with the throughput.

As follows, we explore the line flux limits in the CLEAR$^{\text{ER}}$ data set. By its construction, this data set contains programs with a range of observing times (Figure 5) and that use one or both of the WFC3 grisms. As such, we want to quantify the line flux limit as a function of on-source exposure time and observed wavelength.

In Figure 11, we show the $S/N$s of the emission lines measured in CLEAR$^{\text{ER}}$ as a function of line flux, exposure time, and observed wavelength. The gray line in each panel is fixed.

The wavelength and exposure time dependence is quantified explicitly in Figure 10, which shows the empirically derived 5σ emission line depth for the lines in four grism wavelength windows. At a given exposure time and wavelength, the $S/N$ of the observed lines is higher at the redder end of each of the grisms. For low exposure times (<4 hr), the 5σ depth ranges from $\sim 8 \times 10^{-17}$ erg s cm$^{-2}$ at the blue end of the G102 grism, to $\sim 4 \times 10^{-17}$ erg s cm$^{-2}$ in the G141 grism. The latter is consistent with the emission line limit derived from the full sample of two orbit G141 data from the 3D-HST survey (Momcheva et al. 2016). For the deepest G102 data included in this paper (which are that way in a large part because they include the ultradeep observations from the FIGS survey; Pirzkal et al. 2017, 2018), the emission line depth reaches $\sim 2 \times 10^{-17}$ erg s cm$^{-2}$. In general, theredder G141 grism is twice as sensitive as the G102 grism and the redder end of each grism is more sensitive than their bluer end.

In Figure 12, we show how the depth of the CLEAR data set maps onto the plane of SFR versus stellar mass for star-forming galaxies. The Hα and [O II] emission lines are detected in more than 50% of galaxies with SFRs $> 1 M_\odot$ yr$^{-1}$ over the redshift range 0.7 $< z < 2.5$. The weaker Hβ line is generally detected only in massive and more star-forming galaxies. The doubly ionized [O III] line is preferentially detected in the upper end of the main sequence. The [O III] line is expected to be brighter above the main sequence than below it, given the scaling between the specific star formation rate (sSFR) and the ionization parameter (Papovich et al. 2022).

7. Science Goals and Results

The original CLEAR proposal listed several primary science goals, including using the grism data set (1) to measure spectroscopic redshifts for hundreds of galaxies in the redshift range $1 < z < 8$ to fainter limits than possible from ground-based spectroscopy; (2) to provide a measurement of the Lyα distribution function; and (3) to measure the evolution in the Lyα equivalent width distribution function as a test of models of reionization. CLEAR also contained many secondary science goals as the data set provides spectroscopic coverage of features from stellar populations and nebular regions from rest-frame UV to NIR (depending on the galaxy redshift). Those secondary goals included studying (4) the properties of stellar populations in galaxies at $1 < z < 2$; (5) the gas-phase metallicities and gas ionization in galaxies at $z > 1$; (6) star formation in galaxies via hydrogen recombination lines, e.g.,
Hα, Hβ, and Paβ, and (7) spatially resolved emission in galaxies at z > 0.5. Largely these goals have been realized. In what follows, we summarize the key science results enabled by the CLEAR team to date, and how these relate to the original science goals.

### 7.1. Measurement of Hundreds of Galaxy Redshifts

The HST/WFC3 G102 and G141 spectroscopy cover emission lines and absorption features in galaxy spectra that enable accurate spectroscopic redshifts for galaxies in the CLEAR fields. In total, the CLEARER data set constrains the spectroscopic redshifts of 3900 galaxies in GOODS-N and GOODS-S using the detection(s) (S/N > 3) of the integrated emission of one or more emission lines (including Lyα, MgII, [O II], Hβ, [O III], Hα+[N II], [S II], [S III], or Paβ). Figure 6 shows the distribution of galaxies with redshifts measured from at least one emission line. The redshift distribution has a median of z = 1.02 with an interquartile range (25–75th percentile) of 0.68–1.44. There is a long tail to higher redshift, where the redshift of the 95th percentile extends to 2.42.

Figure 6 also shows the distribution of galaxies with detections of Hα+[N II], [O III], Hβ, or [O II]. Only galaxies with S/N > 3 in the labeled emission line are included. The redshift span of these subsamples shifts with the rest wavelength of the respective lines (as the different lines are detected in the grism spectra over different redshift ranges, see Figure 3), with Hα+[N II] being available for z ∼ 0.5–1.5, [O III] for z ∼ 0.7–2.4, Hβ for z ∼ 0.6–2.3, and [O II] for z ∼ 1.1 to z > 3 in some cases. We detect (at S/N > 3) 1724 galaxies in Hα+[N II], 1225 galaxies in [O III], 678 galaxies in Hβ, and 1076 galaxies in [O II]. Note that these are not unique samples as some galaxies are detected in multiple lines at >3σ significance.

### 7.2. Constraints on the Lyα Equivalent Width Distribution Function

A main goal of CLEAR is to constrain Lyα emission from galaxy candidates at z > 6. There are several advantages in the use of space-based, slitless spectroscopy to explore Lyα in galaxies, and these are related to the fact that the HST/WFC3 grism data provide independent observations of the emission in these galaxies without many of the biases, backgrounds, and selection effects that can impact ground-based studies. These include the following.

First, space-based slitless spectroscopy eliminates systematic differences caused by using different optical/NIR spectrographs from ground-based telescopes. Typically optical spectrographs are used to study Lyα at z < 7, and NIR spectrographs are used for higher redshifts (see, e.g., Jung et al. 2018, 2019; Pentericci et al. 2018; Hoag et al. 2019; Fuller et al. 2020).

Second, ground-based surveys have different instrumental sensitivities, seeing variations, and variable night-sky line emission. This produces a time- and wavelength-dependent flux sensitivity to emission lines with variations of a factor of >5 (e.g., Treu et al. 2012; Jung et al. 2020). Furthermore, there is evidence that emission line fluxes from ground-based spectra are 2–4× lower than slitless, space-based data, which can result from (seeing dependent) slit-loss corrections and difficulties in flux calibrations in the case that only an emission line is detected with no continuum (e.g., see Masters et al. 2014).
contrast, space-based observations provide a smoothly varying line flux sensitivity function (e.g., Jung et al. 2022).

Third, slitless spectroscopy targets galaxies indiscriminately—with no target preselection. A single HST/WFC3 G102 observation is sensitive to Lyα from all galaxies with 6.0 < \( z \) < 8.2 (see Figure 3). Ground-based observations require slits and potentially imperfect/biased selection. For instance, there is a natural bias to place slits on brighter objects (and these often have the lower Lyα equivalent width; see Stark et al. 2010; Finkelstein et al. 2013; Oesch et al. 2015; Jung et al. 2018, 2020). Slitless, space-based spectra therefore provide less biased samples compared to ground-based spectra in this regard.

In reality, there are few objects with plausible detections of Lyα in galaxies at \( z > 6 \) in the CLEARER data set. Jung et al. (2022) identify several possible candidates, including one galaxy at \( z \approx 6.51 \). This (lack of) strong detections is consistent with other searches for Lyα from HST/WFC3 grism data, where there are only a handful of detections in galaxies at \( 6 < z < 8 \) (e.g., Schmidt et al. 2016; Tilvi et al. 2016; Larson et al. 2018; Jung et al. 2022). The basic conclusion is that the Lyα emission from all galaxies is substantially weaker than in lower-redshift galaxies (whose Lyα equivalent width distributions were used to predict Lyα line fluxes at \( z > 7 \)). This is particularly true for galaxies with lower UV luminosities, as these are expected to be strong Lyα emitters.

Using CLEAR we have obtained improved constraints on the evolution of the Lyα equivalent width distribution function. In Jung et al. (2022), we combined observations from CLEAR with ground-based data sets to measure this evolution. We found that for all galaxies at \( z > 6 \), Lyα emission is significantly suppressed compared to samples at \( z < 6 \). Interestingly, however, Jung et al. (2022) argue there is tentative evidence that the suppression of Lyα is stronger for galaxies with lower UV luminosities. This means that there is additional attenuation/absorption of Lyα photons in lower-luminosity galaxies. This can be explained if reionization is highly inhomogeneous, where the more UV-luminous galaxies blow larger ionized “bubbles” around them (e.g., Finlator et al. 2009; Pentericci et al. 2014; Katz et al. 2019). Once these bubbles reach sizes of \( \sim 1 \) physical Mpc, then the Lyα photons from the source have been sufficiently redshifted (compared to the Hubble flow) that attenuation is mitigated (e.g., Mason & Gronke 2020; Park et al. 2021; Qin et al. 2022; Smith et al. 2022). In this way, the Lyα photons from UV brighter objects are less impacted than lower-luminosity objects that have small ionized bubbles surrounding them. However, as we discuss in Jung et al. (2022), the constraints based on the current data sets are still too small given the sample size, but this makes predictions for both JWST and NGRST that should identify Lyα emission at fainter flux sensitivities and for vastly larger samples.

7.3. Studies of Stellar Populations in Distant Galaxies

Several studies have studied the stellar populations of galaxies as derived from their stellar continuum features in the CLEARER data set.

For high-redshift galaxies, the G102 and G141 grism data probe many of the well-known spectral features of stellar populations in rest-frame optical. Due to uncertain and time-variable sky backgrounds, these features are difficult (but not impossible) to detect in high-redshift galaxies from the ground.
Figure 13 shows a stacked spectrum of galaxies selected to be “quiescent” based on their \( UVJ \) rest-frame colors. These are a subset of those galaxies studied by Estrada-Carpenter et al. (2019, 2020). The shape of the spectrum and strength of spectral features are sensitive to age and metallicity (see also the discussions in Estrada-Carpenter et al. 2019, 2020). The spectrum on the blue side is consistent with solar metallicity. The shape of the spectrum on the red side is more consistent with superSolar abundances. These facts could be an indication of higher \( \alpha/Fe \) ratios (such as titanium, magnesium, oxygen) at a fixed [Fe/H], consistent with other observations of quiescent galaxies at low and high redshifts (Conroy & van Dokkum 2012; Choi et al. 2014; Kriek et al. 2019). This could also be related to age as the galaxies in the stack span a range of redshift, with higher-redshift galaxies contributing more at rest-frame blue wavelengths. At higher redshift, quiescent galaxies show evidence of younger stellar populations (e.g., Estrada-Carpenter et al. 2019) and so the differences in the spectra could be representative of differences in population age.

Estrada-Carpenter et al. (2019) used G102 data for a sample of quiescent-selected galaxies at \( 1 < z < 1.8 \) to measure constraints on the ages, star formation histories, and metallicities of the galaxies. They showed that massive quiescent galaxies at these redshifts already harbor older stellar populations (with light-weighted ages indicating formation epochs \( z_f > 2.5 \)) and that they had already enriched to stellar metallicities approaching or exceeding solar (\( \approx Z_\odot \)). This indicates that the massive \( z > 1 \) quiescent galaxies experienced very early star formation and chemical enrichment. Estrada-Carpenter et al. (2020) used models with flexible star formation histories to show that quiescent galaxies with the earliest formation have more compact morphologies, indicating a correlation between formation and compactness. The stellar population constraints on stellar mass, dust attenuation, and SFRs have also been used to study galaxy gas-phase metallicity–mass and ionization–mass relations (Papovich et al. 2022; see also the next subsection) and to study the evolution of “green valley” galaxies in transition from the star-forming to quiescent sequences (V. Estrada-Carpenter et al. 2023, in preparation).

One of the big remaining questions is, where are the progenitors of the massive, quiescent galaxies with solar enrichment? Measurements of stellar metallicities (derived from continuum spectroscopy, including our own work with CLEAR) of massive, quiescent galaxies at \( z > 1 \) are consistent with solar abundances (e.g., Onodera et al. 2015; Kriek et al. 2016, 2019; Estrada-Carpenter et al. 2019; Lonoce et al. 2020). Most studies of the gas-phase metallicities of star-forming galaxies at \( z = 1–2 \) (including our own analysis from CLEAR) find that star-forming galaxies with stellar mass \( \log M_*/M_\odot \geq 10.5 \) are subsolar (e.g., Steidel et al. 2014; Sanders et al. 2015; Strom et al. 2017; Henry et al. 2021; Papovich et al. 2022), with \( 12 + \log(O/H) \) approximately 0.2–0.3 dex below the solar value \( (12 + \log(O/H))_\odot = 8.69; \) Asplund et al. 2009). Therefore, we are missing those star-forming galaxies at \( z \geq 2 \) that have high levels of enrichment, similar to the metallicities inferred for quiescent galaxies at this epoch. This discrepancy is compounded by evidence that the \( \alpha/Fe \) ratios are enhanced (Steidel et al. 2016; Strom et al. 2018, 2022; Topping et al. 2020), which requires even lower iron abundances (which typically dominate \( Z_\odot \); see Estrada-Carpenter et al. 2019). Oxygen (an \( \alpha \) element) stems primarily
from observations of nebular gas, while the iron abundance dictates the shape of the stellar continuum, and direct measures of \( \alpha/\text{Fe} \) from stellar continuum are only possible for rare cases of bright galaxies, see, e.g., Kriek et al. (2016, 2019). As illustrated in Figure 13, the shape of the spectrum and the strength of the spectral features are sensitive to changes in the metallicity. This presents an opportunity for observations from future studies with JWST and the Nancy Grace Roman Space Telescope that will enable both higher-quality data and larger samples of galaxies to explore this problem of the “missing,” high-metallicity progenitors of the quiescent galaxies at \( z > 1 \).

### 7.4. Emission Line Diagnostics of Galaxies

The CLEARER data set enables a large range of science using galaxy emission lines as diagnostics of star formation, gas conditions including ionization and metallicity, and black hole activity.

In Figure 14, we show stacks of the 1D CLEARER spectra in bins of stellar mass and redshift. Each row in the stack contains \( \sim 70 \) individual galaxy spectra. The stacks illustrate the change in the relative luminosity of the emission lines as a function of stellar mass (left panel) and redshift (right panel), and indicate which strong rest-optical lines are accessible (although not necessarily detected in individual galaxies) in the CLEARER spectral range.

Several papers have used the CLEARER data set to investigate how star formation proceeds in galaxies from the local universe to cosmic high noon (Cleri et al. 2022a; Matharu et al. 2022). Cleri et al. (2022a) identified galaxies at \( z < 0.29 \) where both \( \text{H} \alpha \) and \( \text{Pa}\beta \) emission are detected in the integrated (1D) CLEAR spectra. Because these are both hydrogen recombination lines, their theoretical line ratio is relatively insensitive to conditions in the H II regions (over a wide range of temperature and density; see Osterbrock & Ferland 2006).

By comparing the \( \text{Pa}\beta \) emission to dust-corrected UV emission, one can probe star formation on \( \sim 5 \) Myr timescales (the lifetimes of O-type stars responsible for ionizing the nebula which produce emission from hydrogen recombination) with \( \sim 100 \) Myr timescales (the lifetimes of B-type stars responsible for the UV continua). Cleri et al. (2022a) showed that low-mass galaxies have more scatter in the \( \text{Pa}\beta/\text{UV} \) ratios than higher-mass galaxies, and they argued that this is a result of increased burstiness in the galaxies’ star formation histories.

Matharu et al. (2022) used the spatially resolved \( \text{H} \alpha \) emission for galaxies in CLEARER to study how star formation proceeds at \( 0.5 < z < 1 \). They showed that the sizes of \( \text{H} \alpha \) disks are larger than that of the stellar continuum (measured in a broadband that covers the same rest-frame wavelengths as the \( \text{H} \alpha \) emission), but that there is redshift evolution compared to samples at \( 1 < z < 1.5 \) (from 3D-HST; Nelson et al. 2016) and \( z > 2 \) (from KMOS3D; Wilman et al. 2020). Matharu et al. (2022) showed that this evolution is consistent with star formation proceeding in an “inside-out” fashion in galaxies.

The CLEARER data have also been used to study the physical conditions of the nebular (Simons et al. 2021; Backhaus et al. 2022, 2023; Papovich et al. 2022) and highly ionized (Cleri et al. 2022b, 2023) gas in galaxies.

Papovich et al. (2022) used measurements of the fluxes (and flux ratios) of the [O II], [O III], and H\( \beta \) emission lines from the CLEAR spectra to study the ionization and chemical enrichment of galaxies over \( 1.1 < z < 2.3 \). They showed that at fixed stellar mass (\( \log M_*/M_\odot \sim 9.4-9.8 \)), higher-redshift galaxies have lower gas-phase metallicities and higher ionization parameters than they do at lower redshift (\( z \sim 0.2 \)). Moreover, at fixed mass and/or at fixed metallicity, higher-redshift galaxies have ionization parameters that correlate positively with their sSFR. Papovich et al. (2022) posit that this correlation could arise because the gas density (and/or gas

**Figure 12.** The emission line detection rate of the CLEARER data set is shown in the plane of galaxy stellar mass and SFR. Left panel: the full CLEAR galaxy sample \((m_{\text{F105W}} < 25)\) is shown in black and the galaxies with emission line detections \((S/N > 5)\) in each of the indicated lines is shown by the color contours. Right panel: the fraction of galaxies in the full CLEAR sample with an emission line detected \((i.e., N_{\text{detections}}/N_{\text{total}})\) in each of the indicated lines. \( N_{\text{detections}} \) is the total number of galaxies in CLEAR with observing conditions that allow for a potential line detection—i.e., for each object, it is considered whether the observed wavelength of the line overlaps with that available G102/G141 grism coverage.
for the stack. The error bars show the standard deviation in the sample. The dotted lines show wavelengths of prominent absorption lines and spectral indices.

graph showing the G102 and G141 data for 64 galaxies selected to be quiescent both increase with increasing sSFR.

Figure 13. The rest-frame average spectrum of quiescent galaxies at 0.8 < z < 2.5 with stellar mass log M*/M⊙ > 10.5 in the CLEAR fields is shown. The spectrum shows the G102 and G141 data for 64 galaxies selected to be quiescent (based on UVJ rest-frame colors) and shifted to the rest-frame wavelength using their measured grism redshifts (heavy black line). The spectrum of each galaxy is normalized in the rest-frame wavelength range 4600–5500 Å before taking the median flux density for the stack. The error bars show the standard deviation in the sample. The dotted lines show wavelengths of prominent absorption lines and spectral indices (see Worthey 1994). The thin, colored lines show simple stellar population models (FSPS; Conroy et al. 2009; Conroy & Gunn 2010), formed in an instantaneous burst with an age of 2.5 Gyr, and metallicity (as indicated, see legend), binned to R ~ 200.

geometry) and the escape fraction of ionizing photons likely both increase with increasing sSFR.

Backhaus et al. (2022) used the CLEAR spectra to study the spatially integrated line ratios [O III]/Hβ, [S II]/(Hα+[N II]), and [Ne III]/[O II] in galaxies spanning 0.6 < z < 2.5. Comparing with photoionization models, they conclude similarly to Papovich et al. (2022) that galaxies with higher mass and lower SFRs have both higher metallicity and lower ionization parameters. To distinguish ionization from star formation and AGNs, they construct the diagnostic diagrams “OHNO” ([O III]/Hβ versus [Ne III]/[O II]) and “unV087” ([O III]/Hβ versus [S II]/(Hα+[N II])) of this sample. The “un” in “unV087” indicates that Hα+[N II] are spectrally unresolved in the HST grisms. While the “unV087” diagram poorly separates the AGN and [Ne V]-emitting galaxies (indicating high ionization) from the star-forming galaxies, Backhaus et al. (2022) argue that the “OHNO” diagram does effectively discriminate these populations. They conclude that “OHNO” will be a useful indicator for AGN activity and the ionization conditions in high-redshift galaxies observed by the JWST.

Backhaus et al. (2023) measured the radial gradient of the [O III]/Hβ ratio in galaxies over 0.6 < z < 1.3 to study the spatial variations of their ionization and to search for low-luminosity AGNs. While the majority of the galaxies are consistent with a zero gradient, they argue that 6%–16% of the galaxies in the sample likely have nuclear [O III]/Hβ ratios that are 0.5 dex higher than they are in their outer regions. Backhaus et al. (2023) argue that these galaxies (which are generally not detected in X-rays) may host low-luminosity AGNs. Furthermore, they did not find evidence for a significant population of sources with off-nuclear ionization.

Simons et al. (2021) used the spatially resolved maps of the [O II], Hβ, [O III], Hα+[N II], and [S II] emission lines from CLEARER to derive gas-phase metallicity maps for 238 star-forming galaxies over the redshift range 0.6 < z < 2.6. They measured the radial gradient of these metallicity maps, and report that the majority of galaxies at this redshift are consistent with a null or positive (aka “inverted,” i.e., more metal rich in galaxy outskirts) metallicity gradient (see also, e.g., Wang et al. 2017, 2019, 2020; Curti et al. 2020; Li et al. 2022). This finding is somewhat puzzling because it runs counter to simple expectations from star formation and stellar evolution. In star-forming galaxies at this redshift, the star formation surface density is on average higher in the galaxy centers than it is in the galaxy outskirts (Nelson et al. 2016; Tacchella et al. 2018). Given that, we might expect for these galaxies to form negative metallicity gradients gradually (more metal rich in the galaxy centers) through stellar evolution and local chemical enrichment. Simons et al. (2021) argue that the ubiquity of null/positive gradients in these galaxies implies that their gas-phase metals are being redistributed on galaxy scales (or their ISMs are being unevenly diluted through metal-poor accretion) on timescales shorter than the short time (≤100 Myr; Simons et al. 2021) it would take for them to develop naturally a declining metallicity gradient through stellar evolution. These metals could be redistributed around galaxies through galactic scale outflows and/or the high levels of turbulence in the ISMs of these galaxies (e.g., Förster Schreiber et al. 2006, 2009; Weiner et al. 2006; Kassin et al. 2007, 2012; Wisnioski et al. 2015; Simons et al. 2016, 2017; Übler et al. 2019; Price et al. 2020).
Finally, Cleri et al. (2022b) used the CLEAR 1D spectra to search for high-ionization [Ne V] (λ3426) emission in galaxies over 1.4 < z < 2.3. [Ne V] has an exceedingly high creation potential (97.11 eV), and is an indicator for highly energetic photoionization—e.g., from AGNs, supernova radiation, and/or a hard ionization spectrum from stars (N. J. Cleri et al. 2023, in preparation). Cleri et al. (2022b) select 25 galaxies in the CLEAR sample with [Ne V] detected. Based on the ratios of the [O III]/H$\beta$ lines in these galaxies, they show that most of the sample is consistent with photoionization from an AGN.

8. Summary

This paper presents an overview of the CLEAR survey—a 130 orbit HST/WFC3 spectroscopic and imaging program. The CLEAR observations include 10 and 12 orbit HST/WFC3 G102 grism spectroscopy and F105W imaging in the GOODS-N and GOODS-S legacy fields, respectively. The full data set discussed here includes the WFC3/G102 grism observations from the CLEAR survey and overlapping WFC3/G102+G141 observations from a number of ancillary programs in the HST archive.

We discuss the design of the CLEAR survey, the data processing and products, and the science that has been carried out by the CLEAR team with this data set. Alongside this paper, we release a number of science-ready data products created from this program, including emission line flux catalogs, updated 3D-HST photometric catalogs, and 2D and 1D extracted spectra for 6048 galaxies. These products are available at MAST25 as a High Level Science Product via doi:10.17909/9cjs-wy94.

Figure 14. Stacks of the CLEAR 1D spectra are shown in bins of stellar mass (left panel) and redshift (right panel). Each row includes ~70 emission-line-selected galaxies. The best-fit continuum is subtracted from each spectrum, and the spectra are normalized by the inverse of their luminosity distance squared—so that the brightness corresponds to the luminosity of the sources. Prominent emission line/line complexes in the optical–NIR are indicated.

25 https://archive.stsci.edu/hlsp/clear/
We thank the anonymous referee and data editor for a constructive report that improved the manuscript. We thank Mark Dickinson, Rachael Livermore, and Ryan Quadri for valuable conversations and contributions to the early development of the CLEAR survey. This work is based on data obtained from the Hubble Space Telescope through program number GO-14227. Support for Program number GO-14227 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. This work is supported in part by the National Science Foundation through grant AST 1614668. R.C.S. appreciates support from a Giacconi Fellowship at the Space Telescope Science Institute. V.E.C. acknowledges support from the NASA Headquarters under the Future Space at the Space Telescope Science Institute. This research made use of Astropy (https://www.astropy.org) a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). Some/all of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific observations analyzed can be accessed via doi:10.17909/t9-cffv-wx60. The science-ready data products can be accessed via 10.17909/9cjz-wy94.

Software: eazy-Py (Brammer 2021), EAzy (Brammer et al. 2008), Grizli (Brammer 2019), AstroDrizzle (Gonzaga et al. 2012), Astropy (Astropy Collaboration et al. 2013, 2018), NumPy (Harris et al. 2020), Matplotlib (Hunter 2007).