CrossMark

The Hobby–Eberly Telescope Dark Energy Experiment (HETDEX) Survey Design, Reductions, and Detections[*]

Karl Gebhardt[1], Erin Mentuch Cooper[1,2], Robin Ciardullo[3,4], Viviana Acquaviva[5,6], Ralf Bender[7,8],
William P. Bowman[3,4], Barbara G. Castanheira[9], Gavin Dalton[10], Dustin Davis[1], Roelof S. de Jong[11], D. L. DePoy[12,13],
Yaswant Devarakonda[12], Sun Dongsheng[14], Niv Drory[2], Maximilian Fabricius[7,8], Daniel J. Farrow[7,8], John Feldmeier[15],
Steven L. Finkelstein[1], Cynthia S. Froning[2], Eric Gawiser[16], Caryl Gronwall[3,4], Laura Herold[17], Gary J. Hill[1,2],
Ulrich Hopp[7,8], Lindsay R. House[1], Steven Janowiecki[2], Matthew Jarvis[10], Donghui Jeong[18,19], Shardha Jogee[1],
Ryota Kakuma[14], Andreas Kelz[11], W. Kollatschny[20], Eiichiro Komatsu[17,21], Mirko Krumpe[11], Martin Landriau[22],
Chenxu Liu[1], Maja Lujan Niemeyer[17], Phillip MacQueen[2], Jennifer Marshall[12,13], Ken Mawatari[14], Emily M. McLinden[1],
Shiro Mukae[14], Gautam Nagaraj[3,4], Yoshiaki Ono[14], Masami Ouchi[14,21,23], Casey Papovich[12,13], Nao Sakai[14], Shun Saito[21,24],
Donald P. Schneider[3,4], Andreas Schulze[21,23], Khavvia Shanmugasundararaj[1], Matthew Shetrone[25], Chris Sneden[1],
Jan Snigula[7,8], Matthias Steinmetz[11], Benjamin P. Thomas[1], Brianna Thomas[26], Sarah Tuttle[26], Tanya Urrutia[11],
Lutz Wisotzki[11], Isak Wold[27], Gregory Zeimann[28], and Yechi Zhang[14]

[1] Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA; gebhardt@utexas.edu
[2] McDonald Observatory, The University of Texas at Austin, Austin, TX 78712, USA
[3] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[4] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[5] Physics Department, CUNY NYC College of Technology, Brooklyn, NY 11201, USA
[6] Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
[7] Max-Planck Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany
[8] University Observatory, Fakultät für Physik, Ludwig-Maximilians University Munich, Scheiner Strasse 1, D-81679 Munich, Germany
[9] Department of Physics, Baylor University, One Bear Place 97316, Waco, TX 76798, USA
[10] University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
[11] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[12] Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242 USA
[13] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242 USA
[14] Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan
[15] Department of Physics, Astronomy, Geology & Environmental Sciences, Youngstown State University Youngstown, OH 44555, USA
[16] Physics and Astronomy Department, Rutgers, The State University, Piscataway, NJ 08854-8019, USA
[17] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-85748 Garching, Germany
[18] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[19] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[20] Institut für Astrophysik, Universität Göttingen, Friedrich-Hund Platz 1, D-37077 Göttingen, Germany
[21] Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), the University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[22] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[23] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[24] Institute for Multi-messenger Astrophysics and Cosmology, Department of Physics, Missouri University of Science and Technology, 1315 N Pine St, Rolla, MO 65409, USA
[25] UC Observatories UC Santa Cruz 1156 High Street Santa Cruz, CA 95064, USA
[26] Department of Astronomy, University of Washington, Seattle, 3910 15th Avenue NE, Room C319, Seattle WA 98195-0002, USA
[27] Astrophysics Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[28] Hobby Eberly Telescope, University of Texas, Austin, Austin, TX 78712, USA

Abstract

We describe the survey design, calibration, commissioning, and emission-line detection algorithms for the Hobby–Eberly Telescope Dark Energy Experiment (HETDEX). The goal of HETDEX is to measure the redshifts of over a million Lyα emitting galaxies between $1.88 < z < 3.52$, in a 540 deg^2 area encompassing a comoving volume of 10.9 Gpc^3. No preselection of targets is involved; instead the HETDEX measurements are accomplished via a spectroscopic survey using a suite of wide-field integral field units distributed over the focal plane of the telescope. This survey measures the Hubble expansion parameter and angular diameter distance, with a final expected accuracy of better than 1%. We detail the project's observational strategy, reduction pipeline, source detection, and catalog generation, and present initial results for science verification in the Cosmological Evolution Survey, Extended Groth Strip, and Great Observatories Origins Deep Survey North fields. We demonstrate that our data

[*] Based on observations obtained with the Hobby–Eberly Telescope, which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen.

reach the required specifications in throughput, astrometric accuracy, flux limit, and object detection, with the end products being a catalog of emission-line sources, their object classifications, and flux-calibrated spectra.

Unified Astronomy Thesaurus concepts: Dark energy (351)

1. Overview

Data from supernova distances (Riess et al. 1998; Perlmutter et al. 1999; Riess et al. 2021) combined with the results from the cosmic microwave background radiation (Bennett et al. 2013; Komatsu et al. 2014; Planck Collaboration VI 2020) show that the universe is undergoing accelerated expansion compared to what is expected for a universe with only radiation (photons and massless neutrinos) and matter (baryons, dark matter, and massive neutrinos). The additional acceleration accounts for about 70% of the current expansion and thus the universe's mass–energy content. The source of this acceleration is called "dark energy", a designation that reflects the gross ignorance of the scientific community. Yet, dark energy has profound implications for the formation of galaxies and the evolution of the universe. The theoretical community has proposed a number of equally compelling ideas, such as vacuum energy (a cosmological constant), a new scalar field (evolving dark energy or quintessence), and a modification to gravity (see the review by Frieman et al. 2008). Any one of these proposals requires a fundamental change in our understanding of the laws of physics, and the general consensus is that the resolution of this problem will be nothing less than revolutionary.

Despite significant differences in magnitude from the theoretical prediction, the most promising candidate for dark energy remains vacuum energy, or Einstein's "cosmological constant", as estimates from supernovae observations and the cosmic microwave background appear consistent with this interpretation (e.g., Betoule et al. 2014; Scolnic et al. 2018; Adachi et al. 2020; Aiola et al. 2020; Planck Collaboration VI 2020; Dutcher et al. 2021). Studies based on measurements of the large-scale clustering of galaxies provide similar conclusions (e.g., Alam et al. 2021; DES Collaboration et al. 2021). There are a number of uncertainties concerning the vacuum energy interpretation, with the most serious being the difference in the Hubble Constant (H_0) using different measurement techniques. Specifically, there are lingering differences between local measurements of H_0 and the value inferred from the power spectrum of the early universe and the assumption of a ΛCDM cosmology (Freedman et al. 2019; Hikage et al. 2019; Joudaki et al. 2020; Planck Collaboration VI 2020; Wong et al. 2020; Di Valentino 2021; Di Valentino et al. 2021; Heymans et al. 2021; Riess et al. 2021). While these differences are small, it the era of precision cosmology, they may signify new physics. To address the problem, a large number of programs has been designed to measure the effect of dark energy, with most focused at late times when dark energy is expected to dominate (see Figure 2 of Vargas-Magana et al. 2019). Te best way to constrain the history of dark energy is to measure the expansion rate of the universe over as wide a time baseline as possible. A cosmological tracer that can be used for this purpose is Lyα emitting galaxies (LAEs). LAEs have been detected over a large redshift range, and their redshifted 1215.67 Å line can easily be detected using low-resolution spectroscopy or narrow-band imaging (e.g., Cowie & Hu 1998; Gronwall et al. 2007; Ouchi et al. 2008; Adams et al. 2011).

The Hobby–Eberly Telescope Dark Energy Experiment (HETDEX) is a spectroscopic survey aimed at measuring the Hubble parameter, $H(z)$, and the angular diameter distance, $D_A(z)$, in order to determine the potential evolution of the dark energy density. Without compelling theoretical guidance or preference for one specific cosmological model compared to any other, we focus HETDEX on being able to provide a direct measure of the dark energy density for a cosmological constant model. Our target specification is a 3σ detection of the dark energy density at $z = 2.4$ assuming a cosmological constant. This accuracy translates into a measurement precision of 0.9% on $H(z = 2.4)$ and 0.8% on $D_A(z = 2.4)$ between redshifts $1.88 < z < 3.52$. All instrumental, observational, and calibration requirements follow from this top-level specification.

To put HETDEX in context with other published, ongoing and planned experiments, the cleanest comparison is to use the expected uncertainties on the distance estimates. Currently, the uncertainties on H and D_A at different redshifts range from 1.8%–3% (see the summary in DES Collaboration et al. 2021). The Baryon Oscillation Spectroscopic Survey (BOSS) and its extension, eBOSS, measure D_A with accuracy of 1.8% at $z = 0.5$ (Bautista et al. 2020; Gil-Marín et al. 2020), 2.0% at $z = 0.7$ (de Mattia et al. 2020), and $\sim 3\%$ at $z = 2.3$ (du Mas des Bourboux et al. 2020), and the Dark Energy Survey (DES) gives an uncertainty of 2.7% at $z = 0.84$ (DES Collaboration et al. 2021). Various ongoing missions, such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016), the Prime Focus Spectrograph (PFS; Takada et al. 2014), and Euclid (Laureijs et al. 2011) are expected to achieve a precision of $\sim 0.5\%$ at $z \sim 1$. Eventually, the Nancy Grace Roman Telescope (Spergel et al. 2015) will produce an uncertainty below 0.5% at $z \sim 1$. The most comparable experiment to HETDEX is DESI, which plans to push out to $z > 2$ with a precision similar to that of HETDEX, i.e., below 1%. The techniques used by DESI and HETDEX are different and well complement each other. The design of HETDEX is to provide a measure of the distance scales at $z > 2$ that is comparable to the most accurate low-redshift measurements. By not relying on theoretical models to design the survey, we aim to provide the most accurate observational comparison to low-redshift programs.

A survey the size of HETDEX will provide significant science beyond just a measure of $H(z)$ and $D_A(z)$, including measurements of additional cosmological parameters, constraints on galactic and active galactic nuclei (AGN) evolution, and information on the halo populations of the Milky Way. For this paper, we do not discuss the full scientific benefit from HETDEX; instead we focus on how our goals for $H(z)$ and $D_A(z)$ set the requirements for all calibrations and analysis. The redshift range and the accuracies expected on the expansion rates are designed to provide a unique and significant measure of the evolution of dark energy.

HETDEX will use a set of 74 FL(IFU) fiber arrays, which are currently installed at the focal surface of the 10 m class Hobby–Eberly Telescope (HET) (see Hill & HETDEX Consortium 2016; Indahl et al. 2016, and Hill et al. 2021, submitted). For the data presented in this paper, the IFUs were installed over several years, and the population of the focal plane ranges from 20 to 71 active IFUs. Each IFU contains a bundle of 448 1″5 diameter fibers as described in
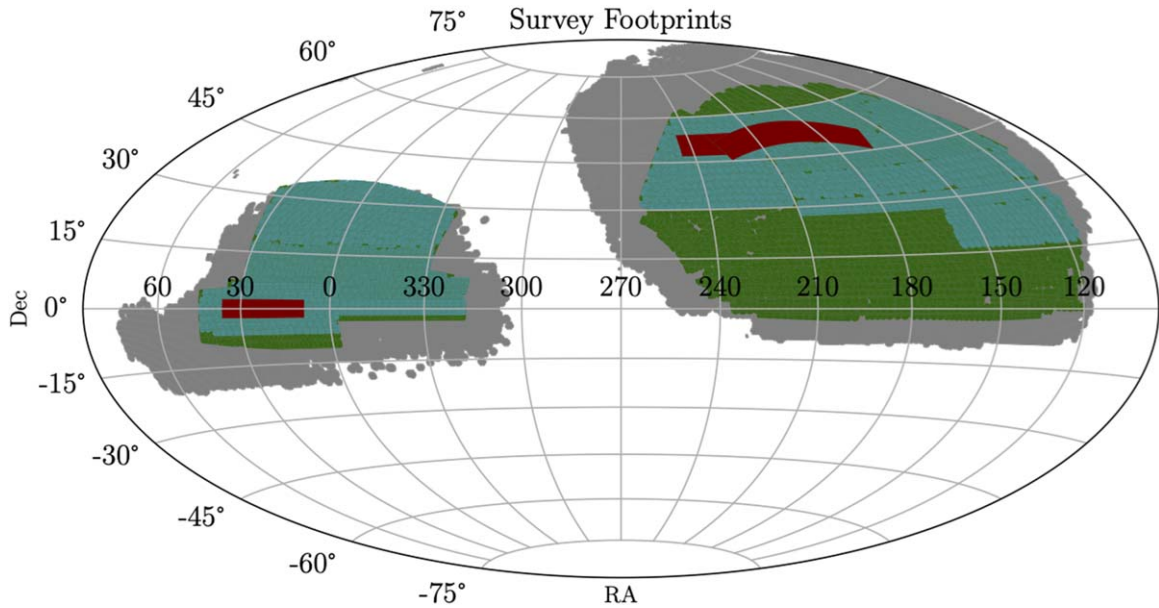
Figure 1. The HETDEX field compared to overlapping large-area surveys. The red regions display the 540 deg² baseline fields of HETDEX. The green, cyan and gray areas show, respectively, the BOSS (Dawson et al. 2013), eBOSS (Dawson et al. 2016), and DESI (DESI Collaboration et al. 2016) footprints.

Kelz et al. (2014). The IFUs feed two low-resolution Visible Integral-field Replicable Unit Spectrographs (VIRUS) covering the wavelength range between 3500 and 5500 Å. The full set of 74 IFUs will contain 33,152 fibers, distributed over the central 18′ of the telescope's 22′ diameter field of view. When used with a standard 3 point dither pattern, the instrument produces a focal surface filling factor of about 1 in 4.6.

The methodology of HETDEX is straightforward. At each location in the sky, three 6 min exposures are taken in a triangular dithering pattern to fill in the gaps between the fibers. Twilight sky frames produce both the wavelength calibration and the fiber-to-fiber normalization, while field stars with known magnitudes and positions primarily from the Sloan Digital Sky Survey (SDSS; York et al. 2000; Abazajian et al. 2009) determine the overall flux calibration and astrometric solution. Each emission-line source falling within an IFU is found via custom detection software, and the objects' counterparts are identified on complementary images acquired from the Blanco, Mayall, Subaru, and Hubble Space Telescopes. Using the Bayesian analysis of Leung et al. (2017), Ouchi et al. (2020), and Farrow et al. (2021) with updates in D. Davis et al. (2021, in preparation), each emission-line source is then classified as either a $1.88 < z < 3.52$ Lyα emitter, a $z < 0.5$ [O II] emitting galaxy, or less often, some other type of object.

The HETDEX survey covers two distinct regions of the sky extending 540 deg². The high-decl. region, referred to as the "spring" field, covers 390 deg², while the equatorial "fall" field, extends over 150 deg². Figures 1 and 2 show the field locations. We expect that around 460,000 IFU observations will be taken within these boundaries, with the exact placement of the fields dependent primarily on the observing conditions and the weather pattern. The expectation is that during the course of the survey, the project will detect over one million Lyα emission lines. The large-scale clustering of the LAEs will then provide the cosmological parameters sought by the experiment.

This paper presents the observational design of the HETDEX survey in Section 2, the survey requirements to reach the target cosmological constraints in Section 3, the instrument layout in Section 4, our dither strategy in Section 5, the reduction procedures needed to reduce and calibrate the spectra in Section 6, line and continuum detection algorithms in Section 7, simulations for completeness in Section 8, and our method of line identification in Section 9. HETDEX will make its catalogs and data public; at present, all data releases have been internal. The latest internal catalog is called HDR2 (for the second HETDEX Data Release).

2. Observational Setup

The science goal of achieving $<1\%$ uncertainty in the cosmological distance measures sets the requirements for the depth and area of the HETDEX experiment. One needs to survey a large enough volume of space to limit the contribution of sample variance and observe deep enough so that the number density of galaxies is sufficient to minimize Poisson shot noise. Thus, the project's exposure times and survey area are defined by our current knowledge of the LAE luminosity function in the redshift range $1.88 < z < 3.52$ (e.g., Gronwall et al. 2007; Ouchi et al. 2008; Ciardullo et al. 2012), the estimated bias of the LAE population (Gawiser et al. 2007; Guaita et al. 2010; Kusakabe et al. 2018; Khostovan et al. 2019), the number of IFUs mounted on the telescope, and the expectation for the allocation of observing time. For reasonable telescope usage, we set the base exposure time to 18 min, which is split into three 6 min dithers with an overhead of 2 min for read time and dithering. Under nominal observing conditions, these exposures should detect ∼2.5 LAEs per IFU. In cases of nonoptimal observing conditions, we use real-time estimates of the image quality, sky transparency, sky brightness, mirror illumination, and target availability to adjust the exposure time.

HETDEX produces a significant amount of data. As of 2020 August 1, the date of the HETDEX Data Release 2, 32% of the planned survey area had been observed. This data set consists of over 3100 telescope pointings containing 160,000 IFU observations (the number of IFUs per pointing changed with time as we added in more units), 215 million spectra (160,000 IFU observations ×448 fibers/IFU ×3 dithers), and 100 TB of data storage. Completion of the full survey is scheduled for 2024.
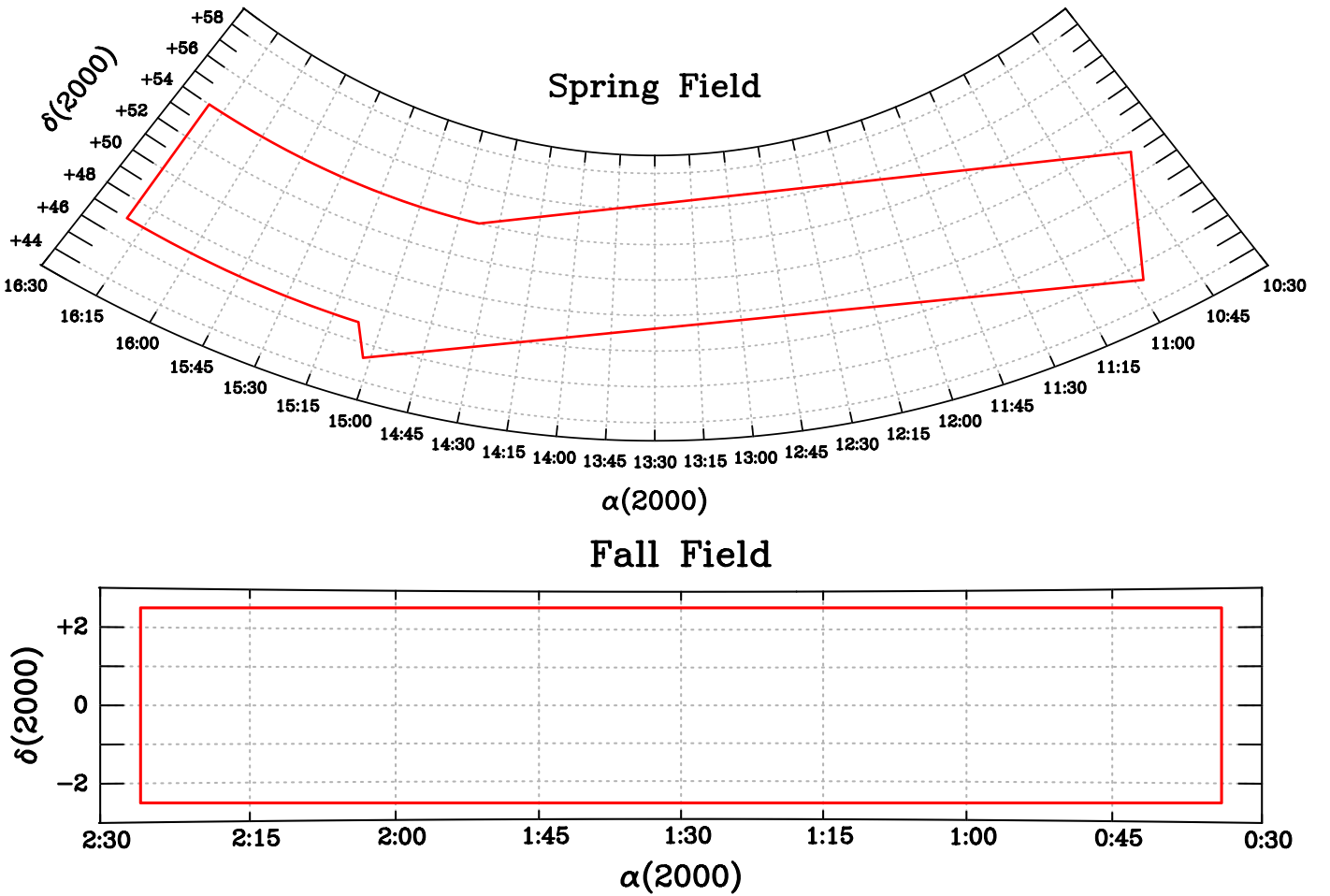
Figure 2. The outline of the HETDEX spring and fall fields. The spring field extends over ∼390 deg² of sky; the fall field, ∼150 deg². Our baseline design contains 4000 pointings in the spring field (covering 15% of the total area) and 2000 pointings in the fall field (covering 21% of the area). Thus our nominal survey will have 90 deg² of spectra when considering the 1/4.6 fill in the focal plane. The lower fill factor in the spring field provides contingency for the survey.

The Hobby–Eberly Telescope is not a fully steerable telescope: while it can rotate to any azimuth, it can only observe at a fixed elevation of 55°. Consequently, observations are only possible when an object passes through the primary mirror's 12° diameter field of view. This means that fields can only be observed twice a night, once with one focal surface orientation toward the east transit and another when the orientation is toward the west transit. Figure 3 shows the final layout of IFUs on the HET focal surface; this semihexagonal pattern ensures that either of the two tracks provides an optimal tiling on the sky. To maintain the regular tiling (Chiang et al. 2013), we populate the hexagon centers (telescope pointings) on the flat rectangle (Figure 2) and project them onto the survey area (Figure 1) by using an area-preserving map (Tegmark 1996). However, because the 74 VIRUS IFUs were installed over a 5 yr period, the early HETDEX observations, which were taken with an incomplete IFU array, had unique on-sky footprints. This necessitated the use of modified field centers in order to optimize sky coverage. Consequently, data taken during the first three years of the HETDEX project do not have a regular sky tiling. These asymmetries do not impact the measurement of large-scale clustering, since the window function of the observations is known very accurately. However, the irregular tiling did produce a significant amount of overlap in the IFU pointings, which allowed us to tune our detection algorithms and improve our understanding of the flux

calibration. Only in the spring of 2020 did the number of active IFUs become large enough to use the preplanned tiling pattern. The window function in the radial (i.e., wavelength) direction is similarly well known, and both directions are considered in the flux limit estimates discussed below.

3. Science Requirements

Shoji et al. (2009), Chiang et al. (2013), and Farrow et al. (2021) discuss the forecasts for the cosmology measures. In order to reach our target accuracies of less than 1% for $H(z)$ and $D_A(z)$, the HETDEX project has a set of science requirements. These are summarized below.

Survey size: The volume and configuration of the survey determine the number of spatial modes that can be used to define the large scale clustering. Since the HET can only access fields that are 35° from the zenith, the footprint of the survey is a compromise between optimizing field observability by having a narrow strip in R.A., avoiding areas of the sky with significant Galactic extinction, and keeping the shape wide enough so that large scale modes can be adequately sampled. The fields shown in Figure 1 sample 10.9 Gpc³ of space between $1.88 < z < 3.52$. The shortest axis of the spring and fall fields are 7° and 5°, respectively. This width is set in order to allow adequate sampling of the largest clustering scales of interest.
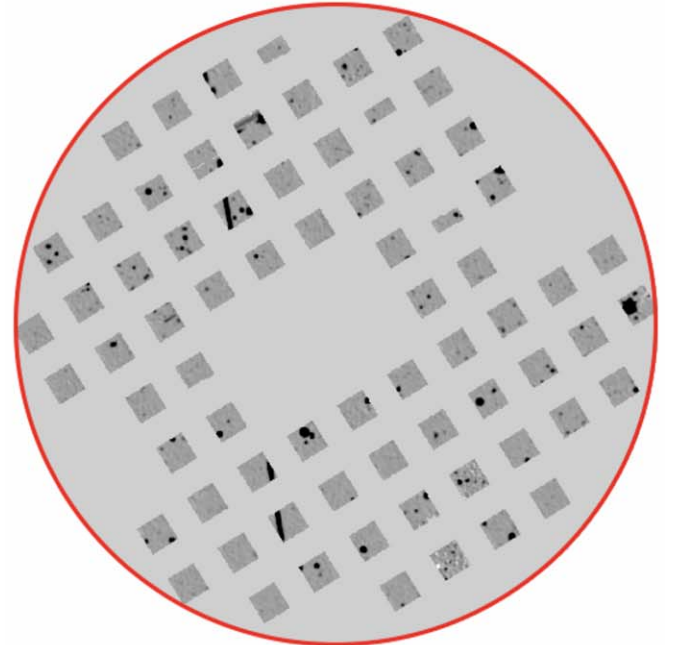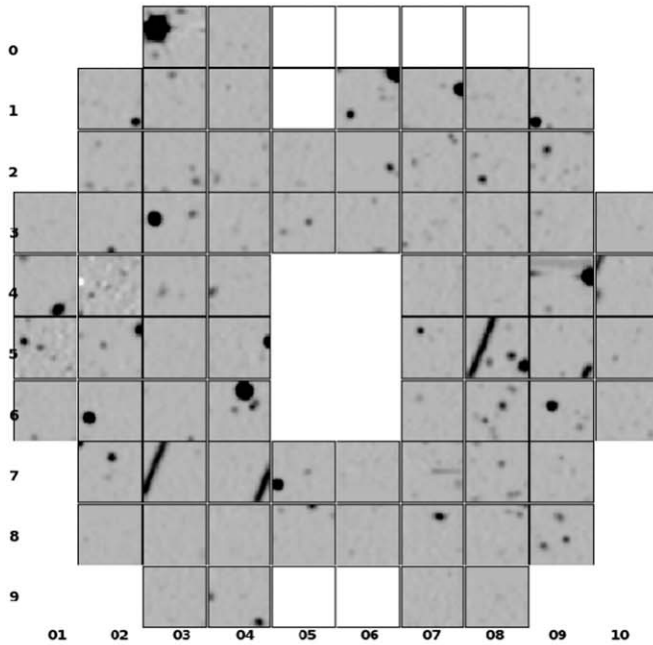
Figure 3. Left: A reconstructed image showing the layout of IFUs on the HET's focal plane for data taken on 2020 June 16. This image is formed by collapsing the 4400–5200 Å spectral region of the VIRUS spectra from a science observation. The image shows the IFUs as being adjacent to each other in order to save space; each IFU is 51″ on a side. The open squares on the outsides of the array denote inactive IFUs; the open rectangle in the middle shows space occupied by other HET instruments. The streak seen in five of the IFUs (037, 047, 085, 095, and 104, where the first two digits represent horizontal numbering and the last digit gives vertical position) is from a moving object, most likely an asteroid or satellite. At the time the frame was taken, 71 IFUs were operational. Right: The same science observation showing the locations of the IFUs on sky. In this image north is up and east to the left; the IFUs are arrayed on a 100″ grid. The red circle has a diameter of 18′. Note that the IFUs are oriented in the direction of the parallactic angle and will therefore change with the azimuth of observation. The moving object is now obvious.

There is a trade between survey depth, survey area, and survey duration. The chosen values for the survey volume (10.9 Gpc^3) and depth (discussed below) mean that the contributions to the correlation function uncertainties from cosmic variance and shot noise are equal. For a specified survey duration, going deeper would decrease shot noise while increasing noise from cosmic variance. A more shallow survey would have the inverse effect. The survey duration is a more subjective choice and is based on our top-level goal of measuring the $z = 2.4$ values of H and D_A to an accuracy of 0.9% and 0.8%, respectively. One obvious contingency is to increase the survey duration, and therefore the survey footprint, if needed.

Number of sources: The accuracy and precision of the galaxy power spectrum depends on the number of LAEs detected, the false-positive rate, and the fraction of misclassified sources. We distinguish between false positives and galaxy misclassifications, since these two errors have different effects on the correlation analysis. For false positives, we are referring to noise or pixel defects that manifest themselves as an apparent emission line. This error should primarily produce white noise and thus lower the signal-to-noise of our measurement. (This assumption will be thoroughly tested.) The misclassification of galaxies is a larger issue, especially when [O II] emitters are designated as LAEs. In this case, the clustering signal of the [O II] galaxies will leave an imprint on the clustering of the LAEs.

The translation of the galaxy power spectrum into cosmological distance estimates further depends on how the LAEs represent the large-scale clustering of the underlying dark matter distribution. This factor, which is known as the galaxy bias, is a physical quantity that we do not have control over, whereas the other properties depend on our observations and

software. For the following analysis, we assume that, for our redshift range of $1.88 < z < 3.52$, the linear bias parameter is between 1.8 and 2.2 (Gawiser et al. 2007; Guaita et al. 2010); the higher the bias, the more accurate the cosmological measurement, as the power spectrum signal is proportional to the bias squared, while the Poisson shot noise stays constant. For the volume described above, the HETDEX goal is to identify 1.1 million LAEs with a false-positive rate of less than 10% and a misclassification rate (of foreground [O II] galaxies as LAEs) of less than 2%. Since the Poisson shot noise is given by the LAE number density within the whole survey volume (instead of the physical number density of LAEs), a finite amount of observing time translates into a trade-off between depth and area (Chiang et al. 2013). In terms of statistics, this trade-off is equivalent to the limitations imposed by Poisson shot noise and cosmic variance.

A sample of 1.1 million sources over 10.9 Gpc^3 provides a density of 1.1×10^5 galaxies Gpc^{-3}. This density optimizes the trade-off between cosmic variance and shot noise for measurements of the clustering strength, as discussed in Chiang et al. (2013). The specification of a 10% false-positive rate keeps the white noise effect below the statistical limits of our measurements; the stringent specification of the 2% contamination limit minimizes the imprint of the [O II] galaxy clustering signal onto the clustering signal of LAEs (e.g., Pullen et al. 2016; Leung et al. 2017; Addison et al. 2019; Grasshorn Gebhardt et al. 2019). Farrow et al. (2021) discuss the implications of having larger or smaller contamination fractions and redshift-dependent contamination fractions. Again, this 2% limit is designed to keep the systematic uncertainties below the statistical uncertainties.

Minimum spatial scale: HETDEX uses the full power spectrum and does not rely solely on measuring the scale of the

baryonic acoustic oscillations (Shoji et al. 2009). This means that in order to reach its cosmological specifications HETDEX must probe down to scales of $5\,h^{-1}$ Mpc. Calibration down to these scales must include the significant nonlinear effects. Jeong & Komatsu (2006, 2009), and McCullagh et al. (2016) show that we can use scales below $5\,h^{-1}$ Mpc in our analysis, and we utilize these numerical studies. Additionally, on the observational side, the window function has to be accurately determined within the focal plane on similar scales. Section 6.17 outlines our ability to reach the required specification on the accuracy of the window function on these small scales.

Wavelength and redshift accuracy: Errors on the redshifts can impact the measurement of large-scale clustering by washing out redshift space distortions, which are a powerful tool for cosmological studies. Lyα-based redshifts have both a systematic offset and a random scatter about the true systemic redshift of the galaxy. This is due to the physics of radiative transfer, and both the offset and scatter have amplitudes of about $200\,\mathrm{km\,s^{-1}}$ (e.g., Shapley et al. 2003; Shibuya et al. 2014; Trainor et al. 2015; Byrohl et al. 2019; Muzahid et al. 2020; Gurung-López et al. 2021). Since we use redshift space distortion in cosmological studies, any smearing of the redshifts, either from physical or systematic effects, will affect our results. Thus, to avoid any further increase in the uncertainty of our redshift determinations, we require that the precision of our redshift measurements be less than $180\,\mathrm{km\,s^{-1}}$. We note that HETDEX redshifts are significantly more accurate than this, with typical uncertainties below $100\,\mathrm{km\,s^{-1}}$. The precision of these measurements is discussed in Section 6.10.

Flux limit accuracy: Any study of the galaxy power spectrum requires measuring the effect that observational selection has on the observed distribution of sources. Thus one needs to know the flux limit versus wavelength at each location in the survey. The required accuracy for these flux limits can be estimated using the expected number of sources per field. The requirement is to not have the flux limit uncertainty be larger than the uncertainty arising from Poissonian errors. For each HETDEX observation we expect about 200 LAEs, which translates into a 7% variation from the Poissonian noise. We note that this uncertainty limit is for a fully populated IFU array and is larger for a focal plane containing fewer IFUs. Thus, we set a limit of 5% on the flux limit, averaged over the whole focal plane, for one observation; at this level, the uncertainty on the flux limit would only modestly increase the uncertainty on the expected number of sources per field. This 5% limit translates then directly into the requirements for the measurement of system throughput. The flux limit and throughput accuracy are discussed in Section 6.16 and Section 6.17. We show that our precision on the throughput measurement and therefore the flux limit is, on average, around 2%. Thus, we are meeting the requirements for the flux limit accuracy.

Dither accuracy: The emission-line detection algorithm relies heavily upon having an accurate knowledge of an observation's point-spread function (PSF), and this model can only be determined from a precise measurement of each frame's dither position. The positions of the dither offsets are also important for a proper uniform sampling of the sky. Our specification on the accuracy on a dither position is $0\overset{''}{.}4$. This accuracy is discussed in Section 5.

Astrometric accuracy: Knowledge of the astrometric accuracy of the HETDEX frames is important for matching emission-line spectra with target lists produced by imaging surveys. Such matchings are used to help discriminate Lyα emission of a high-z source from the [O II] flux of a foreground galaxy. The astrometric accuracy of individual sources is discussed in Section 6.14 and is much better than our specification of $0\overset{''}{.}5$.

Imaging survey: HETDEX does not need to select targets beforehand; all the objects within our survey's footprint are observed. Consequently, we do not need an imaging survey to identify high-z galaxies; instead we require images to assist with line identification. Over our $2000\,\text{Å}$ spectral range, most HETDEX sources have just a single emission line, produced primarily either by Lyα or [O II] $\lambda 3727$. (There are other features that may appear in the spectra but these are the dominant lines.) Furthermore, the low resolving power of the VIRUS units does not allow us to split the [O II] doublet nor resolve the skewed line profile common to Lyα (Runnholm et al. 2021). Thus, without additional information, these two lines can be confused, thereby imprinting the power spectrum of foreground [O II] sources on top of that of the LAEs (and vice versa).

By measuring the continuum of an object, using either its HETDEX spectrum or its flux in a broadband image, we can estimate an emission-line equivalent width. This single piece of additional information is extremely useful for helping to discriminate the unresolved [O II] emission of a foreground galaxy from unresolved Lyα at high-z (e.g., Rhoads et al. 2000; Gronwall et al. 2007; Leung et al. 2017). The rest-frame equivalent width distribution for Lyα at $z \gtrsim 2$ is quite different from that of nearby [O II] galaxies, and the $(1 + z)$ boosting that occurs in the observer's frame greatly increases this offset (e.g., Gronwall et al. 2007; Ciardullo et al. 2013). As a result, a comparison of the emission-line strength to the continuum flux density can produce a clean separation of the two lines. For this calculation, we do not remove the contribution of emission lines to our broadband flux density measurements, since their effect is generally small and our calibration of the LAE/[O II] galaxy discriminant is empirical. For the uncertainty in the continuum measurement to not dominate that of the HETDEX emission line, the reference broadband images must reach a limiting magnitude of $g \sim 25$.

The imaging surveys and their flux limits are presented in D. Davis et al. (2021 in preparation) and will be described in detail there. The imaging comes from a variety of sources, including our own observations with HyperSuprimeCam (HSC) on the Subaru telescope and the Mosaic II camera on the Mayall 4 m telescope, and archival data from the Wide Field Camera 3 and Advanced Camera for Surveys imagers of the Hubble Space Telescope (Koekemoer et al. 2011), the DES (DES Collaboration et al. 2021), the Dark Energy Camera Legacy Survey (DECaLS; Dey et al. 2019), the Canada–France–Hawaii Telescope Legacy Survey (CFTHLS; Coupon et al. 2009), and the SDSS (York et al. 2000; Abazajian et al. 2009). When multiple imaging surveys are available for the same source, we use the data that give us the best depth and image quality. For the vast majority of fields within HETDEX, at least one of the images reaches our specification of $g = 25$. The HETDEX spectra themselves, when collapsed over the g-band region, typically reach about $g = 24$; while this is not quite deep enough for our flux-density requirements, it is useful as a comparison benchmark for the products of the imaging
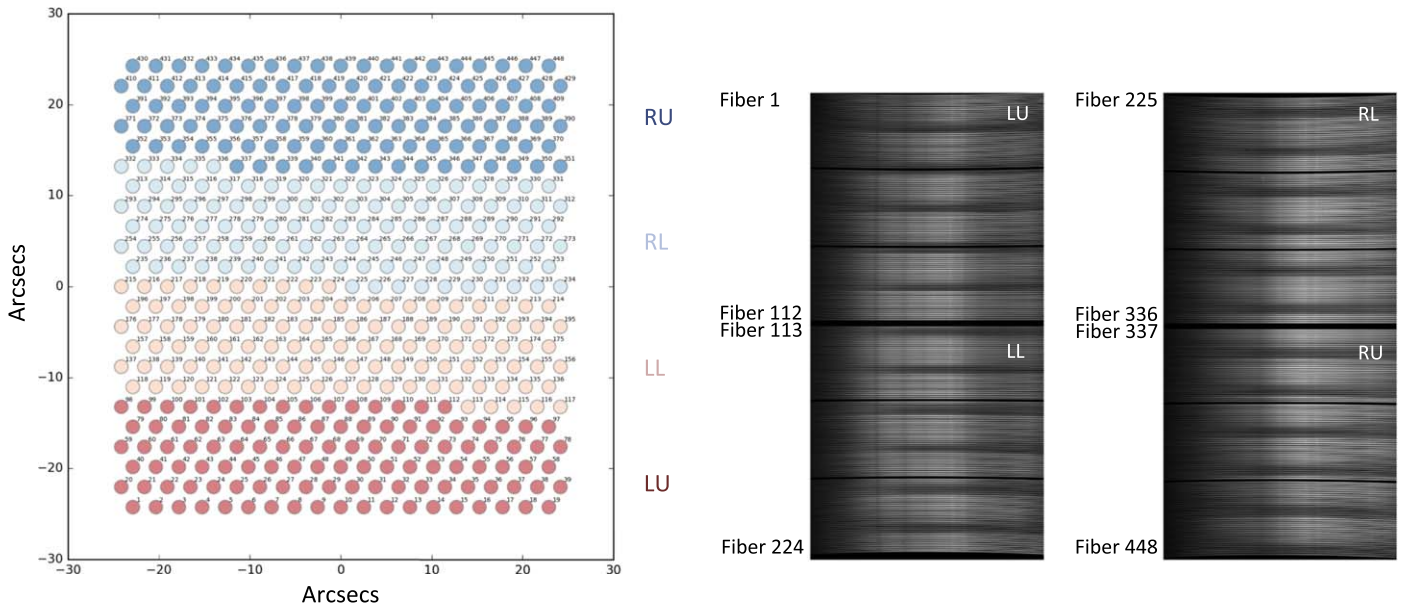
Figure 4. The layout of the fibers associated with each IFU. The left-hand panel displays the nominal fiber arrangement in the focal plane and their corresponding detectors and amplifiers. The fibers are numbered 1 to 448. This image represents the standard alignment, though there are a handful of IFUs with different layouts. The right-hand panel shows the layout of the spectra from the fibers on the detectors. The horizontal is the wavelength direction, and the vertical is the fiber direction. Each spectrograph is split into two sides, each with its own detector, "L" and "R", and each detector (or side) has two amplifiers, designated "U" and "L".

surveys and allows us to confirm that the imaging data have a common photometric zero-point.

Setup time: In order to reach specifications for observing efficiency, the instrument setup time, which we define as the interval between the end of one 3 dither sequence and the start of the next, must be no more than 4 min. As shown in Hill et al. (2021, submitted), HETDEX observations are close to this specification.

By meeting the science requirements above in an observing program consisting of about 460,000 IFU observations (baselined originally with over 6000 observations), HETDEX expects to produce a combined distance measure (i.e., a spherically averaged distance, typically called D_V) of 0.8% in the $1.88 < z < 3.52$ universe. Given normal weather statistics, instrument stability, and expected observing time allocations, the survey is expected to be complete in 2024.

4. IFUs and Detector Designations

The focal surface of the Hobby–Eberly Telescope is capable of housing a grid of 78 IFU fiber arrays, each with 448 fibers covering $51'' \times 51''$ on the sky. (References for the telescope and instrument can be found in Hill et al. 2021, submitted.) To reach the nominal specification for HETDEX, we require that 74 of these potential IFUs be operational. The individual IFUs are designated by a three-digit `ifuslot` in the focal surface, with the first two digits representing the column number and the last digit defining the row (see Figure 3). Except for the slots near the center of the focal surface, which are reserved for use by other HET instruments, the center of each IFU is separated from that of its nearest neighbor by $100''$, thus leaving $49''$ of unused space between the fiber bundles. As a result, in a fully populated focal plane, the VIRUS IFUs take up $\sim22\%$ of a circle where the diameter of the circle is defined by the corners of the outermost IFUs. (The actual HET focal surface is larger than this, as the telescope's guide cameras extend over a region further out.)

Each IFU in the above array feeds its own spectrograph unit, which is designated by a `specid`. Each unit has two spectral channels or sides, each with its own detector, designated "L" and "R". Finally, each detector (or side) has two amplifiers, designated "U" and "L". Thus, during each exposure, four separate detector images are generated by each VIRUS spectrograph; for example, a file containing the sequence 073RU is from the "U"-amplifier of the "R"-side CCD of the spectrograph fed by the IFU in column 7 and row 3. A single VIRUS exposure with 74 IFUs generates $74 \times 4 = 296$ data files. Figure 4 displays the nominal layout of the fibers for each IFU, along with their CCD and amplifier. A few IFUs have slightly different alignments, which we handle on an individual basis. The fiber numbers run from 1 to 448 and are labeled in the figure.

5. Dither Sequence

Since the center of each $1''\!.5$ diameter fiber is separated from its nearest neighbor by $2''\!.543$, a dither sequence is needed to fill the gaps between fibers. These dither offsets are performed by shifting the fiducial position of a star on the focal surface guide camera, thereby forcing the telescope to move accordingly. Each HETDEX field is observed at three different positions, for a total exposure time of 18 min. The commanded dither pattern is triangular: dither 2 is offset from dither 1 by $1''\!.27$ in x and $0''\!.73$ in y, while dither 3 is offset from dither 2 by $1''\!.46$ in y. The resultant overlap in dither positions is thus very small, about 2% in area. This dither sequence provides complete spatial coverage over the $51'' \times 51''$ region of each fiber array.

Figure 5 demonstrates the consistency of the dither pattern by illustrating the dither positions for all HETDEX data sets taken between 2020 January 1 and 2020 June 30. To create the figure, the locations of every continuum source on the VIRUS frames are compared to those derived for the observation's other dithers. Over the entire sample, the mean difference between the commanded and measured dither position is 6%, i.e., the mean offset is $0''\!.08$ smaller than the expected value of $1''\!.47$. After accounting for this scale difference, the scatter of
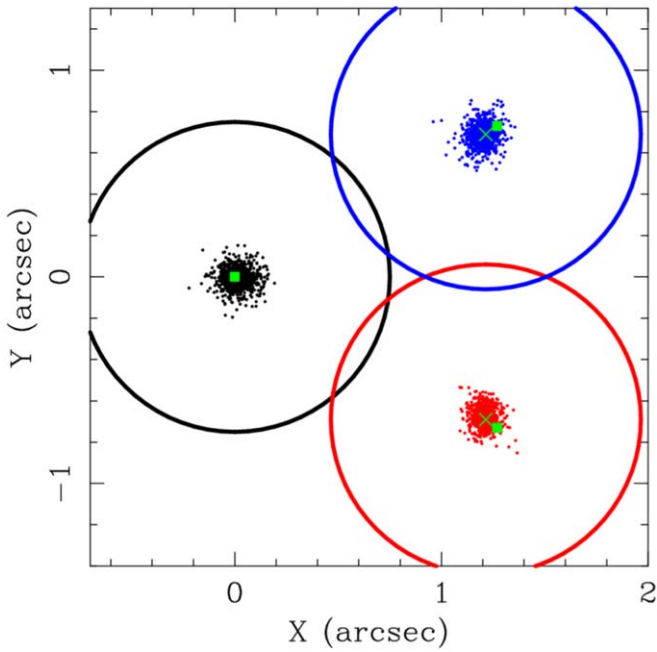
Figure 5. The measured locations of the dither offsets derived from all continuum sources in fields observed between January 1 and 2020 June 30. The large black, blue, and red circles represent the size of a fiber and the assumed position in the standard three-point dither. The black, red, and blue points are the measured positions for dithers 1, 2, and 3, respectively, where we have centered the three dithers on their average position. The light green crosses show the centroids of dithers 2 and 3, while the light green squares illustrate the desired center; the offset between these positions is likely due to a small difference in plate scale. For HETDEX reductions, it is not important that the commanded and actual dither positions match, but it is critical that we know what the actual dither position is. After accounting for an overall scale factor, the measured offsets are consistent with the requested centers, and the measured scatter of the dither centers is the same as our measurement accuracy. This agreement allows us to use the requested dither offsets in our data reduction; the measured dither positions are only used as a check on the data.

the measured positions, $0.''15$, becomes equal to the measurement uncertainty of the individual dithers, which is also $0.''15$. This consistency implies that the true position of the telescope can be measured to a similar accuracy. All subsequent data processing uses the commanded dither offsets while accounting for the small reduction in scale. Thus, dither 2 is assumed to be offset from dither 1 by $1.''215$ in x and $0.''70$ in y, and dither 3 is considered to be offset from dither 2 by $1.''40$ in y.

The large circles in Figure 5 represent the size of the fibers and the location for the assumed three-point dither sequence. Ideally, there would be no overlap, but the dither sequence shows a few percent overlap. This small overlap is taken into account in our flux calibrations.

6. Basic reductions

As mentioned above, the HET can only observe objects that are 35° from the zenith. The telescope has a 11.1 m spherically shaped primary mirror, which consists of 91 separate hexagonal segments. To observe an object, the telescope structure rotates to the appropriate azimuth, and the tracker at the top end of the telescope follows the object in x, y, and z as it moves across the primary mirror's 12° diameter field of view. To correct for the primary mirror's spherical aberration, this tracker contains a wide-field corrector (WFC) capable of improving image quality over a 22′ diameter field of view.

Since the WFC's pupil diameter is 10 m, the VIRUS fibers see different sets of mirror segments at different times during a track. Moreover, for long exposures, the effective aperture of the telescope changes, especially near the ends of the track, since part of the WFC's pupil falls off the projection of the primary. More details about the tracker and WFC can be found in Hill et al. (2021, submitted).

For HETDEX, we need to take into account differences in mirror illumination for all exposures and use observations that go to the edge of the track. This means we need to measure the effective integrated throughput during each exposure. Similarly, there are also slight changes in the telescope plus instrumental resolution in the spectral direction. As discussed in Section 7.4, we compensate for this by not using a fix line width in the emission-line detection algorithm. Thus small changes in the spectral shape do not affect our ability to detect emission lines.

Because of the HET's dynamic nature, the calibration of its data products is tricky. HETDEX has a complicated data model, as it is important to calibrate all of its ~33,000 fibers to an accuracy of a few percent and to keep the residuals associated with the sky subtraction to less than 3%. The latter systematic is especially important, since the primary targets of HETDEX have line fluxes well below that of the sky. For example, a typical $z \gtrsim 2$ galaxy may have the flux from its Lyα emission line distributed over ~150 pixels obtained from ~5 different fibers associated with the observation's three dithers. Since a detection signal-to-noise of 5 corresponds to a total count level of around 300 ADU spread over about 100 individual detector pixels (three dithers, and including the spectral and fiber direction on the detector), this means controlling pixel level issues to a precision below one count per detector pixel. For comparison, the background counts on a dark night range from 10 to 30 per detector pixel away from sky lines. Below, we describe all aspects of the spectral reductions including overscan removal, bias subtraction, flat-fielding, sky subtraction, flux calibration, astrometric calibration, and object detections for both continuum and emission line sources.

6.1. Data Processing Requirements

All HETDEX reductions and data storage use resources at the Texas Advanced Computing Center (TACC). The raw data coming off the telescope are transferred from the HET to the TACC as soon as the CCDs are read out. While quick-look quality checks are performed at the telescope upon readout, the primary data reductions are run on a monthly cadence on the TACC computers. The CPU resources are significant, but given the excellent facilities at TACC, the project can process 4 yr of observations in 1–2 weeks using ~200 processors simultaneously. A full processing of the data is generally performed multiple times, as with each iteration our understanding of the behavior of the detectors improve. Although nearly all aspects of the reduction scripts on TACC are automated, the large number of files associated with HETDEX reductions, the wide variety of instrumental differences across the spectrographs, and the lack of stability in some of the first generation detectors mean that a significant amount of individual attention is needed throughout the process.

6.2. Reductions to Sky Subtraction

Basic detector characterization uses pixel flats, bias frames, dark frames, twilight sky exposures, and the sky background of science exposures. The twilight frames provide the primary

calibration for the fiber profile, spatial trace maps, spectral trace maps, fiber-to-fiber normalizations within a given IFU, and IFU-to-IFU normalizations across the focal surface. (These frames also allow us to measure the resolving power of the instrument.) This information is then slightly modified using the data of each individual science frame: since for a typical observation, about 50%–70% of the fibers are looking at blank sky, these on-sky data allow us to make small adjustments to the calibration.

The VIRUS spectrographs are mounted on the side of the HET and are very stable, with calibrations that do not change significantly over many months to years (Hill et al. 2021, submitted). Given this stability, calibrations frames averaged over a month are superior to those taken daily. Moreover, the tight packing of fibers on the CCDs means that there is overlap in the fiber profiles at the level of a few to 10%, and being able to measure this well requires a large number of data sets. As a result, individual calibrations are not as robust as monthly averages; this improvement is noticeable on the sky-subtracted frames.

The sequence of reduction steps for the twilight sky frames is overscan subtraction, bias subtraction, pixel flat correction, background light model subtraction, measurement of the fiber traces, measurement of the fiber profiles, derivation of the relative wavelength solution, fiber extraction, and the assignment of the full wavelength solution. The reduction order for nighttime science frames is overscan subtraction, bias subtraction, pixel flat correction, background light model subtraction, adjustment of the fiber trace position, fiber extraction, adjustment of the full wavelength solution, sky measurement, sky subtraction in both the 2D and extracted 1D spectra, and flux calibration.

6.3. Bias Frames

The HETDEX project takes 11 bias frames every day, or about 330 bias frames per month. An interesting feature of these frames is the presence of a low-level, temperature-dependent interference pattern in some of the detectors. Left unaccounted for, this pattern can create false positives for emission line detections.

Unfortunately, the interference pattern is not stable enough to be removed via bias frames taken at different times of the night. Thus, our master bias must preserve the broad-scale features present on the individual biases but not include the transient small-scale interference patterns. To do this, we smooth over the bias pattern on the individual biases using a 1×9 pixel boxcar average. The dimensions are designed to not mix the bias from individual columns (i.e., the $1\times$) and to remove the interference pattern (i.e., the $9\times$). Then, on every detector readout throughout the night, the interference pattern (if any) can be measured and recorded. If the pattern is present, the 1 to 2 count noise increase associated with its presence can be incorporated into the analysis. We also keep track of the bias patterns and trace their behavior with time. Since these patterns reflect the matched pairing of a controller and a detector, any change to the pattern may indicate a failure in the controller. There are some detector amplifiers that have an unstable interference pattern, and these must be removed from the analysis.

6.4. Dark Current

Our array of 74×4 amplifiers have a range of dark currents. Although we take dark exposures every day, the daily variations are large enough to preclude the use of daytime darks for nighttime observations. Instead, the daily darks

provide quality checks, which alert us to issues with individual amplifiers. We then fold the dark frame information into the background light analysis discussed below.

6.5. Pixel Flats

Given that a single weak emission line may be spread over about 150 pixels on three dithered frames, we require excellent knowledge of the response of each individual pixel. Thus an important aspect of data reduction is the application of pixel flats for each detector. The initial step in deriving these flats is to examine the high signal-to-noise flatfield frames acquired in the lab before the CCDs are installed at the telescope. These detector flats cannot be used on their own, since the devices are temperature cycled before installation, but the lab flats are useful for identifying the locations of hot pixels on the CCDs.

Pixel flats generated with the instrument on the telescope are the most important component for the flat-field correction. These are produced using defocused spectra, as it is important to illuminate those pixels in between the fibers on the detector and still maintain the spectral dispersion. To defocus the light, we use a set of spacers that increase the separation of the IFU head attachment and the spectrograph. We then take a set of images from a laser-driven light source (LDLS) with an integrating sphere. This setup allows light to easily reach into the fiber gaps, enabling the creation of an accurate set of flatfield frames. From start to finish, this procedure takes a few hours of daylight time per detector. Since this is a time-consuming process, we only perform these observations upon detector installation and once every 12 months thereafter. We have looked carefully at pixel flats taken over three years of operation, and the flats are remarkably stable, to better than 0.1% on average of the pixel flat value. When significant changes are found, they are all traceable to instrument maintenance, and we monitor these changes with new flats.

We reduce the flatfield frames by first dividing each row by a smoothing spline and then repeating the procedure for each column. The resultant frame residuals are then examined for pixels lying more than 3σ above or below the predictions of the spline; when such pixels are found, they are masked and a new spline is generated. This process is repeated until convergence is achieved for all the individual pixels. The result is a highly accurate pixel flat along with a variance frame, as determined from the individual exposures. In general, the uncertainties associated with our pixel flats are below the 1% level.

Many VIRUS CCDs have significant features, including large dust spots, many charge traps, and a "pox" contamination where the quantum efficiency of individual pixels can be suppressed by 10%–40%. These issues were quite common on the first generation of detectors and are still present in a few of the later units. This "pox" tends to be located on the corners of the detectors and is particularly difficult to deal with. The HETDEX project has been removing badly affected detectors and slowly building a set of CCDs that do not have significant pox. While we will never have a completely pox-free data set, the worst units are being addressed. We have run extensive simulations for object detections, including regions affected by the pox, and, as expected, the pox regions have a higher flux limit.

The pixel flats come from the LDLS, which provides high-count level observations and correspondingly high signal-to-noise data. However, the nighttime science data are always in the low-count regime, and flats generated with low-light levels do not exactly match the bright-light LDLS flats, especially for

pixels with relative throughput values below 0.6. We therefore flag the low-throughput pixels and do not use them in subsequent analyses. Including cosmic rays and flagged pixels, we generally exclude about 3% of the pixels in any exposure.

6.6. Background Light

There is extra light in most detectors that needs to be modeled. This background light has contributions from the extended wings of the PSF, scattered light, errors in modeling the wings of the fiber profiles, bias counts not included in the master biases, and controller issues. Because these effects involve a mixture of additive and multiplicative sources, their individual contributions are not easily modeled, and we do not attempt to measure the relative importance of each component. Instead, we rely on an empirical approach that combines all the effects. This procedure is not exact, but it allows us to reduce the background subtraction residuals to below a few percent. If one requires background light removed at a level below this, then an additional correction is likely needed.

Our background modeling uses all the twilight sky and nighttime science exposures for a given month; this sums to roughly 500 twilight frames and ∼1000 night sky observations. After subtracting the overscan and bias from each frame, we locate regions on the detector that should not contain any light from fibers: these are gaps at the bottom and top edges of the detector and two or three gaps in the middle. We measure the light in the gaps as a function of the wavelength for every exposure. This background light allows us to calculate a full-frame background model via interpolation. Thus, every science exposure and twilight exposure provides information for the background light model.

The background model is not linear with the count level. To deal with the nonlinearity, we use three different flux levels for the science and twilight frames. For each level, we compute the biweight average (Beers et al. 1990) from the hundreds of frames taken that month. This gives us a set of three background models and their corresponding flux levels for each detector.

To apply this correction, we find the average count level in the science frame using a specified region of the chip. We then interpolate this count level using the three background models to determine the appropriate background to apply. For a typical science exposure, the background light is about 2 counts per pixel, with a variation that depends on the sky brightness and the brightness of nearby continuum sources.

The top left panel of Figure 6 shows the month-to-month variation in the background count level of the models for IFUslot 046. The colors correspond to the four different amplifiers used in this IFU.

Our background models only provide a full-frame correction and do not account for scattered light that is local. Photons from bright objects, such as stars, galaxies, and meteors, can scatter off a variety of surfaces, such as the IFU plate, the optical and mechanical elements of the telescope, and the optics of the spectrographs themselves. We see evidence for all three types of scattering in the data, and there are likely more. We do not apply a local scattered light model in the current data, and it is a subject to be explored in future analyses.

The loss of light due to scattering is automatically included in the flux calibration. If a bright continuum source creates excessive scattering over the whole detector, that detector is removed from consideration. For the emission line detections, we exclude from consideration those sources close to bright objects. Since our focus for the cosmology is high-redshift objects, we only remove sources that happen to be projected in the same region of sky as a bright source. (For science focused on nearby galaxies, one should rely on a separate catalog for emission lines near bright sources.) These bright sources will create holes in the survey, and these are included in the window function. For calibration, the brightest sources are not used since they exceed the acceptance criteria for extraction from 2D to 1D spectra. Thus, while scattered light from local sources is not included in the analysis, it does not create problems for the calibration of the spectra or the detection of objects.

6.7. Fiber Trace

In order to perform our spectral extractions, it is crucial that the fiber traces be as accurate as possible, and that they routinely reach a precision of 0.02 pixels in their centering. To achieve this level of accuracy, we rely on exposures of the twilight sky. As a first pass for the fiber trace, we use the initial measures for the fiber positions, which were determined when the spectrographs were first installed on the telescope. We then use the twilight sky data to compute the location of the peak flux of each fiber in 10 spectral bins by weighting the summed-pixel positions in the spatial direction by the square of the total flux. This calculation begins at one end of the spectrum and then increments to the next spectral bin, using the previous bin's centroid as a starting point. When the full detector field has been measured, we fit the resultant array of fiber centroids versus wavelength with a smooth spline. Finally, using this initial fiber trace as a starting point, we repeat the procedure, this time weighting the points by the fiber profile itself, instead of the simple square of the flux weighting. This iterative method is extremely accurate: based on the measured wavelength-to-wavelength variations, which we track over periods of months, the centroids of our fiber traces are known to a precision of ∼0.02 pixels.

Although the shapes of fiber traces are extremely stable, there are nightly shifts in their position: the trace positions have zero-point offsets and a breathing mode, where the fiber separations expand and contract with time. To handle these variations, we first compute the biweight average of all twilight trace positions taken over a month, assuming that each fiber trace is the same except for a possible zero-point offset in its position. Then, for each individual nighttime exposure, we take these fiducial fiber positions and adjust the traces with a linear model. (In other words, we apply a zero-point offset and a scale factor.) These nightly motions can be up to 0.1 pixel in amplitude and must be included for a proper sky subtraction.

The middle left panel of Figure 6 shows the evolution with time of the positional difference between the first fiber and the last (112th) fiber on one of the amplifiers for a CCD in IFUslot 046.

6.8. Fiber Profile

A key element required to achieve an accurate sky subtraction is knowledge of the fiber profiles. We rely on the twilight images exclusively for this information. The shape of a typical profile is flat-topped with steep wings, and, despite many attempts, we were unable to robustly describe the profiles analytically with a limited number of parameters. Thus, we model the fiber profiles in both the spectral (along the fiber) and spatial (across the fiber) directions without having to rely on an
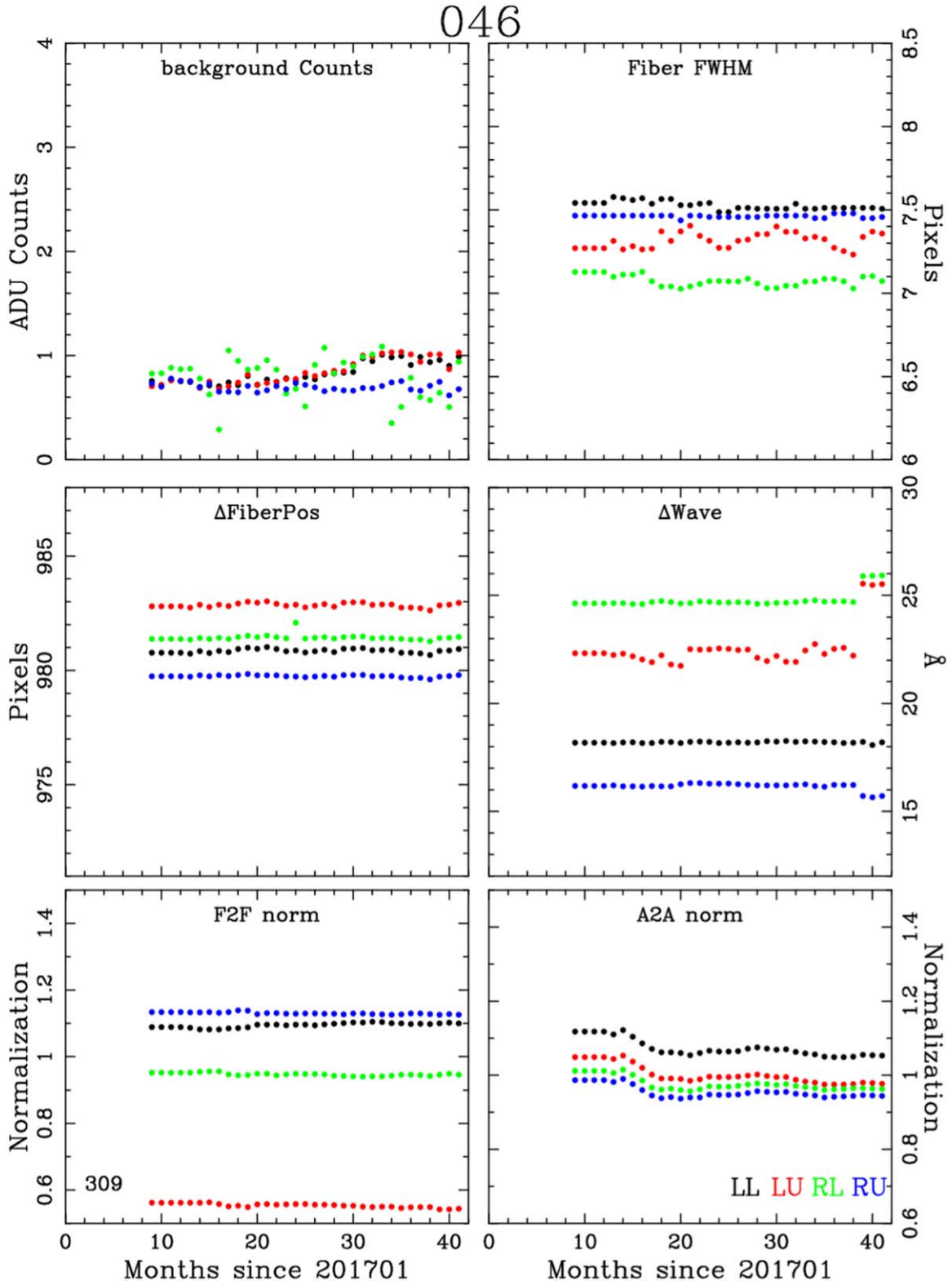
Figure 6. Variations in the calibration of one typical IFU over three years of operation. Plots like these enable us to look for trends in the calibration of all IFUs in the VIRUS array. This set of data is for IFUslot 046, which feeds spectrograph 309. The six checks we apply are: background count level (top left), FWHM in pixels across the fiber profile (top right), separation in pixels between the first and last (112th) fiber on an amplifier (middle left), change in the wavelength solution (in Å) from the corner fibers of the IFU (middle right), normalization differences between the ratio of the corner fibers of the IFU (bottom left), and the overall normalization of the particular amplifier compared to the average of all the amplifiers in all the spectrographs (bottom right). The colors represent the four amplifiers of the IFU, with LL in black, LU in red, RL in green, and RU in blue.

analytic form. The most robust approach is to fit each profile nonparametrically using 13 spatial bins across the fiber. At each spectral location along the fiber, we use our knowledge of the fiber trace to fit the fiber profile with a smooth spline (Wahba 2019) using a 10 pixel bin size in the wavelength direction. After performing ∼100 of these fits along the fiber, we interpolate through the individual spectral bins to create a set of 13×1032 entries for each fiber. The 1032 is the number of pixels on the detector array in the wavelength direction. Doing this for all fibers provides the full frame fiber profile, with $13 \times 1032 \times 112$ entries per amplifier. For further refinement, we use the residuals from these fits and fit a spline to the residuals, adding the results back into the models. We then iterate our fiber profile with the fiber trace and repeat both measurements. The results typically converge in two iterations, but we perform five iterations for all solutions. This procedure is done on all twilight frames and all amplifiers for a month. Finally, we compute the biweight of the hundreds of twilight frame fiber profiles to provide the month's master fiber profile.

We note that because of the high density of fibers on the CCDs, there is overlap of the fiber profiles between adjacent fibers. We do not explicitly take this overlap into account when deriving the fiber profile. Instead, we rely on the background model to correct for the fiber-to-fiber contamination. The overlap in counts is typically 2%–3%.

The top right panel of Figure 6 shows the variation in fiber FWHM in pixels for IFUslot 046 at a single location on the detector.

6.9. Fiber Extraction

The fiber extraction algorithm uses a weighted sum across each wavelength bin. Using the fiber trace position as the centroid and the fiber profile as the weights, we perform an optimal spectral extraction as outlined by Horne (1986). For each resolution element, we also track the rms and the χ^2 deviation from the fiber profile. These values are used directly in the object detection algorithms. Most importantly, the reduced χ^2 measurements allow us to accurately remove cosmic rays. Any resolution element with $\chi^2 > 10$ is assumed to be affected by a cosmic ray hit and is flagged for subsequent removal in the analysis. Furthermore, the χ^2 value provides a method for the identification of low-level charge traps. The VIRUS detectors suffer from a number of such traps, which, if not identified as such, can result in false emission-line detections. The sum of the χ^2 values over all the data provides a clean way to find these defects.

6.10. Wavelength Solution

Traditionally, one would use arc-lamp exposures to determine both the wavelength solution and the spectral resolution of an instrument. In our case, the integrating sphere has both vignetting and spectral features that create differences between the wavelength solutions based on arc lamps and solutions produced by on-sky calibrations. As a result, arc-lamp calibrations cannot attain the accuracy required to reach the HETDEX specifications. We therefore rely on the twilight sky to define the wavelength scale.

The wavelength calibration of HETDEX spectra is performed in two steps. First, we adjust each fiber to a local calibration for its individual amplifier. Second, we fit all the spectra over the full CCD to a standard calibration, thereby placing the full set of fibers on the same system.

We begin by defining the sum of the central three fibers on an amplifier as the fiducial spectrum from the frame. Each extracted fiber spectrum is compared to this fiducial over 10 equally spaced wavelength bins between 3500 and 5500 Å in the twilight sky. To determine the wavelength offset between each fiber and the fiducial, we use the full wavelength range in those 10 bins and find the offset that gives the minimum rms between the two spectra. These wavelength bins define a 2D plane of wavelength offsets for all 112 fiber spectra on the amplifier. We then interpolate this low-resolution map onto individual 2 Å wavelength bins. For the edges of the detectors, we perform an extrapolation based on the slope derived from the last few bins. We create this map for all twilight frames taken over a month and then average the maps (after adjusting for a zero-point) to create a master relative wavelength solution for the amplifier. Wavelength differences from this master solution, measured on adjacent corners of the amplifier, then give us a measure of the stability of the solutions.

The second step uses each of the 112 fibers to build a highly sampled twilight sky spectrum. We compare this spectrum to that of the Kurucz et al. (1984) KPNO solar spectrum (convolved down to the resolution of the VIRUS spectrographs) by fitting the wavelength offsets of the 10 bins and interpolating those solutions onto each 2 Å pixel. This results in residuals to the solar atlas that are roughly 0.3 Å in amplitude. The wavelength solution is similar for all the spectrographs; it is very linear between 3500 and 4800 Å before experiencing significant curvature at the reddest end of our spectral coverage. There are differences between the solar atlas and the twilight sky, but the solar light dominates enough that the comparison is robust. Thus, each observation for a night is tied to that particular evening's twilight.

We routinely check whether the wavelength solutions obtained from the twilight exposures adequately represent those for the nighttime observations by inspecting the sky-subtracted images for unexplained residuals. We observe no systematic residuals over all amplifiers. In addition, as an end-to-end test, we compare the radial velocities derived for HETDEX field stars with higher-precision velocities obtained from the Sloan Extension for Galactic Understanding and Exploration (Yanny et al. 2009) and Large Sky Area Multi-object Fibre Spectroscopic Telescope (LAMOST; Luo et al. 2015; Xiang et al. 2017) surveys. As shown by Hawkins et al. (2021), the resultant $30 \, \mathrm{km \, s^{-1}}$ rms of the VIRUS spectra is within our specifications, as are the small systematic offsets compared to LAMOST ($\sim 13 \, \mathrm{km \, s^{-1}}$) and SDSS (a few $\mathrm{km \, s^{-1}}$).

We do not apply a heliocentric correction to the final spectral database, although that information is provided in Hawkins et al. (2021). However, that correction will be included in the catalog paper of M. Cooper et al. (2021, in preparation).

The middle right panel of Figure 6 shows how the absolute value of the wavelength difference between the corners of the 112 fibers for IFUslot 046 changes with time. For this data set, there is a jump in early 2020, which corresponds to when the spectrograph was taken off the the telescope and then returned. This jump is included in the data reductions. We note that these twilight calibrations are used as starting points for the nightly reductions, and we refine each science frame calibration as small modifications to the twilight calibration. Thus, small changes such as those seen in the wavelength shift are incorporated.

6.11. Instrumental Resolution

There are two aspects for the instrumental resolution that we consider: the line width in the spectral direction and the source profile in the spatial direction. In detector pixels, these should be similar, but we note that the focus in the spectral direction tends to be slightly better than in the fiber direction. We discuss both of these and their implications.

The line width in the spectral direction is not critical for the HETDEX project, as the parameter is not used by our detection algorithms. As outlined below in Section 7, emission-line detections are initially performed using a spectral sum over 3–4 pixels in wavelength and is then refined using a fit where one of the free parameters is the line width. Thus, the instrument's spectral resolution is incorporated into the HETDEX data products via a cataloged line width. If a subsequent study requires knowledge of a source's intrinsic line width, then the instrumental values will need to be considered. Hill et al. (2021, submitted) show that the instrumental resolution of VIRUS is fairly constant as a function of the wavelength, with a FWHM around 4.7 Å, giving a resolving power that varies from 750 to 950, blue to red. There are some spectrographs that have a larger change in the spectral resolving power over the fibers. We track these units for possible refocusing in the future.

More important is the VIRUS instrumental resolution in the spatial direction. Since the fiber packing of each IFU is relatively tight with 10%–20% overlap, poor focus in the spatial direction will cause point sources to spread into neighboring fibers, thereby lowering the signal-to-noise ratio. Our quality control involves examining the FWHM in the spatial direction of every fiber in every IFU. As can be seen in the top right panel of Figure 6, the spatial FWHM for fibers in IFUslot 046 is about 7.5 pixels; the full range of FWHMs for all the IFUs extends from ∼6.5 to ∼8.5 pixels, although some individual fibers can have larger values. We use the cumulative distribution function for the instrumental fiber FWHM for all fibers in an IFU to help determine whether a unit needs upgrading. The fiber profiles are stable with time to within the measurement uncertainty.

6.12. Fiber-to-fiber Flats

The twilight-sky frames also provide a measure of the fiber-to-fiber relative throughput over the full HET field. Once again, we break this measurement into two steps. First, we measure the fiber-to-fiber variation over an individual amplifier, and then we measure the amplifier to amplifier variation over the full field of all IFUs.

In the first step, we use the extracted spectra of every twilight sky exposure. We initially scale each of an amplifier's 112 fiber spectra using the biweight average of the pixel values between 4300 and 4900 Å. We then use the 112 spectra (all of which have slightly different wavelength solutions) to make a single, highly sampled sky spectrum that has 112×1036 elements, and take the biweight average over 11 pixels to make an array with approximately $112 \times 1036/11$ elements. The 1036 comes from the wavelength resampling, which covers 3470–5540 Å in steps of 2 Å. This becomes our fiducial scaling profile. The original individual fiber spectra are then divided by the fiducial to produce a measure of the fiber-to-fiber variation as a function of the wavelength. This procedure is then iterated five times to produce a fiber-to-fiber map for the twilight spectrum under

consideration. Finally, we take the biweight of all the twilight fiber-to-fiber maps created over a month to produce our final fiber-to-fiber map. The bottom left panel of Figure 6 shows the variation in the relative fiber normalization between the first fiber on the blue edge of the amplifier and the 112th fiber on the amplifier's red edge for IFUslot 046.

After normalizing all the fibers with a single amplifier, the next step is to calculate the relative normalization for each amplifier in the array of VIRUS units. In a similar fashion to that described above, we generate a master sky spectrum using the results from all the amplifiers. With a fully populated IFU array, this involves ∼33,000 spectra. The fiber-to-fiber map is divided by the master spectrum of each amplifier to provide the relative amplifier normalization. As above, we iterate the solution using the new normalization and compute our final estimate as a function of the wavelength. The biweight of the amplifier normalizations found from every twilight exposure over a month then gives the master profile.

The bottom right panel of Figure 6 shows the variation in the relative amplifier normalization compared to the average of the full field for IFUslot 046. Our analysis of these ratios suggests that the amplifier normalizations of our older units decrease with time. We attribute this change to better performance from the newer IFUs and possible accumulation of dust on the older spectrographs. Any change over time in the overall performance of the observations is included in the calibrations and source simulations (presented in Section 8).

6.13. Sky and Background Subtraction

One of the more critical reduction steps for the HETDEX program is sky and background light subtraction. At very low signal-to-noise ratios, any residual introduced by sky subtraction will reduce our ability to find faint objects and create a significant increase in the number of false detections. All previous reduction steps are therefore tuned to make background light subtraction as robust as possible over the entire range of detector properties.

The majority of the background light comes from the night sky, but there are other sources to consider, especially when trying to measure down to levels of less than one count per pixel. Some of these other effects include scattered light, dark current, unaccounted for wings in the telescope PSF, unaccounted for wings in the fiber profile, and stray charge.

The fiber extractions produce 1D arrays of flux versus wavelength. After this step, we estimate a local sky for each individual amplifier by using our knowledge of the fiber fluxes to identify and exclude all discrete continuum sources in the fibers. We start with a background light subtraction in an amplifier, where there are 112 fibers. We compute the biweight average over all fibers and flag any source that is more than 3 times the biweight scale from the average as a continuum source. For the remaining fibers, we assume that faint continuum sources may still be hiding within the data. This hypothesis is confirmed via deep HST imaging; for HETDEX the fields with HST overlap, and we find that about 10% of the fibers that made it through the initial continuum cut have faint sources in the HST images. Thus, to correct for this systematic, we also remove 10% of the remaining fibers with the highest count rates. This approximation will likely introduce a very small residual zero-point in the background, and studies that are very sensitive to the background should consider applying an additional contribution to the zero-point. After this last cut,

the biweight average is used to determine the sky value at each wavelength. To ensure the robustness of our sky measurement, we require at least 30 fibers be involved in our background light estimation; for most frames, 80–90 fibers are used.

For the sky-background estimate, we do not attempt to remove spectral regions in fibers that have detected emission lines. This can be a problem for large objects, where the emission covers most of all of an IFU (at least $13'' \times 50''$). Large Lyα blobs, diffuse nebulae, and the halo regions of nearby galaxies can, in theory, fall into this category. For these objects, the problem can be mitigated by performing a full-field sky subtraction derived from the entire array of IFU spectrographs (see below). While we have not seen emission lines affect the sky subtraction of our frames, we realize that this might happen at a low level, and we continue to monitor such issues.

In cases where an object extends over a substantial fraction of full IFU array, then there is no easy solution for sky subtraction, and the observation is excluded from the cosmological analysis. The area of the HETDEX spring field that includes M101 falls into this category.

Figure 7 shows the sky and sky-subtracted residuals for one fiber from one exposure. The black line is the sky model using the local estimate that is based on the 112 fibers of the amplifier. The residual spectrum in red is the sky model subtracted from the extracted fiber spectrum. The variation in the sky residuals is consistent with that expected from the counting statistics of the sky.

In general, our sky-subtraction technique is extremely robust and accurate to the photon noise level. Aside from cases where the data are affected by controller problems, the main failure mode is when a bright star or a large nearby galaxy extends over a substantial fraction of an IFU. With our minimum criteria of 30 fibers required for a sky estimate, the inclusion of object flux over so many fibers causes the background to be overestimated. Even then, our detection algorithms (described below) can still pick out emission lines on top of the oversubtracted, negative continuum.

The excellent sky subtraction is partly a product of having a very robust measure of the fiber profile and the fact that we are working in a spectral regime with no strong sky lines. In fact, the strongest line visible in our sky spectra is from mercury vapor at 5461 Å. Sky lines do create correlated residuals in the spectral dimension for some of our amplifiers. These residuals are included in the noise estimation, with the noise being increased in those regions.

In addition to creating a sky estimate for each amplifier based on its own sky fibers, we also produce a full-frame sky subtraction based on the combined data of all the amplifiers. This global sky estimation, which is especially useful for removing the effects of bright stars and extended galaxies, is produced in a manner very similar to that of the local sky estimate. We take all fibers, in this case up to ~33,000, flag the fibers containing continuum sources, and generate a background light spectrum using the 90% of the remaining fibers with the lowest flux (generally 20k to 25k fibers). Both the full-frame and local sky subtraction are kept and used in subsequent analysis.

6.14. Astrometry

We require that the astrometric positions derived from the HETDEX spectra match those from imaging catalogs to $0\rlap{.}''5$. This specification is driven by our need to separate the Lyα
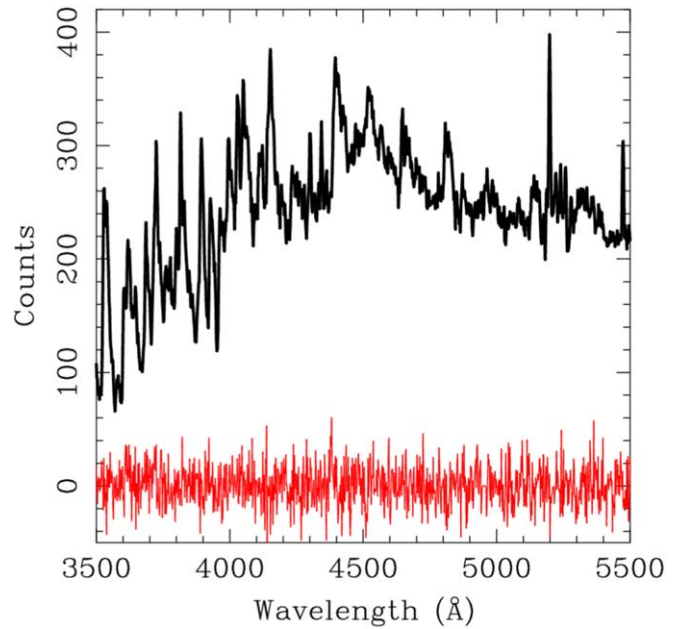

Figure 7. Sky model and sky-subtracted residual (in counts per pixel) for fiber 55 on amplifier RU of IFUslot 093 (spectrograph 507). The sky model, in black, comes from the local-sky estimate and is based on all 112 fibers of the amplifier. The sky-subtracted residual is in red and is derived by subtracting the sky model from the extracted fiber data. The spectra are from the first dither sequence of exposure number 016 on 2021 July 31. The variation in the residual is equal to that expected from the counting statistics of the sky counts.

emission from $1.88 < z < 3.52$ objects from the emission lines of foreground contaminants. Because the VIRUS spectrographs only cover the spectral range from $3500 \, \text{Å} \lesssim \lambda \lesssim 5500 \, \text{Å}$, virtually all LAEs and most [O II] galaxies between $0.13 < z < 0.47$ will have only a single emission line. (At redshifts below $z < 0.13$, Hβ becomes a second confirming line, and at $z < 0.10$, [O III] shifts in our instrument's spectral range.) To discriminate between the cases where only one line is detected, we need to compare the line fluxes obtained from the spectra with continuum flux densities measured from broadband imaging (see Leung et al. 2017). Given the large ($1\rlap{.}''5$ diameter) size of the fibers, and the fact that some of our imaging comes from the Hubble Space Telescope, the astrometric uncertainty of VIRUS sources will always be larger than that obtainable from imaging. However, simulations demonstrate that with three dithers, we can determine the centroid of any detected source to $\sim 0\rlap{.}''5$. (Doing better would require subsampling of the dithers, and even then, the improvement would be marginal at best, to $0\rlap{.}''4$.) This is sufficient to allow robust matching for most of the detected emission line objects.

We note that Lyα from high-z sources can be offset from the centroid of their host galaxy (Bond et al. 2009; Shibuya et al. 2014; Lemaux et al. 2021) although this offset is generally much smaller than $0\rlap{.}''5$. This scatter, along with the astrometric error, is taken into account when establishing the probability of a counterpart. Moreover, on those occasions where a HETDEX emission line has more than one possible counterpart, we can employ a Bayesian decision algorithm similar to that used in the HETDEX Pilot survey (Adams et al. 2011).

Obtaining $0\rlap{.}''5$ precision on VIRUS sources requires that the uncertainty associated with each field's global astrometric solution be negligible compared to the centroiding error of any
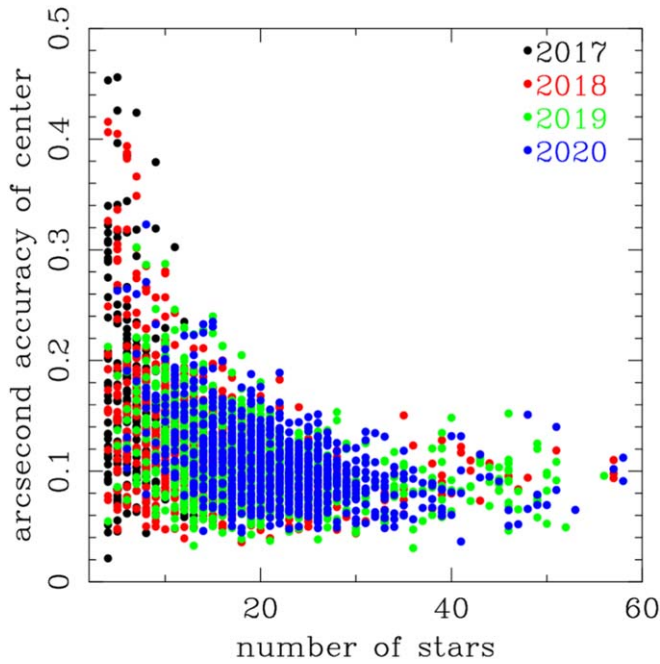
Figure 8. Accuracy in arcseconds of the focal plane center against the number of stars used for the astrometric solution. These data are for HETDEX observations for all of HDR2. The color corresponds to the year, where black is for 2017, red for 2018, green for 2019, and blue for 2020. The 2017 data had far fewer units installed and therefore fewer stars, which leads to the increased astrometric uncertainty.
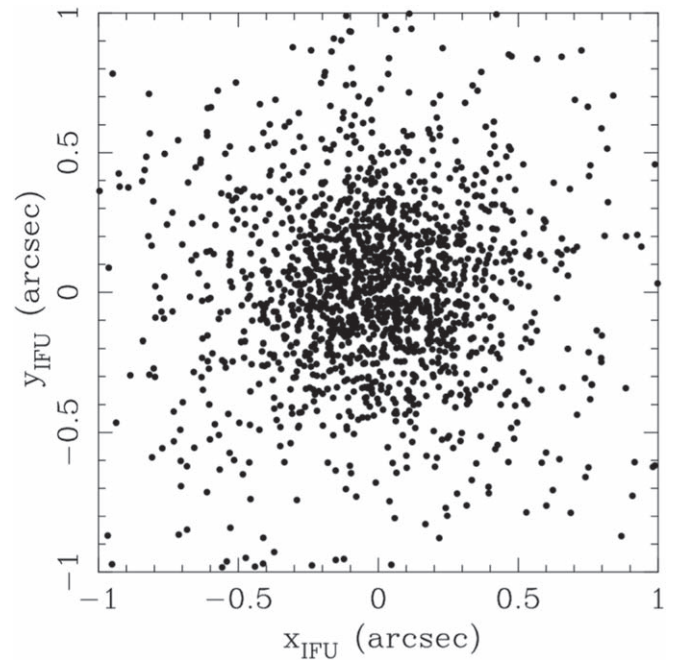

Figure 9. Offsets in arcseconds for stars selected for the astrometric measurement of the IFU seat position for IFUslot 046. There are 1800 stars in this figure covering all HDR2 fields, with magnitudes ranging from 14–20 in the g-band. The rms is $0\rlap{.}''35$.

individual source. There are multiple parameters that affect our ability to create such a solution. We consider four aspects that drive our astrometric analysis: (1) our knowledge of the field center, (2) the accuracy of the IFU seat positions, (3) our understanding of the dither offsets, and (4) our ability to reconstruct source positions for astrometric reference stars in the field. In fact, it is this last issue that limits our ability to determine the other three parameters. Specifically, in order to achieve $0\rlap{.}''5$ astrometric precision for individual sources, we need to define each field center, IFU seat position, and dither offset to better than $0\rlap{.}''2$. We typically reach $0\rlap{.}''1$ precision in these measurements.

To obtain an astrometric solution for each field, we first average each fiber's counts between 4400 Å and 5200 Å, and spatially interpolate those counts over the three dithered exposures to construct a pseudo "image" of the sky. An example of such an image is shown in Figure 3. We then use the PSF-fitting routines of DAOPHOT (Stetson 1987, 1990) to measure the IFU positions and fluxes of all the continuum sources, both on the interpolated image and on the individual dither frames. These lists are fed into the DAOPHOT routines `master` and `match` to determine the astrometric offset of each dither (thereby confirming the dither pattern) and best-fit (x, y) locations of the continuum sources on the HET's focal surface. Finally, the derived (x, y) positions of stars on the IFUs and the nominal (x, y) locations of the IFUs in the focal plane are compared to the objects' equatorial ICRS coordinates in the SDSS DR15 (Abazajian et al. 2009), Gaia DR2 (Gaia Collaboration et al. 2018), or PanStarrs (Chambers et al. 2016; Flewelling et al. 2020) catalogs.

High-galactic-latitude observations generally contain around 30 stars bright enough ($g \lesssim 22.5$) to use in an astrometric solution. With this number of stars, the global solution is usually accurate to better than $0\rlap{.}''2$. Figure 8 demonstrates this by showing the

accuracy of the astrometric solutions for all the HETDEX fields in HDR2 plotted against the number of stars used. The colors represent the year, with black for 2017, red for 2018, green for 2019, and blue for 2020. The older data had far fewer active units and therefore fewer stars to use for the measurement of the focal plane center. For the majority of the observations, we are meeting the required specification for our astrometric accuracy.

As noted above, to perform this measurement, we need to know the relative positions of each IFU in the HET's focal plane. The approximate IFU seat positions are known from laboratory measurements, and the strong expectation is that these relative positions do not move. Since the IFUs and spectrographs have been added steadily over the years, we regularly have the chance to test this assumption by remeasuring all the IFU seat positions using on-sky observations. Our data confirm the constancy of the IFU seats.

Our knowledge of the relative IFU seat positions improved over the years using the ensemble of HETDEX observations. A typical HETDEX field contains between 0.5 and 1 star per IFU in the magnitude range $14 < g < 20$. By tracking the astrometric offsets between the stars' cataloged positions (primarily from Gaia) and the positions found from our nominal astrometric solutions, we built up an offset map for each IFU. Figure 9 shows an example data set for IFUslot 046, where each point represents the offset of a HDR2 star in (x, y) coordinates. Since the focal plane center is determined using star positions on all the other IFUs, any change in the relative position of one affects the relative positions of all the others. Convergence therefore requires about three iterations, and at present the IFU seat positions are known to better than $0\rlap{.}''05$.

This same procedure also allows us to track our ability to measure a stars' positions in the absolute sense. As illustrated in Figure 9, our overall astrometric solutions have an rms accuracy of $0\rlap{.}''35$. This number can be broken down into its component terms as follows:

1. Fiber positions within an IFU: $0\rlap{.}''05$
2. Equatorial position of field center: $0\rlap{.}''1$
3. Focal surface position angle: $0\rlap{.}''2$. This assumes a $0\rlap{.}°05$ uncertainty in our knowledge of the field rotation, which is determined from the astrometric solutions for each observation at a mean radius of $4'$.
4. IFU seat positions: $< 0\rlap{.}''05$
5. IFU seat rotations: $< 0\rlap{.}''08$. This assumes $0\rlap{.}°2$ rotation error at a radius of $19''$ within an IFU.
6. Statistical: $0\rlap{.}''2$ for a three-point dither sequence on an individual star. This depends weakly on the signal-to-noise ratio of the continuum sources.
7. Cataloged star positions: $0\rlap{.}''15$. The main sources of scatter for this term are visual binaries/optical doubles and nonstellar sources.
8. Source proper motions: $< 0\rlap{.}''05$

Each of terms 1–7 has been measured with the current data, while the mean proper data come from the Gaia DR2 catalog (Gaia Collaboration et al. 2018). The sum of these terms in quadrature results in a formal accuracy of $0\rlap{.}''35$, in excellent agreement with the results shown in Figure 9.

6.15. Image Quality

In order to properly extract the spectrum of a point source, we need to measure the image quality of an observation, quantify how the image quality changes with the wavelength, and model the effects of differential atmospheric refraction (DAR). All of these quantities are derived from the VIRUS data themselves.

The image quality for each observation comes from modeling the distribution of light expected in the fiber array from a point source. To do this, we use the bright stars in the field; typically $\sim$30 objects are suitable for analysis. For each star, we integrate the spectral flux between 4500 and 5000 Å and fit for the star's equatorial coordinates, total flux, and best-fit PSF FWHM using a Moffat function with a fixed $\beta = 3.0$. This beta value is an average best fit from multiple exposures where we have determined the minimum χ^2 to the stellar profiles; we keep β fixed for robustness, although we expect the true value varies by a small amount. The PSF fit itself is performed via a grid search where we integrate the model PSF over the face of the fibers and then calculate the amount of light expected in each fiber. The best-fit PSF is the one that minimizes the χ^2 between the model and the data. We then use the PSFs of all $\sim$30 stars to determine the mean FWHM and scatter for the frame. Once the FWHM is known, we remeasure all the positions and fluxes for the stars and use the mean FWHM in all subsequent analyses.

We also use the brightest stars to measure the behavior of the FWHM with the wavelength. By repeating the above procedure within ten 200 Å wide spectral bins from 3500 to 5500 Å, and comparing the FWHM values, we recover the standard Kolmogorov turbulent atmosphere result, FWHM $\propto \lambda^{-0.2}$ (Roddier 1981). This relation is included in all subsequent analyses.

The Hobby–Eberly Telescope does not have a atmospheric dispersion corrector, and we must account for the shift of an object's position as a function of the wavelength. The same bright stars used for the PSF analysis provide a very accurate measure of this DAR. Our data demonstrate that from 3500 to 5500 Å, a source position moves by $0\rlap{.}''95$, a value in agreement with the calculations of Filippenko (1982). Given the large fibers, the large separation of the fibers, and the amplitude of this offset, it is essential to consider this systematic in any

spectral extraction. In fact, in order to accurately determine an object's flux, this criterion demands that a source's position be known to a precision of at least $0\rlap{.}''2$.

6.16. Throughput

In order to measure the power spectrum of $1.88 < z < 3.52$ LAEs over 540 deg^2 of sky, the flux limits of each observation must be known to high precision. Our target goal for this limit is 5%, averaged over the full field and all wavelengths. For each pointing, we must make sure that the uncertainty in throughput does not dominate over the error associated with the field's counting statistics. Given that the HETDEX survey consists of over 6000 individual pointings taken over many years, this is a formidable task.

Our definition of throughput is based on a 50 m^2 clear aperture and a set of three dithers, each with a 360 s exposure taken in clear weather; for exposures through the best conditions, this value is $\sim$0.16 at 5140 Å. Since our goal is to collect a data set that is as homogeneous as possible with respect to this throughput, we adjust the HETDEX exposure times to compensate for transparency variations, image quality changes, primary mirror illumination fraction, and sky background. In practice, we cannot obtain a constant flux limit over the full survey, due to limitations imposed by field availability and rapid changes in the observing conditions. Thus, for the purposes of our experiment, we define the "throughput" by treating each exposure as if it were 360 s long through a 50 m^2 clear aperture. We also refer to the throughput curve as the response function, as it allows for the translation of counts to flux units.

The 5% value on our flux limit precision is derived from the expected number of sources and from simulations. Our goal is to keep the error associated with the flux limit below that caused by Poisson noise of an observation, and it is based on the expectation of detecting 150–200 LAEs per field. A flux precision of 5% is good enough to produce only a modest increase in the overall error. This is confirmed by our simulations, which show that a 5% uncertainty on the flux limits has a negligible effect on our ability to measure the cosmological distance indicators.

We approach the flux and throughput calibration in two ways. First, for the scales within the focal plane, we measure the variation of a star taken in multiple frames, after removing an overall relative normalization. Second, for the uncertainty of the absolute normalization of a given field, we use the variation of all stars within that field. The ultimate test of the procedures then comes from comparisons with externally calibrated spectra.

For scales smaller than $23\,h^{-1}$ Mpc, we rely on different tests for the spatial and spectral variations. Since this scale is within the focal plane, we first remove the overall normalization of the measurements and look at the spectral variations of a star taken in multiple exposure sets. This is done by observing the same star on many IFUs and measuring the rms scatter in the count rate between observations while taking into account atmospheric diffraction in the spectral extraction process. This test typically produces rms scatters of 2% in 2 Å bins and is consistent for many different stars and over time. This is well within our specification. Thus, the limiting accuracy for our spectral calibration is determining the absolute flux calibration and the integrated flux limit. Figure 10 shows the typical variation between spectrographs in the spectral direction, as determined from the twilight sky exposures.
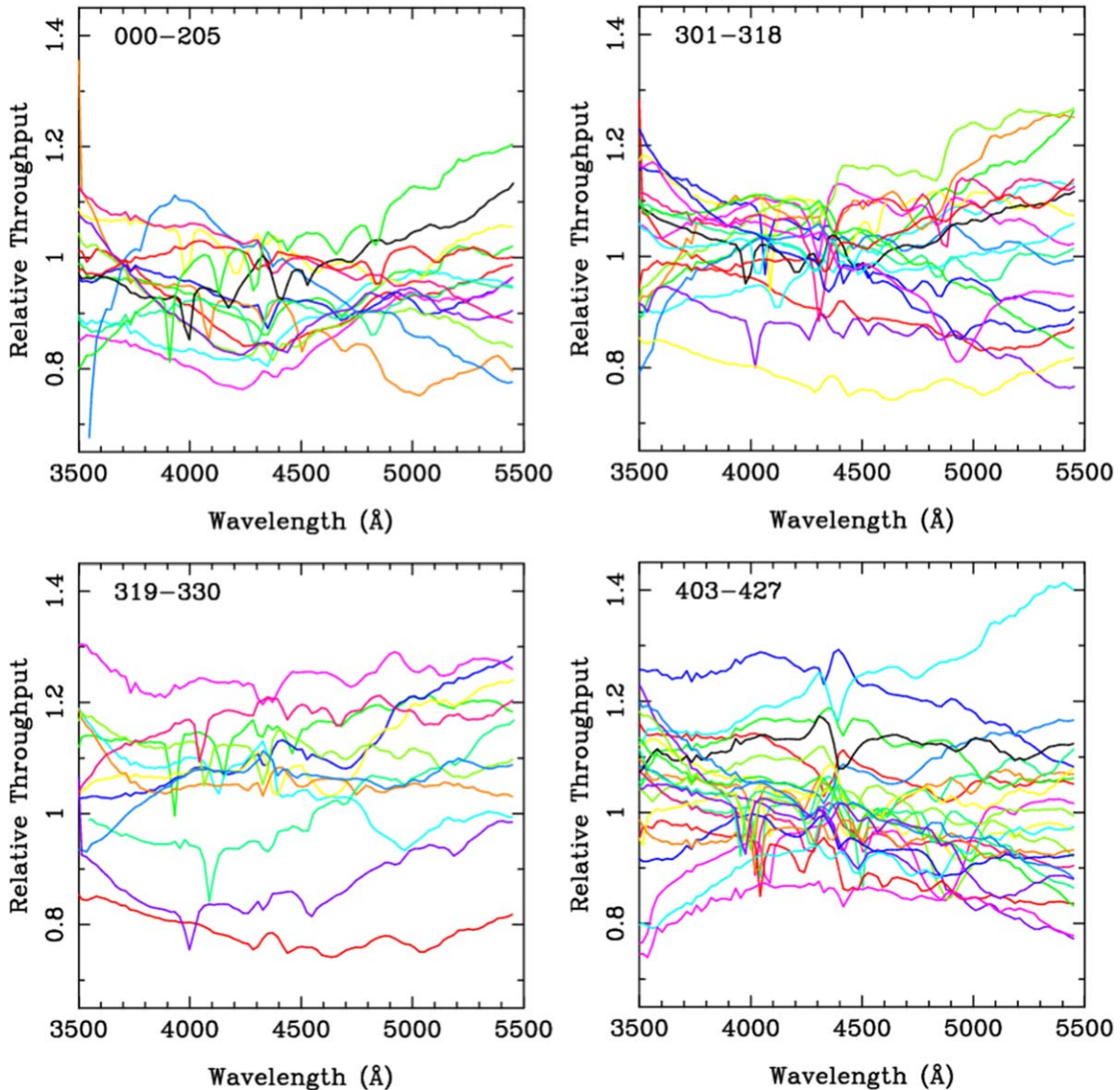
Figure 10. Relative throughput measures of amplifier LL versus wavelength for each spectrograph. Each panel presents data from a different set of spectrographs, where the spectrograph identification is given in the label near the top. The spectrographs with higher numbers are those installed at a later time and generally have higher relative throughput by about 10%. Each curve is based on the average of 112 fiber-to-fiber profiles within the amplifier. We use these curves to determine the relative throughput calibration over all IFUs.

Our calibration of spatial scales within an individual observation is performed using flux limit cubes created using the measured noise properties of the data. These cubes are calibrated using input source simulations specific for the observation and are especially important for addressing issues such as edge effects, transitions over amplifiers, fibers with little or no transmission, and the effect of bright stars and nearby bright galaxies on our measurements. The cubes are created with scales of 2″ per pixel, but this size can be adjusted for specific cases, as the only limitation is computer time. These cubes allow us to measure spatial variations within a given observation on scales similar to those for the spectral direction. The limiting accuracy for both spectral and spatial calibration is determined by the overall absolute flux limit. For the remainder of this section, we focus on obtaining the broadband absolute flux calibration of our spectroscopic data, integrated over a full frame and over all wavelengths.

6.17. System Flux Calibration

The Hobby–Eberly Telescope's design complicates the derivation of an absolute flux calibration for its spectra. Since the HET is set at a fixed zenith angle of 35° with a range of ±6°, differential atmospheric attenuation varies by at most 20%, as all observations are taken between 1.14 and 1.32 airmasses. (In fact, the variation is generally much smaller than this, since most HETDEX observations are taken close to the center of the track.) However, during an observation, an object's movement across the telescope's 11.1 m primary mirror may cause the effective aperture of the system to change, as different segments of the primary (with potentially different reflectivity coefficients) come into view. As a result, while observations of spectrophotometric standard stars might be used to infer the nominal wavelength dependence of the throughput curve, the applicability of these data to any single HETDEX observation is uncertain. Fortunately, the HETDEX spring and equatorial survey regions have been

well-characterized by the SDSS (Abazajian et al. 2009). Since the flux calibration of bright SDSS field stars ($g \leqslant 18.5$ mag) is good to <1% in *griz* and <2% in *u* (Padmanabhan et al. 2008), we can treat these stars as in situ standards. By integrating their stellar counts in 3″ radius apertures, modeling the effective aperture of the telescope, and taking advantage of the stars' SDSS photometry (along with photometry from other wide-field surveys), we can obtain an accurate characterization of the system throughput for each individual HETDEX pointing.

6.17.1. Methods of Flux Calibration

There are a number of ways to exploit the availability of SDSS photometry for HETDEX flux calibration. Below we describe three methods for performing this calibration, and the results of various tests of their robustness. We note that, since each method shares a dependence on the SDSS *g*-band magnitude, the three calibrations are not completely independent. Nevertheless, the techniques are sufficiently different so that each procedure provides a crosscheck on the other two methods, and the combination of all three yields a robust measure of our throughput curves.

Each method relies on generating the expected spectral energy distribution (SED) of field stars. First we choose stars from a catalog formed from a combination of Gaia (Gaia Collaboration et al. 2018), Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Chambers et al. 2016), SDSS (Abazajian et al. 2009), and United States Naval Observatory (Roeser et al. 2010) astrometry. We extract the spectrum of each star using a 3″ radius aperture (see Section 7.3) and generate a throughput curve using the techniques described below. By combining the throughput curves derived from all the stars present in a field, we obtain a first-iteration mean response curve for the observation, along with a measure of the star-to-star variance.

After deriving these mean response curves for all the observations, we normalize the curves to a common throughput and combine the data to produce a final global response function versus wavelength. We then compare this global response function to the curve derived for each individual observation and find the best-fit normalization and slope that transforms the individual curves into the global curve. This final fit is what we use in our analysis. Thus, our model for the system response is based on a two-parameter model (normalization and slope) fitted on top of a global response curve. Below we provide the details for the three possible methods for this flux calibration.

System throughput from g-band normalization.

On fully steerable telescopes, the response curve for most spectrographs is derived through the observation of spectrophotometric standard stars. In the case of the HET, where objects track onto and off of the different mirror segments, this simple procedure is inadequate. However, a slight modification of this standard procedure can produce a first-order estimate of the system response.

Observations of faint ($m_V \gtrsim 15$) spectrophotometric standard stars (Massey et al. 1988; Oke 1990; Allende Prieto et al. 2009; Bohlin et al. 2014; Narayan et al. 2019), if performed using the same three-position dither pattern employed by HETDEX, can be used to derive the nominal wavelength dependence of the instrument's throughput curve (see Figure 11). Measurements of SDSS field stars can then be used to renormalize this flux calibration curve to each individual pointing. In particular, since the *g* filter's bandpass falls entirely within the spectral range of the VIRUS spectrographs, one can compensate for a
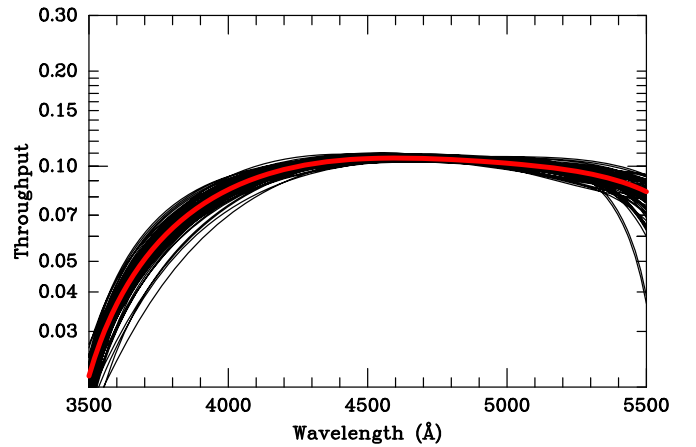

Figure 11. Individual response curves derived from 89 observations of spectrophotometric standard stars between 2018 January 7 and 2018 November 2. The mean response curve is shown in red. All the curves have been normalized to a common *g*-band filter response.

change in the effective telescope aperture or a grayshift in sky transparency by simply multiplying the *g*-band filter's transmission curve (Doi et al. 2010) by the flux-calibrated spectra of SDSS field stars, computing the stars' expected *g*-band magnitudes, and comparing these magnitudes to those listed in the SDSS catalog (Abazajian et al. 2009). Any offset can then be applied to the nominal throughput curve derived from the spectrophotometric standards.

The advantage of this approach is that the method can be applied to field stars as faint as $g \sim 23$, ensuring that more than enough objects are always available, even at high galactic latitudes. The disadvantage is that the calibration assumes that any change in the system response is achromatic. Such an assumption may be problematic, as the aerosol content of the air may change during the night. (Ozone absorption may also vary significantly on timescales of hours, but its effect largely lies outside the wavelength range of the VIRUS spectrographs. For a discussion of the various atmospheric components, see Hayes & Latham 1975, and references therein.) Moreover, since the effect of aerosols is most important in the blue and ultraviolet, observations taken through clouds may introduce a redshift-dependent systematic error into the HETDEX analysis.

System response from stellar absorption lines.

A second way of obtaining a full-spectrum estimate of system response is by modeling the absorption lines of individual SDSS stars. Objects with $g \lesssim 17$ have sufficient counts in their VIRUS spectra to allow a detailed comparison to libraries of stellar spectra, such as those of ATLAS9 (Kurucz 1970, 1979; Castelli & Kurucz 2003) and Medium resolution Isaac Newton Library of Empirical Spectra (MILES) (Cenarro et al. 2007; Falcón-Barroso et al. 2011). These fits, along with the stars' precision SDSS magnitudes, can then yield an estimate of the objects' effective temperatures, metallicities, reddenings, and, most importantly, their expected flux distributions above the Earth's atmosphere. A comparison to the observed spectra then yields the system response at each wavelength.

The weaknesses of relying on individual stellar absorption lines include the relatively low resolution of the VIRUS spectra, which results in the blending of several key spectral features, and the low sky density of bright stars at high galactic latitudes. Specifically, for most HETDEX observations, there is only $\sim$0.01 $g < 17$ star per IFU. Thus, even with the fully complement of 74 spectrographs, there is typically only $\sim$1

suitably bright comparison star per telescope pointing, and many fields contain no such stars.

Another major issue associated with using a single calibration star is the systematic error associated with its spectral extraction. The large fibers, the large fiber spacing, the mean $1\farcs8$ FWHM seeing of the survey, and need to compensate for atmospheric dispersion, all contribute to the overall shape of the extracted spectrum. As a result, even a small ($0\farcs05$) error in the assumed stellar position can produce a many percent shift in the blue-to-red normalization of an extracted spectrum. This systematic is further complicated by the fact that, even if the astrometric position of the star were known perfectly, we do not know the relative position of the individual fibers to requisite accuracy. Thus, the uncertainty in the inferred response curve will be dominated by the systematics caused by our imprecise knowledge of the stellar centroid and the effects of atmospheric dispersion, rather than simple counting statistics. As a result, throughput curves that rely on measurements of a single star are problematic and cannot produce robust calibration.

System response from multiband stellar photometry.

The third approach to HETDEX flux calibration involves the creation of synthetic VIRUS spectra of field stars based on their available multiband photometry and prior knowledge of the distribution of stellar luminosities, temperatures, and metallicities along the line of sight. These data can be used to predict the spectra of the SDSS field stars, which can then be compared with their actual VIRUS spectra.

To create the synthetic spectra of the field stars, we begin by generating a set of 1024 template stars with high-quality predictions for their photometric and spectroscopic properties. Our choice of stars is dictated by the Galactic model of Green et al. (2014) along two lines of sight taken to be representative of the HETDEX spring and fall survey regions. Using this model, we can predict the distribution of absolute magnitudes (M_g) and metallicities ([Fe/H]) of $g < 24$ stars in the HETDEX field. These predictions can be translated into model spectra using Padova and Trieste Stellar Evolution Code isochrones and luminosity functions (Bressan et al. 2012; Marigo et al. 2017), a grid of MILES stellar spectra (Cenarro et al. 2007; Falcón-Barroso et al. 2011), and foreground reddening.

Our analysis uses a constant value of $E(B - V) = 0.02$ across the field. This is, of course, an approximation, and there are reddening maps that could be used to infer extinctions to individual sources. However, these maps are for sources outside of the Galaxy (Schlegel et al. 1998; Schlafly & Finkbeiner 2011) or are limited by the low density of field stars at high galactic latitudes (Capitanio et al. 2017). More importantly, because the reddening in the HETDEX fields is already small and our procedure is probabilistic in nature, improved reddening estimates do little for the overall quality of our solutions. As shown below, we reach our specification on the flux calibration without the use of more complex, 3D reddening maps.

Using our grid of stellar spectra, we compute the most-likely spectrum for each field star based on its observed SDSS *ugriz* magnitude (Abazajian et al. 2009), its Gaia distance (Gaia Collaboration et al. 2018), and our assumption of foreground reddening. By comparing the predicted spectra for an ensemble of field stars to their observed spectra, we can derive an estimate of throughput versus wavelength.

There are two obvious improvements that can be made to this procedure. First, it is fairly straightforward to include other data sets in the calculation of the spectral calculation. For instance, *griz* photometry from PanSTARRS-1 (Chambers et al. 2016), near-UV and far-UV data from the Galaxy Evolution Explorer (Morrissey et al. 2007; Bianchi et al. 2017), near-IR JHK_s measurements from 2MASS (Skrutskie et al. 2006), and 3.4 and 4.6 μm flux densities from the Wide-field Infrared Survey Explorer (Wright et al. 2010) can all be used to help identify the most-likely entry in our spectral grid. Second, rather than using just two sets of priors (one for the spring field and one for the fall field), the priors on M_g and [Fe/H] could be recomputed for each individual line of sight. Our simulations suggest that variations in the priors over the HETDEX fields are minimal, but this flexibility would be required for observations at lower Galactic latitude.

The multiband stellar photometry approach can be applied to stars as faint as $g \sim 23$, which at high galactic latitudes corresponds to about 0.35 stars per IFU, or $\sim$30 stars per pointing for the final array of 74 IFUs. An example of a calibration curve derived via this method is shown in Figure 12. We show example fits for two different nights in 2020, one with average throughput and one affected by clouds. The individual response curves derived for each source are shown as either lines with color, which represent sources used in the fitting, or thin gray lines, which indicate objects that were excluded from the analysis via our flux calibration software. Most of the rejected sources are either galaxies, stars with poor extractions due to their proximity to the edge of a detector, bright saturated objects, or detector artifacts. In the early 2017 data, when only a small fraction of the IFUs were installed, generally only 5–10 stars were available for this calibration, and the flux calibration of these fields has a relatively large error. By 2019, when the IFU array was nearing completion, a typical calibration involved 20 or more objects and produced results that were extremely robust.

It is this multiband photometric method that is used to flux calibrate the vast majority of the data generated by the HETDEX survey. Of the 3300 fields observed in HDR2, less then 20 do not have a sufficient number of stars for the multiband photometric approach; for those frames, we relied on g-band normalization. The expected calibration error on these frames is illustrated in Figure 11.

6.17.2. Final Flux Calibration

As discussed above, the throughput curve derived from $\gtrsim$20 individual field stars (i.e., Figure 12) provides an initial estimate of an observation's response function. The next step takes advantage of the fact that, over small portions of the VIRUS spectral range, we expect the relative response curve of the instrument to be approximately constant over time. Our final flux calibration therefore comes from the analysis of all observations and applies only a normalization and a linear gradient to our survey's response curve; the latter corrects for any large-scale trend in the blue-to-red throughput that might be caused by atmospheric aerosols or systematic error in our spectral extractions.

The results of combining all the throughput curves derived from multiband photometry are shown in Figures 13 and 14. To create the figures, the stellar count rates of each frame's field stars were measured in a series of 100 Å wide bandpasses and compared to their predicted flux distributions based on the field's priors for stars of the given color and absolute magnitude. We then took the biweight average of these normalized throughput curves, and fit a fourth-order polynomial through the data. This polynomial serves as the global throughput curve and defines the instrument's nominal blue-to-red response. We then divide each individual
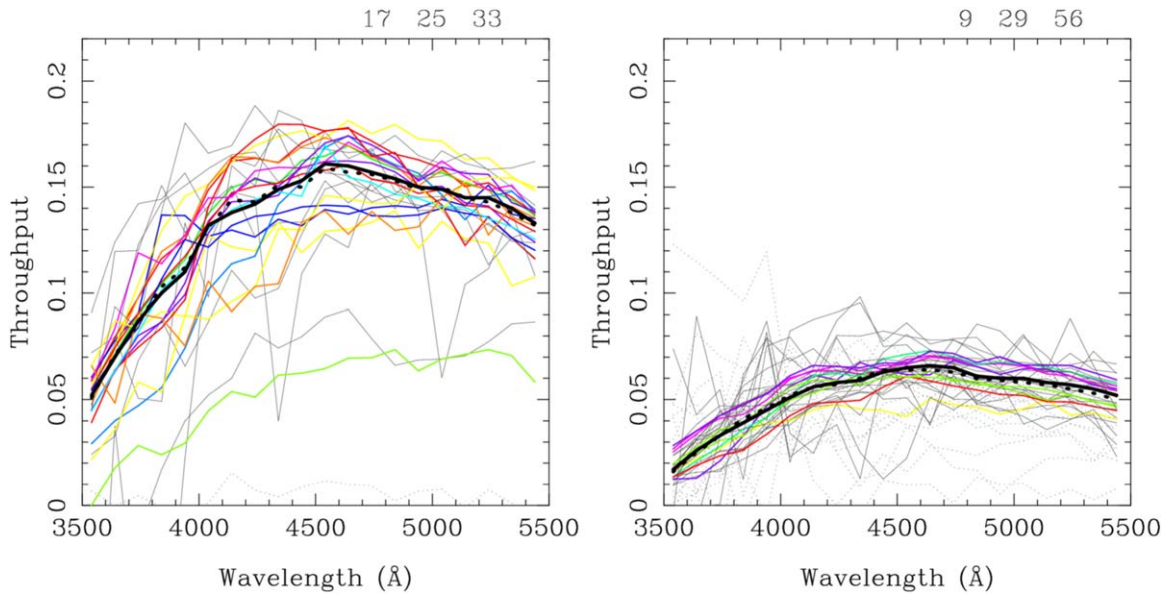
Figure 12. Two examples of the multiband photometry method for absolute flux calibration. The plot on the left is from observation 017 on the night of 2020 June 10; these data were taken under conditions typical for the survey. The plot on the right is from observation 021 on 2020 June 14, which was taken through clouds. The thin lines are the throughput curves derived for individual stars using the multiband fits: the thin gray lines, solid or dotted, show curves that have been rejected either due to poor extractions or poor predictions as to their intrinsic flux distribution, while the lines with color display are throughput curves that passed our quality criteria. The numbers on top, from left to right, are the numbers of continuum sources used in the final fit (lines with color), the number of continuum sources including those rejected based on their rms (lines with color and thin solid gray lines), and the total number of detected continuum sources, which includes stars, galaxies, and sometime meteors (all thin lines). The thick solid black line is the average derived from the accepted stars, while the thick dotted line is the average including the rejected sources. The agreement between the two means demonstrates the robustness of the fitting.

response curve by the global response function and fit the residuals with a linear model. The resultant normalization and slope were then applied to the global curve to yield the observation's final response function, and this function was applied to all the extracted spectra.

The left-hand panel of Figure 13 shows the general behavior of the solutions by normalizing the individual throughput curves for 894 fields observed in 2020 to the response between 4600 and 4800 Å and then scaling the curve to the mean throughput value of 0.15. The solid line shows the median of the fits, the dotted lines display the region containing 68% of the solutions, and the points illustrate the individual measurements that went into creating the curves. The right-hand panel uses these same data to compare the throughputs at 4040 Å to those at 5200 Å. This figure confirms that, although discrepant points are relatively common and there is considerable noise in the individual 100 Å wide data points that go into making the response function, the fitted curves derived from the ensemble of many field stars are relatively robust. Although the data were taken over a six-month period under a variety of observing conditions, the throughput ratio has little dispersion, and the flux calibration at any wavelength is stable to better than ~5%.

Figure 14 uses the throughput curves found from the multiband photometric approach to examine the system throughput for all 1442 HETDEX observations obtained between 2017 January 1 and 2019 Feb 9. The figure displays a histogram of the effective system response at 4940 Å, assuming a nominal three-dither sequence of 360 s exposures and a telescope collecting area of 50 m^2. Under most conditions, the effective throughput of the atmosphere+telescope+fiber +spectrograph system is greater than 13%; for reference, the HETDEX survey requires the effective throughput at 4500 Å to be greater than 10%. When combined with Figure 13, the data

demonstrate that even in the extreme blue, the throughput of the VIRUS spectrographs is generally greater than 5%.

6.17.3. Evaluating the Flux Calibration and Uncertainties

We use a variety of measures to estimate the uncertainties for the full system throughput. The discussion below provides the error associated with each component of the measurement.

To investigate the relative instrumental throughput versus the wavelength over the focal surface, we use all the twilight frames taken throughout a month. We create two masks from the data, one showing the relative fiber-to-fiber variation and the other displaying the detector-to-detector consistency. Based on the standard deviation of these ~100 frames, our relative calibration is accurate to 0.5%, with slight increases (up to 1%) at the edges of the field or detectors. This 0.5% variation represents a floor to our instrumental calibration and is significantly smaller than our measurement accuracy.

The measurement of the overall absolute throughput is affected by two different issues: the accuracy of the fits to the stellar SEDs and the errors associated with our spectral extractions. Both of these issues are significant.

To assess the accuracy of our throughput curves, we use the individual response functions derived from each star on a given frame and determine the standard deviation of the mean of those curves. In other words, since a field's response curve is a mean value based on SED fits to ~20 field objects, we use the response estimates from each individual star to measure the dispersion about this mean and therefore the error associated with the mean value. This simple analysis suggests that the overall accuracy of flux calibration is generally between 2-5%, with a small tail extending up to 10%. (The 10% values are from the very early data when only a few field stars are available). In fact, the overall accuracy is better than this, since our fits model the response curve as a fourth-order polynomial
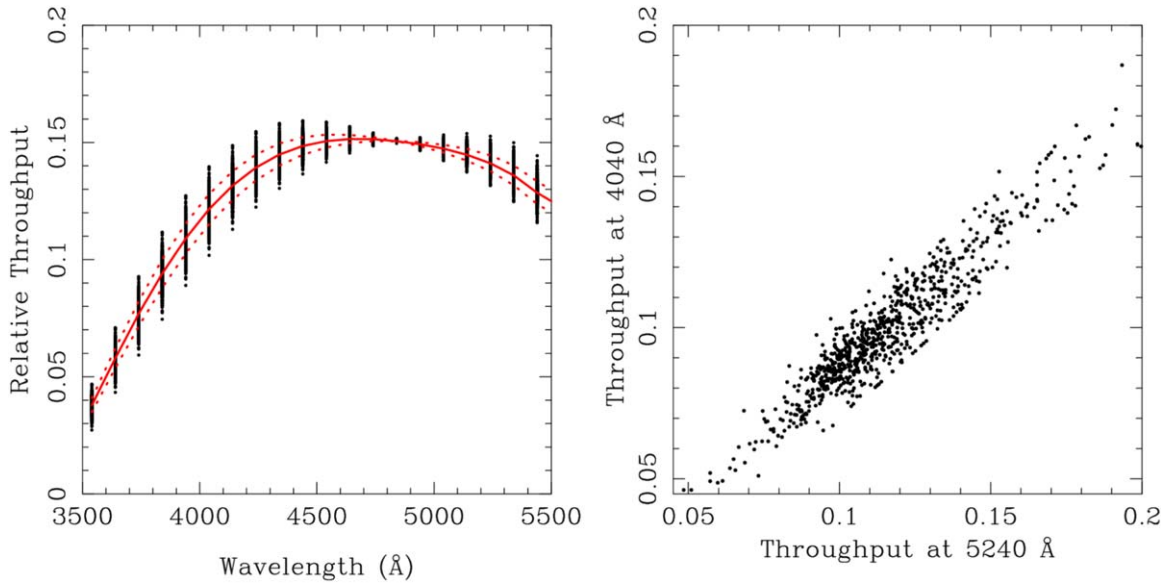
Figure 13. Throughput measurements derived using multiband stellar photometry. The left-hand figure normalizes each curve to 0.15 between the wavelengths 4600 and 4800 Å and shows the median fourth-order polynomial fit (solid line) and the region occupied by the central 68% of the curves (dotted lines). The individual measurements that went into creating the curves are also plotted as the black points. The right-hand panel compares the throughput at 4040 Å to that at 5240 Å. These data come from all the HDR2 observations in 2020.

and are therefore insensitive to high-frequency changes in the throughput. This analysis therefore represents the systematic floor to our throughput accuracy and is the dominant source of noise in our determination of the absolute throughput.

Another way of seeing this result is to plot the standard deviation of the mean response at 4940 Å against the total g-band throughput for all the HETDEX frames in HDR2. This is done in Figure 15, where the colors correspond to the year of observation (black is for 2017, red for 2018, green for 2019, and blue for 2020), and the mean throughput is estimated from simply scaling the stellar flux densities to match the stars' SDSS g-band magnitudes. Since the target accuracy for the absolute flux calibration is 5% on average, the plot suggests that we are well within this specification.

It is difficult to improve on this systematic floor to our flux calibration. For any individual source, extraction errors dominate; these errors include local pixel issues (charge traps, etc.), inaccuracies in the PSF, edge effects, image crowding, astrometric uncertainties, and, in particular, small uncertainties in the relative fiber positions. (Concerning this last issue, we could improve our knowledge of the relative fiber positions by using images of the fiber array taken in the lab. However, since we are currently within our overall specification, this improvement has not yet been implemented.) Based on repeat observations of bright sources located on or near the same region of our detector, our spectral extractions have a minimum variation of ~5%. For random sources in the field, this number increases to a mean value of 11%, with amplitudes of ~14% in the blue and ~10% in the red. Thus, for each exposure, it is important to average over many stars in order to reduce the errors associated with individual spectral extractions.

An obvious external test of the accuracy of our flux calibration is to compare the stellar flux densities derived by HETDEX to similar measurements from the SDSS (Adelman-McCarthy et al. 2008). In HDR2, there are over 3300 field stars with a high enough signal-to-noise ratio to make this comparison. Figure 16 shows nine such objects, with the SDSS spectra displayed in red. In the bottom row, the first two
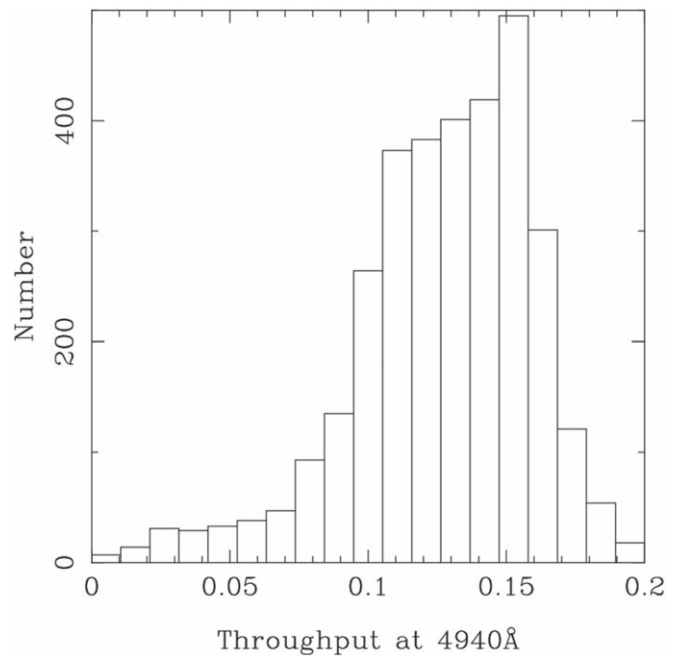
Figure 14. The effective response (in flux per count) at 4940 Å for all HETDEX observations in HDR2. The data assume a standard three-dither sequence with 360 s exposures and a telescope aperture of 50 m^2. This distribution is close to our predictions from the pilot survey (Adams et al. 2011). Observations with throughputs below 0.08 are not used in the final catalog.

stars are at the edge of (or even just outside) an IFU, making our flux calibration suspect. The last panel on the bottom row shows an example where there is a significant discrepancy between the two spectra; this is usually due to image crowding or a bad CCD (although, in a few cases, it is due to stellar variability—approximately 3% of SDSS sources have been found to be variable; Bhatti et al. 2010). We do not attempted to identify the objects that deviate significantly from their expected SEDs, as we have enough good stars in each
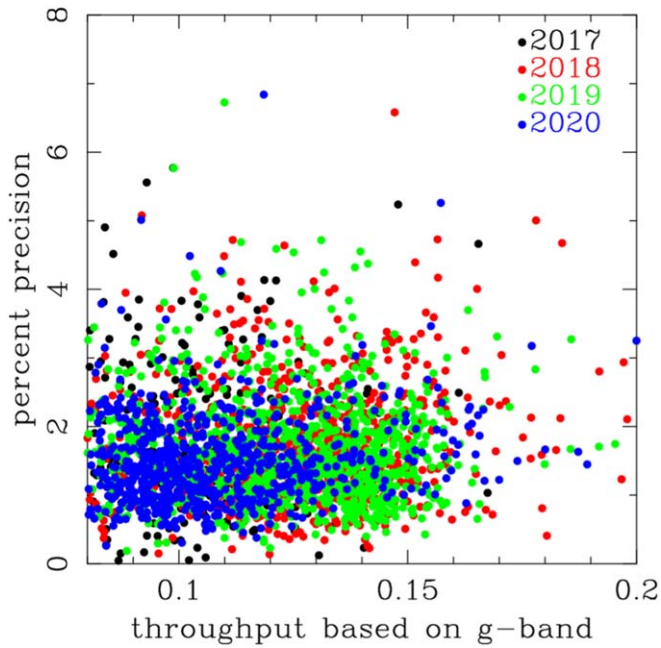
Figure 15. Precision of the throughput measurements versus the value of the throughput at 4940 Å, as determined from the *g*-band normalization of stars in the frames. The points include all HDR2 fields; black is for 2017, red for 2018, green for 2019, and blue for 2020. The throughput is measured by scaling the individual stars' fluxes to their *g*-band magnitude, and the precision is the uncertainty on the mean value (i.e., variation of the individual stars divided by the square root of the number of stars). Our target accuracy is 5% on average, with individual observations going up to 10%. In terms of the average absolute accuracy, we are meeting the project's specification.

observation to mitigate their contribution. As can be seen from the top two rows of the figure, most of the comparisons are quite good, both in a relative sense, and in terms of the absolute throughput.

Figure 17 summarizes the result of this test. The top panel displays the average ratio between HETDEX and SDSS spectra for 3363 sources observed between 2017 January 1 and 2019 February 9. To within the precision of the comparison, the relation is flat, demonstrating that we do not need to apply any postreduction correction. The black curve in the bottom panel of Figure 17 shows the dispersion between the HETDEX and SDSS spectra. This curve represents the expected flux accuracy for using any single star as a flux standard. The red curve represents the precision obtained by our adopted procedure of calibrating HETDEX using the SED fits of ∼20 stars with *ugriz* photometry, while the blue curve is derived by using the response curve found by scaling the mean throughput derived from spectrophotometric standard stars to match the *g*-band photometry of SDSS field stars. Note that the two procedures that rely on individual stars, i.e., the curves based on the observations of spectrophotometric standards and the stars with SDSS spectroscopy, have a higher uncertainty. This is as expected, due to the systematic issues discussed above. Our SED-fitting procedure mitigates this problem, as it averages the results of many stars in the field. Since a typical HETDEX exposure contains at least 20 stars bright enough to be fit, we should expect about a factor of ∼4 reduction in the uncertainty in the flux calibration. This is close to what is measured.

Yet another way of examining the robustness of the multiband photometry calibration is to observe a field containing a spectrophotometric standard star and compare the field's

throughput curve, as derived from the multiband stellar photometry, with that found directly from the standard star. The results of this experiment are displayed in Figure 18, where we plot the response curves derived for observations of the Oke (1990) standard star G193-74. Five nights of data are shown, with the left-hand panels comparing the response curve derived from the analysis of G193-74 itself to the results from the field stars. Both curves are represented by fourth-order polynomials. The right-hand panel shows the ratio of these two curves. At almost all wavelengths, the multiband procedure generates a response curve that is within 10% of that inferred from the spectrophotometric standard. There are some systematics associated with the fit, as the low-order polynomial used for the fitting does not completely capture the true shape of the curve. (The worst deviation is at the extreme red end of the spectrum.) Also, since G193-74 is at a slightly lower Galactic latitude than the HETDEX fields ($l = 166°$, $b = 31°$), the HETDEX spring-field priors used for the calculation are not strictly appropriate for the observation. Nevertheless, the comparison does suggest that the field-star colors provide enough information to flux-calibrate the HETDEX data to better than 7% precision.

We note that it is possible to combine the results of multiband SED fitting, stellar-absorption-line analysis, and spectrophotometric standard star observations to produce a set of throughput curves that are more accurate than those found from multiband photometry alone. While this may be an option for future HETDEX analyses, it is not needed at the present time, as our current procedure is already meeting or exceeding survey requirements for the vast majority of pointings. Since the multiband photometry method also yields quantifiable uncertainties, it is currently our baseline technique for calibrating HETDEX frames.

6.18. Galactic Extinction

All HETDEX spectra are flux calibrated but are not corrected for foreground Galactic extinction. Such a correction is employed in our flux calibration process, as our SED matching algorithm needs to know the stars' true colors. However, the factor is small (we assume $E(B - V) = 0.02$ throughout), and we do not correct our observed spectra for this effect. In the final HETDEX catalogs, estimates of Galactic extinction will be included in the database.

6.19. Post Corrections

Since the fiber-to-fiber flatfield calibrations are primarily based on twilight sky frames, and since these frames may need different corrections for illumination and scattered light, there could be frame-to-frame differences in their calibration. Our data suggest that the residuals from these effects are no more than 5%, and the worst cases are identified during the reduction process. However, there are regions of the HETDEX spectra that can be systematically off by a few percent. One reason for this is scattered light in the blue; another is associated with an absorption feature near 4350 Å produced by the VIRUS optical coatings. Both of these effects are difficult to remove in the pipeline reduction. To correct for these systematic offsets, we compare the HETDEX and SDSS spectra and average the fractional residuals over time, matching our spectra to the SDSS calibration. Specifically, we take all the data acquired over a year (over 1000 exposures) and compute the fractional residual correction that must be applied to all science frames to
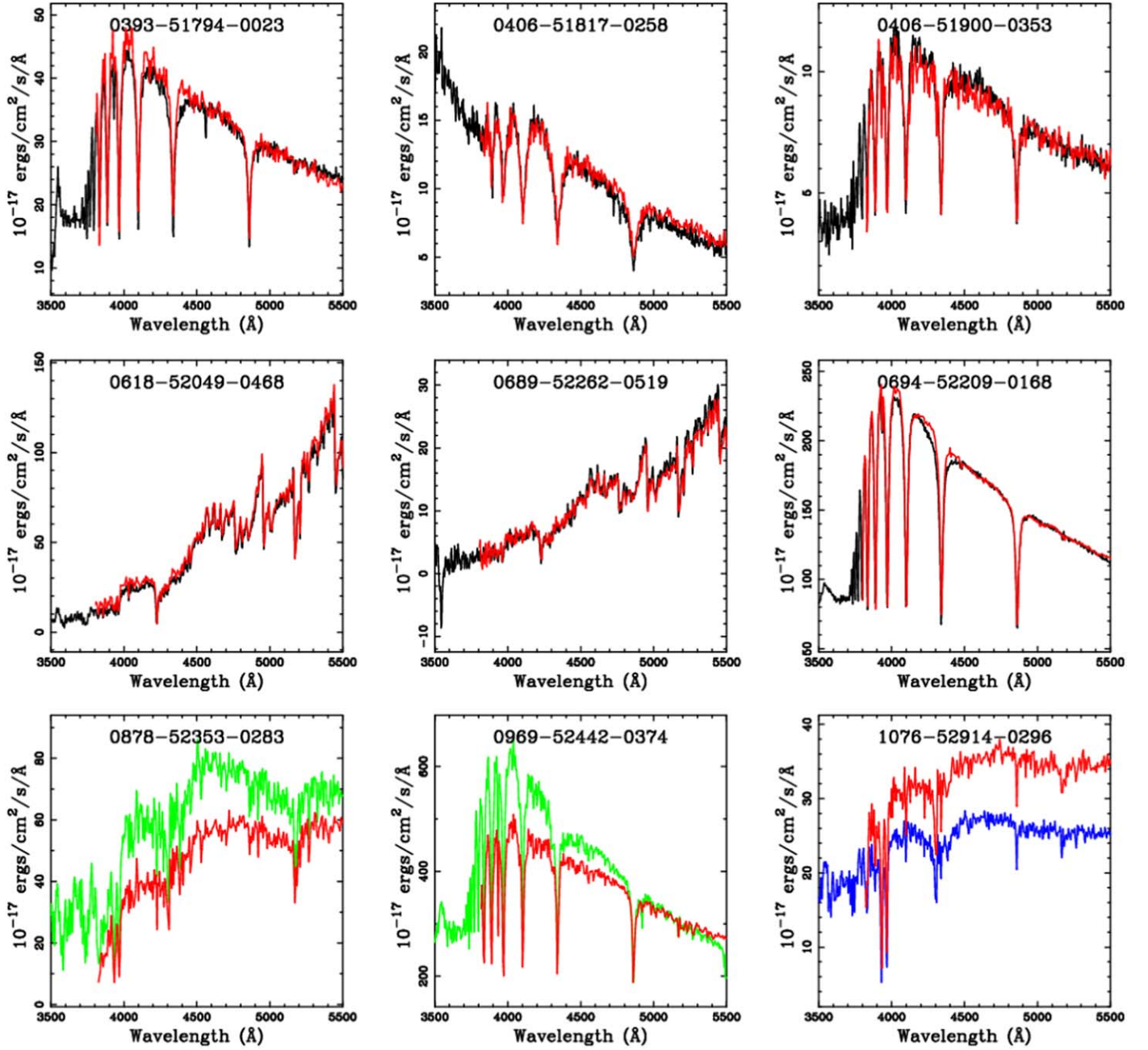
Figure 16. Example comparisons between SDSS spectra (in red) and spectra obtained by HETDEX (in black). The spectra in green represent data from sources at the edge of an IFU; the spectrum in blue shows a source where the SDSS and HETDEX results disagree significantly. We do not track the reasons for the disagreements, as we have enough stars for a robust calibration. The name at the top of each panel is the SDSS ID (plate, MJD, fiberid). The stars in the bottom row are included in the measure of our uncertainties for the overall calibration for a single object, and we are within specification. Since we use a calibration averaged over many sources, we can tolerate individual sources being off by up to 20%.

remove this small (few percent) systematic. Since the residual maps are stable over years, we can apply the same map to all HETDEX data. As the top panel of Figure 17 illustrates, there is little need for this post correction in HDR2.

6.20. Noise Model

An understanding of the element-by-element noise in our sky-subtracted spectra is essential for our line-detection algorithm and the generation of expected source counts. To acquire this knowledge, we compare three different noise models and use the differences between the results to identify potential reduction issues. The three noise models come from (1) the propagation of photon

noise throughout the system, including sky subtraction errors, (2) an examination of local noise based solely on spectral and spatial variations within and across fibers, and (3) the analysis of a global model, which tracks the variance of each spectral element over time. The third technique is by far the most robust, while the classical model of propagating photon noise is the least reliable. Consequently, we rely on the third model for our HETDEX analyses and use the measurements of the local noise (model 2 above) to help flag reduction issues. We discuss each model below.

The classical approach to measuring noise is straightforward: one simply propagates the photon noise through the reductions and tracks the additional uncertainties that come with each step.
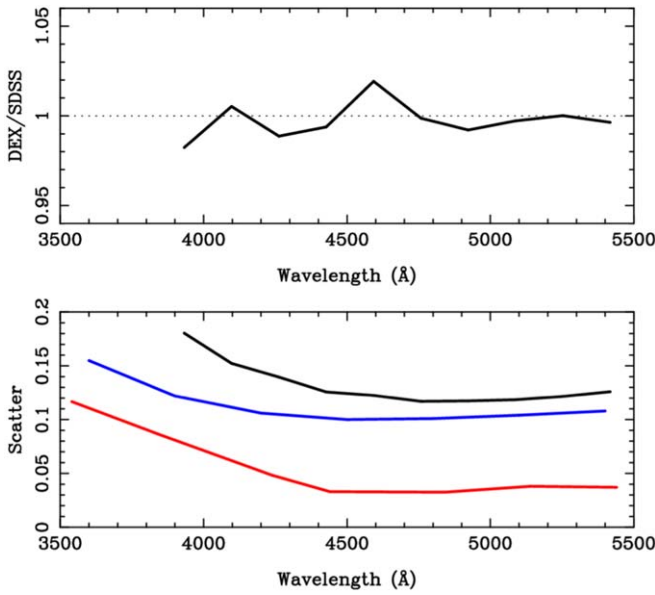
Figure 17. The top panel displays the average flux in a HETDEX spectrum divided by a corresponding measurement in the SDSS. This curve is based on 3363 sources observed between 2017 January 1 and 2020 June 6. The bottom figure is the scatter of the flux calibrations produced by our different methods. The red line is the average accuracy as a function of the wavelength, measured by performing SED fits to the SDSS *ugriz* photometry of field stars; the blue curve is derived by normalizing the response curve found from observations of spectrophotometric standard stars to the counts of SDSS field stars in a synthesis *g*-band filter. The black line is derived by using field stars with SDSS spectroscopy as in situ flux standards. The multiband photometric technique, which relies on observations of an ensemble of stars, is more accurate than methods that rely on a single spectral extraction.

Figure 18. The left panels show the HET+VIRUS throughputs derived from observations of the spectrophotometric standard star GD193-74 (blue) and the field stars surrounding the standard via multiband stellar photometry (red). The right-hand panels show the ratios of the curves. The measurements were performed over five nights in 2018 January. This comparison is limited by the errors associated with the spectral extraction of the standard star, but the agreement is still good to better than ~7% over most wavelengths.

The final noise model is then based on the counting statistics of the object, the sky, and the calibration frames, with the largest noise contributor coming from the sky subtraction. We assume the fiber profile is uncertain to 2%, and we include an additional component of 2% of the sky in counts, i.e., $\sigma^2 = \sigma^2_{\text{classical}} + 0.02*$ sky. Though mathematically simple, this approach tends to underestimate the actual noise in our spectra. The most likely cause of the underestimation is a combination of scattered light, dark current, sky-subtraction errors, fiber profile errors in the wings of the PSF, and detector controller issues that are not captured in the bias. We use these errors as a starting point when we track expected uncertainties over time (discussed below).

The second approach is empirical as it relies on measuring the local variations in the sky in both in the spectral and spatial directions. We rely on this technique to help flag data reduction issues. This procedure utilizes 100 pixels in the spectral direction and the 7 nearest fibers to inform the uncertainty of each pixel in the survey. We run this analysis box across each pixel for an observation, and define the noise to be the greater of two estimates: (1) the biweight scale over the 100 spectral pixels for a given fiber, or (2) the biweight scale using the 7 nearest fibers with 2 spectral pixels each. For bright continuum sources, this model overestimates the noise, since absorption lines and other continuum features increase the scale; for these objects, we simply use the noise found from counting statistics. This local empirical model provides a very robust measure of the statistical noise, but it can miss systematic noise introduced by sky-subtraction errors. Thus, even though this method is a significant improvement over the classical model, it

underestimates the noise around bright-night-sky lines. We therefore cannot rely on this approach for object detection, and only use it to help track down reduction uncertainties.

The third approach to noise estimation, and the one we adopt, is to merge the noise determined from photon counting with a global noise model. For every observation, we produce a sky-subtracted frame with the expected uncertainty based on the propagation of photon statistics. For each exposure and each extracted spectral element, we then examine the ratio of the sky residuals to the sky measurement, which we call the residual ratio. Finally, we measure the variance of each spectral and fiber element by combining the residual ratios for a year's worth of observations (i.e., many thousands of exposures). The result is a map of the actual variance compared to the expected variance for all ~33,000 fibers and 2000 spectral elements.

This noise ratio map provides an estimate of how well an element reflects it expected noise properties. For each pixel in an exposure, we multiply the noise from photon counting by the noise ratio map to generate an estimate of the true noise for the pixel. For most of the pixels of most of the detectors, the ratio is ~1.0. However, there are some detectors that show overall offsets, some that display blue-to-red variations, and some that exhibit large offsets around (relatively) bright-sky lines. Most of these variations are understood (i.e., due to scattered light, a systematic issue with the wavelength solution, or errors in the fiber or fiber profile), and we simply use the ratio as measured in all subsequent analyses.

Unfortunately, some of the HETDEX detectors go through phases when the background and amount of electronic

interference increases. The root cause of these excursions is likely an instability in the controller temperature. In these systems, the variance measured across a year's worth of observations will overestimate the true noise, as most of the observations have noise values that are much lower than that of the ensemble average. The occasional noise increases in the problematic detectors are easily seen when comparing our adopted noise model to the local noise model. That is, there are some emission-line detections that have a reduced χ^2 (from a Gaussian fit in the spectral direction) much lower than one, reflecting an overestimation of the uncertainties. We do not correct these cases and instead adopt the conservatively large uncertainties. The flux limits are derived from the same noise model, so the higher flux limits are included in our source simulations.

6.21. Flagging Pixels, Amplifiers, Detectors, and Bad Data

The final design of HETDEX uses 74 IFUs, 33,000 fibers, 148 spectrographs, 592 amplifiers, 592 fits frames, and 0.62 billion pixels and will include about 30,000 science exposures. Each component can create its own particular problem, and, given the size of this experiment, if a problem is possible, HETDEX will experience it. Thus, we must maintain a high level of quality control, especially down to the pixel level since our primary targets are only a few pixels across. Pixel defects can manifest as detected sources.

We maintain a barrage of measures to flag bad or compromised data. Most are based on statistical analyses of the individual frames, including tracking the number of detections on a CCD amplifier versus time and versus the number of detections on other amplifiers.

For each frame and each fits image (i.e., an image from a single amplifier), we track the sky level relative to all other fits images, the pixel rms after sky subtraction relative to the values on all the other frames, and the number of object detections. We then track this information over time, look for obvious trends, and, if necessary, remove the data of individual amplifiers based on this statistical analysis. Some amplifiers can come and go over time, and we maintain a database for whether one should be used.

We start our analyses with the pixel flats, where we identify bad pixels via their low throughput (a relative value generally <0.6) or "hot" response above a threshold (generally >1.2). Another significant problem are charge traps, and we generally remove the whole column when a charge trap is identified. The existence of charge traps on a detector range can remove anywhere from 0% to 20% of the area. There are also detector regions that are excised due to extreme pox or poor performance.

Some amplifiers have a low-level (a few counts per pixel) interference pattern on the detector. This pattern is a few pixel across and can cause significant a number of false positives. We understand this interference to be due to poor grounding within a given spectrograph, and in some cases we remove the whole spectrograph from the analysis. This obviously creates holes in the survey that are included in the window function.

Individual fibers can be bad for a few different reasons, such as being broken, having low or variable throughput, or just being poorly calibrated. These bad fibers are identified and flagged as well.

Another issue is that, on occasion, the HET's guider will vignette a corner of one of the IFUs. This happens about 1% of the time, and we flag the data from that IFU for subsequent analysis.

In our baseline 6 min integration, stars brighter than $g = 13$ will generally saturate. A small amount of saturation is not a problem as it will just result in a stellar spectrum that is not usable in any analysis. However, brighter stars can cause significant issues when trying to determine the local sky value. For example, the light from a $g = 11$ star can result in an entire amplifier being flagged as bad, and the scattered light from stars brighter than this can affect multiple IFUs. In fact, since the HETDEX spring field goes through Ursa Major, there are cases where the bright stars of the constellation have caused the majority of the IFUs to be flagged as bad. Bright meteors can have a similar effect and cause much of an observation to be lost.

Concerning this last point, our HETDEX database contains a large number of meteors, observed at a very high signal-to-noise ratio. Since the spectrum of a meteor is basically pure emission with very little continuum, these objects create false positives and are often initially flagged as LAEs. However, since we know the average spectrum for a meteor, we can flag these objects in software. Once alerted of a possible meteor, we can visually inspect the full frame and remove the whole trail. The width of a typical flagged meteor trail is about $10''$, though this number can be larger in extreme cases.

All of the effects above are flagged and removed in the subsequent analysis.

7. Object Detection

The above reduction steps produce a set of calibrated fibers with flux, flux uncertainty, a set of flagged pixels, and the fiber's equatorial coordinates on the sky. We search this data set for both emission-line and continuum sources using a model based on the PSF, which is measured from the data themselves (see Section 6.15). The procedures used for the emission-line and continuum source detections are very similar; we highlight the differences below.

The primary targets of HETDEX are LAEs between $1.88 < z < 3.52$. Since the typical image quality for our observations (FWHM $= 1''8$) corresponds to about 15 kpc at these redshifts, we assume that the spatial profile of most LAEs will be consistent with that of a point source. This is good assumption; although galaxies at $z \gtrsim 2$ do have extended Lyα halos, the surface brightness of a typical $z \sim 3$ LAE will e-fold more than 3 times within the radius of our median PSF (Ouchi et al. 2020), and a $3''$ extraction radius (see Section 7.3) will contain virtually all the Lyα flux of a typical object (Wisotzki et al. 2016). Moreover, as shown in Herenz & Wisotzki (2017) the detection signal-to-noise ratio for a faint LAE is virtually unaffected by a modest template mismatch; we will still detect our target galaxies at close to the maximum signal-to-noise ratio using a point-source model. While some of the brightest LAEs may have more extended emission (e.g., Ouchi et al. 2020), the use of the point-source assumption should not introduce any biases for the clustering analysis.

HETDEX does not rely on images to inform the placement of spatial apertures for extraction. We instead search all spatial and spectral resolution elements in the survey for emission-line or continuum flux. Object detections involve a three-step process of generating an initial sample of targets of interest, refining the selection around those targets, and culling the sample to generate a final catalog. Simulations inform each of these steps. For the emission lines, the properties that determine

whether a resolution element is retained are its signal-to-noise ratio, the χ^2 of a Gaussian fit in the spectral direction, and the fitted line width. For continuum sources, presence in the final catalog is simply a function of the signal-to-noise ratio. We discuss each step in turn.

7.1. Emission-line Grid Search

The emission-line detection algorithm begins by searching all spectral and spatial resolution elements associated with each IFU. At each fiber location, we spatially extract a spectrum based on the PSF and attempt to fit an emission line at each location in the spectrum. This is initially done using a grid that is $0''5$ in the spatial direction and 8 Å in the spectral direction. Based on simulations where we attempt to recover artificial sources placed on the sky, this spatial and spectral sampling is optimal for the initial search. This grid has $448 \times 3 \times 500 \times 9 = 6M$ elements per IFU (where the numbers correspond to the number of fibers, dithers, spectral elements, and spatial elements); since the current data release has $\sim$160,000 IFU pointings on the sky, our initial object search involves analyzing $\sim 10^{12}$ elements. We do note that a slightly finer grid would increase our detection sensitivity, but the computation time needed for such grids becomes prohibitive. Moreover, our numerical experiments demonstrate that $0''5$ spatial bins produce source lists that differ from the ideal case by less than 1%.

At each resolution element, we fit a Gaussian line profile with a line width fixed to the instrumental line width. We then keep all sources that have signal-to-noise ratios >4.0 and $\chi^2 < 3.0$ for follow-up analysis. These numbers are quite generous and produce a catalog that is about $10\times$ larger than the final catalog. To produce a more robust source list, we then conduct a secondary search using a 5×5 raster centered on the position of each candidate, with $0''15$ spatial steps. The location within this raster that provides the highest signal-to-noise ratio of an emission line is assumed to be the true source position. During this step, the same input source may appear more than once in the output catalog; to deal with this, objects that are within some specified distance of each other, both spatially and spectrally are combined. For our initial catalog, any two sources that are within $3''$ and 3 Å of each other are considered to be one object. The position that has the highest signal-to-noise ratio is then defined to be the source's location.

For a single IFU, the initial grid search provides about 50–100 sources of interest. The secondary grid search generally reduces this number to about 20 sources per IFU. This decrease is due primarily to more aggressive requirements on the signal-to-noise ratio and χ^2; we prefer a more lenient limit for these properties in the initial search, as our fits generally improve when the spatial location is better centered. (A difference of $0''2$ can make the difference as to whether a source makes it into our sample.) A source measured in the secondary search must have a signal-to-noise ratio >4.8 and $\chi^2 < 1.2$ to make it into our final catalog (see M. Cooper et al. 2022, in preparation). Currently, we maintain a source list that goes down to signal-to-noise ratios of 4.5, but objects below the threshold of 4.8 have not yet been vetted to our satisfaction. The translation from the signal-to-noise ratio to the flux changes from field to field and over IFUs within a field, and we base the individual IFU catalogs on the signal-to-noise ratio selection. In the final catalog, we generally have about five

sources per IFU, coming from an initial search of six million resolution elements.

7.2. Continuum Grid Search

While the focus of HETDEX is on emission lines, there is substantial ancillary science and calibration improvement that comes from the creation and search of a continuum catalog. Of course, we could rely solely on deep imaging surveys to select regions for the extraction of continuum sources, but there are advantages to providing a continuum catalog based entirely on a grid search of HETDEX data. Some examples include measurements of variable stars, transient sources, and moving objects and identification of issues associated with our flux calibration. In fact, the combination of results from a continuum grid search with the spectral extraction of sources with known positions will ultimately provide the most science. Here we focus on defining a continuum catalog based on a grid search, without any reliance on an imaging catalog. The first step in such a search is to define a lower bound for continuum detection. For the first pass, we use a conservative cut, corresponding to around $g = 22.5$. In future analyses, we will extend our cut to lower fluxes and combine these data with dedicated extractions based on an imaging catalog.

For each of the 448 fibers in an IFU, we measure the detector counts in a 200 Å window in the blue (from 3700 to 3900 Å) and in the red (from 5100 to 5300 Å). If either region contains more than 50 counts per 2 Å pixel on average (about $g = 22.5$), we flag it as a possible continuum source. As mentioned, this 50-count limit is arbitrary and designed to be conservative; objects can be detected more than magnitude fainter than this limit. Once we detect a possible source, we then search about the fiber position, using a 15×15 element raster and $0''1$ spatial bins. The spatial location that achieves the lowest χ^2 fit to our PSF model defines the center of source, and a point-source extraction at that position (see Section 7.3) generates the spectrum. This peak-up procedure is obviously different from the emission-line peak-up algorithm where we center by using the highest signal-to-noise ratio. Currently for the continuum source grid search, we only employ point-source model extractions, which are clearly not adequate for resolved galaxies. Refinements for extended sources are planned for the future.

At our current threshold, we generally find about 0.5 continuum sources per IFU, with about 80% triggered from the red bandpass and 20% triggered from the blue spectral region. In the current continuum catalog, we have about 80,000 sources. We are currently refining our ability to robustly extract fainter sources, and we expect to extend down to around $g = 23.5$ in the near future. Thus, we plan to increase significantly the continuum sources found from a grid search. Once combined with an imaging catalog, we should produce around three continuum sources per IFU.

As with its emission-line counterpart, the HETDEX continuum catalog does not rely on input positions from an external database. Obviously, many of the detected continuum sources are stars in the Milky Way; for example, Hawkins et al. (2021) crossmatched the HDR2 continuum source catalog with objects in the Gaia DR2 database (Gaia Collaboration et al. 2018) to examine the spectra of stars with a detected proper motions. However, the HETDEX continuum source catalog contains much more than just stars; it has galaxies, transients sources, and moving objects, such as asteroids, comets, satellites, and meteors. Figure 19 shows typical
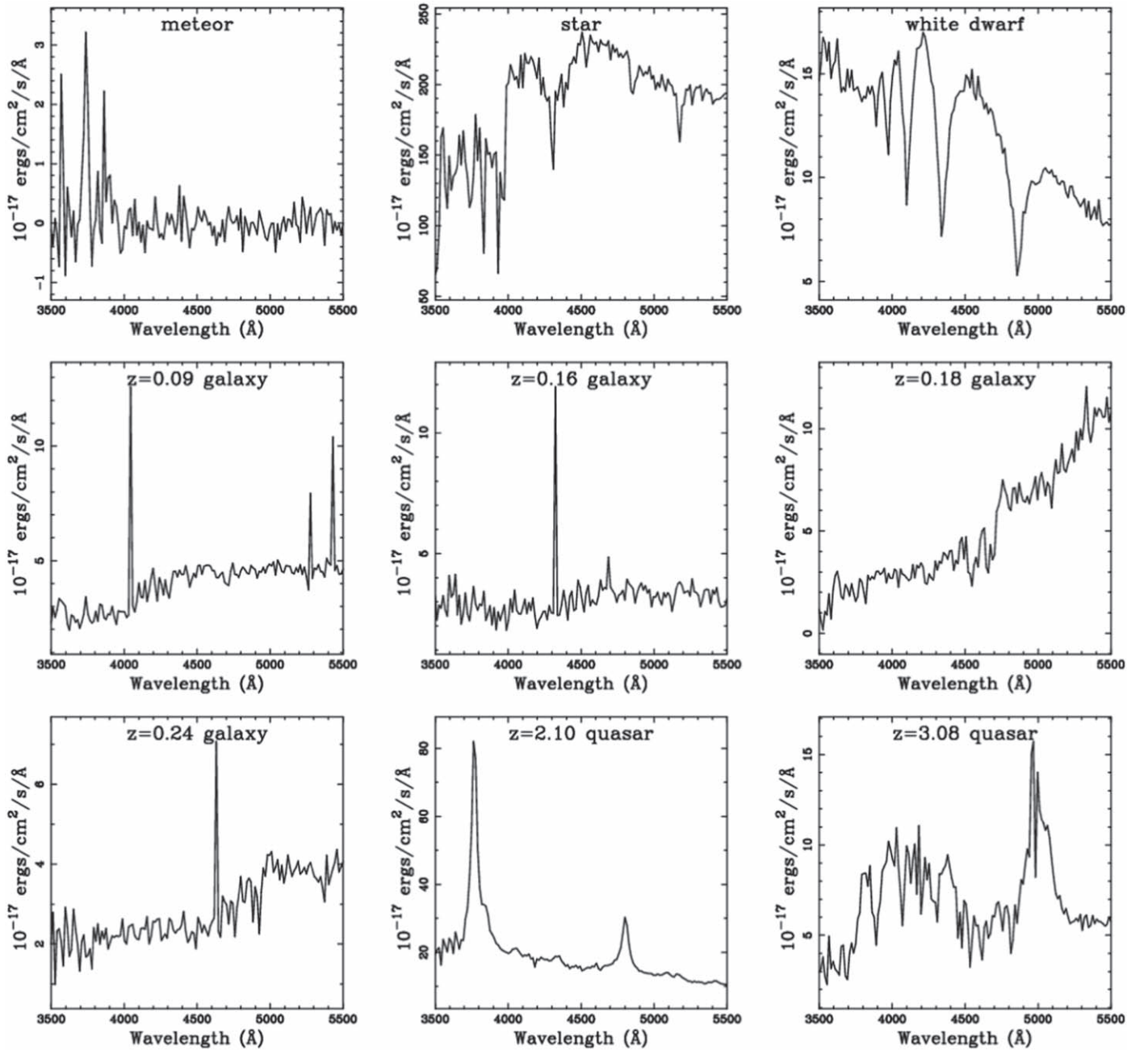
Figure 19. Nine examples of continuum sources. The top left spectrum is that of a meteor; all meteors make it into our emission-line catalog, and some are detected as continuum sources. The remaining eight sources are stars and galaxies, with their classification given at the top of each panel.

spectra for continuum-selected objects identified in 2020 January. Some objects, such as meteors, emission-line galaxies, and quasars, can be present in both the continuum and emission-line source catalogs. Others are pure continuum sources.

7.3. Point-Source Extraction

The size of our HETDEX extraction aperture is dictated by the design of the instrument and the typical seeing at the telescope. For the LAEs, we use a point-source model, which will bias the measured fluxes for spatially resolved sources. Since the HET does not have an atmospheric dispersion corrector, the offset between an object's position at 3500 Å and that at 5500 Å is $\sim 0\rlap.{''}95$ (see Section 6.15). This value, combined with the typical image quality of the observations

($\sim 1\rlap.{''}8$) leads to our use of a 3″ extraction radius. All fibers within 3″ of a source position are taken into account in our analysis. A typical object therefore contains counts from ~ 15 different fibers. We note, however, that given the typical PSF of the observations, fibers that are close to 3″ away contribute very little flux due to the point-source model used in the extractions.

The initial step for spectral extraction is to define the weights given to each fiber as a function of the wavelength. This is computationally intensive, as the mean PSF (as computed in Section 6.15) of a typical three-dither observation has a FWHM similar to diameter of a HETDEX fiber, $1\rlap.{''}5$. Thus, to determine the proper weights for an extraction, we must use an object's centroid, as determined from our $0\rlap.{''}15$ step-size raster search, and integrate the PSF across the face of each

fiber. This means that after performing our initial six million element object search and our higher-resolution 5×5 raster search, we must integrate the PSF of each source over the ~15 fibers contained within the aperture.

After determining the appropriate weight for each fiber, we combine this number with the fiber's noise properties. Recall that the fibers that go into an extracted spectrum come from three dithers and are spread across the detector. Their noise properties may therefore be quite different, with variations up to a factor of ~2. Furthermore, about 3% of all pixels in an image are flagged as bad data for various reasons (e.g., cosmic rays, charge traps, pixel defects, bright star contamination, etc.). This information modifies the weights defined by the observation's PSF and generates the final weights for each fiber. These weights are then used to produce the object's 1D spectrum via the optimal extraction algorithm of Horne (1986).

7.4. Emission-Line Fitting

The final step in the detection of emission lines is to examine the 1D extracted spectra and apply a matched filtering algorithm to the data. For each extracted spectrum, we run a wavelength window over the full spectral range, varying the initial central wavelength and range of the search. By breaking the spectrum into different regions, we can identify multiple lines in the same spectrum and mitigate the effect of local systematics introduced by the data reduction.

We fit the lines using a step size of 4 Å and a window of ± 50 Å. The first spectral fit uses an initial guess of 3500 Å for the wavelength of the emission line and uses a range that extends from the minimum wavelength for the spectral extraction (generally around 3490 Å) to 3550 Å. Each successive window then increments the red end of the range by 4 Å, until the window becomes 100 Å wide; after that, this 100 Å wide window moves across the spectrum in 4 Å steps, until it gradually tapers down to 50 Å at the red end of the spectrum. The spacing of 4 Å leads to a large number of multiple hits for the same source; these are trivially combined into a single source at the detected wavelength. We have measured the fraction of missing sources as a function of the step size. Increments of 8 Å and smaller never miss a source, whereas larger step sizes will sometimes cause the fitter to not peak up on an emission line.

For our initial run, we adopt a Gaussian line width of $\sigma = 3.0$ Å; for subsequent fittings in a refined search, we allow this line width to float. The advantage of fixing the line width is that it improves our ability to detect faint emission lines matching the line width value. For example, if we allow the line width to float initially, a low-level continuum source can be identified as an object with a wide emission line; in this case, a high line width fit would be reported, and a potential real source with a smaller line width might be omitted from the sample. Based on our simulations of input sources with varying line widths, we have settled on the approach of initially setting the line width at 3 Å, a value that still allows us to recover sources with wider emission lines.

The detected emission lines are fit with a Gaussian model using Levenberg–Marquardt minimization. In the initial fit with $\sigma = 3.0$ Å, there are three parameters: continuum level, total line flux, and line centroid; in subsequent fits we treat line width as an additional parameter. We have experimented with additional terms, i.e., using Gauss–Hermite polynomials and fitting more than one line, but have found that a single Gaussian fit is adequate. In future work, we may explore other types of line profiles, such as the asymmetric Gaussian used by Shibuya et al. (2014) or a mean profile determined by from our own data. However, since the resolution of the VIRUS spectrographs is low ($R = 750$–950) and most of our sources have low signal-to-noise ratios, such improvements are unlikely to add much to the detection optimization. We note that fits with symmetric profiles will be poor if the source has a high signal-to-noise, skewed line profile. For example, AGN and quasars will not be well fitted by a Gaussian. Additionally, for some extreme LAEs that have very skewed line profiles, the adopted profile may not be adequate. For cosmology measurements, this effect should be insignificant, since we generally care about whether the source is detected and not the precise amount of flux recorded in the emission line. There is the secondary consideration that we need to know the completeness for any detected source, and there might be a small systematic in the case of faint, but extremely skewed Lyα profiles. For AGN and quasars, C. Liu et al. (2021, in preparation) refits the lines using the full spectrum and also derives a refined completeness limit.

We define the signal for the source using the area under the Gaussian fit; we have compared this area to the result obtained by simply summing the flux within an emission line, and the results are nearly identical. Of the 15 fibers that go into an extracted spectrum, generally 3 have weights above 10% and at most 5 have weights above 1%. These numbers, of course, are dependent on the PSF of the observations and the angular size of the object. Any object with a high-weight ($w > 10\%$) pixel within 3 Å of the line centroid that has been previously flagged as bad is removed from the source list. Cosmological parameters can be sensitive to false positives, and this step helps reduce the number of false detections in our sample.

The uncertainties on the fitted line parameters come from a series of Monte Carlo simulations. We do not rely on the covariance matrix that we generate from the initial fit, although we do compare the predictions of this matrix to those from the numerical simulations in order to look for inconsistencies. (The simulated uncertainties are always higher than those derived from the covariance matrix.) For our Monte Carlo simulations, we use the input line profile and its uncertainty, which come either from the rms of the fit or our noise model, whichever is greater. Generally speaking, the use of the higher noise value improves the rejection rate of false positives at the expense of missing some real sources. We note that it is not unusual for the noise model to be higher than the rms; since this model is based on data taken over a year, detectors whose noise properties change with time will occasionally be better than characterized (see Section 6.20). We track these noise differences to identify variability in the detectors.

The Monte Carlo simulations start with the four fitted parameters, generate a Gaussian line profile for the source, use the noise model to create a synthetic spectrum, and then fit the spectrum's emission line using the same algorithm as for the original measurement. This process is repeated 100 times to determine 68% uncertainties on the emission line's location and strength. This information, along with the line's measured wavelength centroid, flux, line width, continuum level, signal-to-noise ratio, and χ^2 is then entered into the emission-line catalog.

Figure 20 shows some typical emission lines and their fits for the night of 2020 April 25.
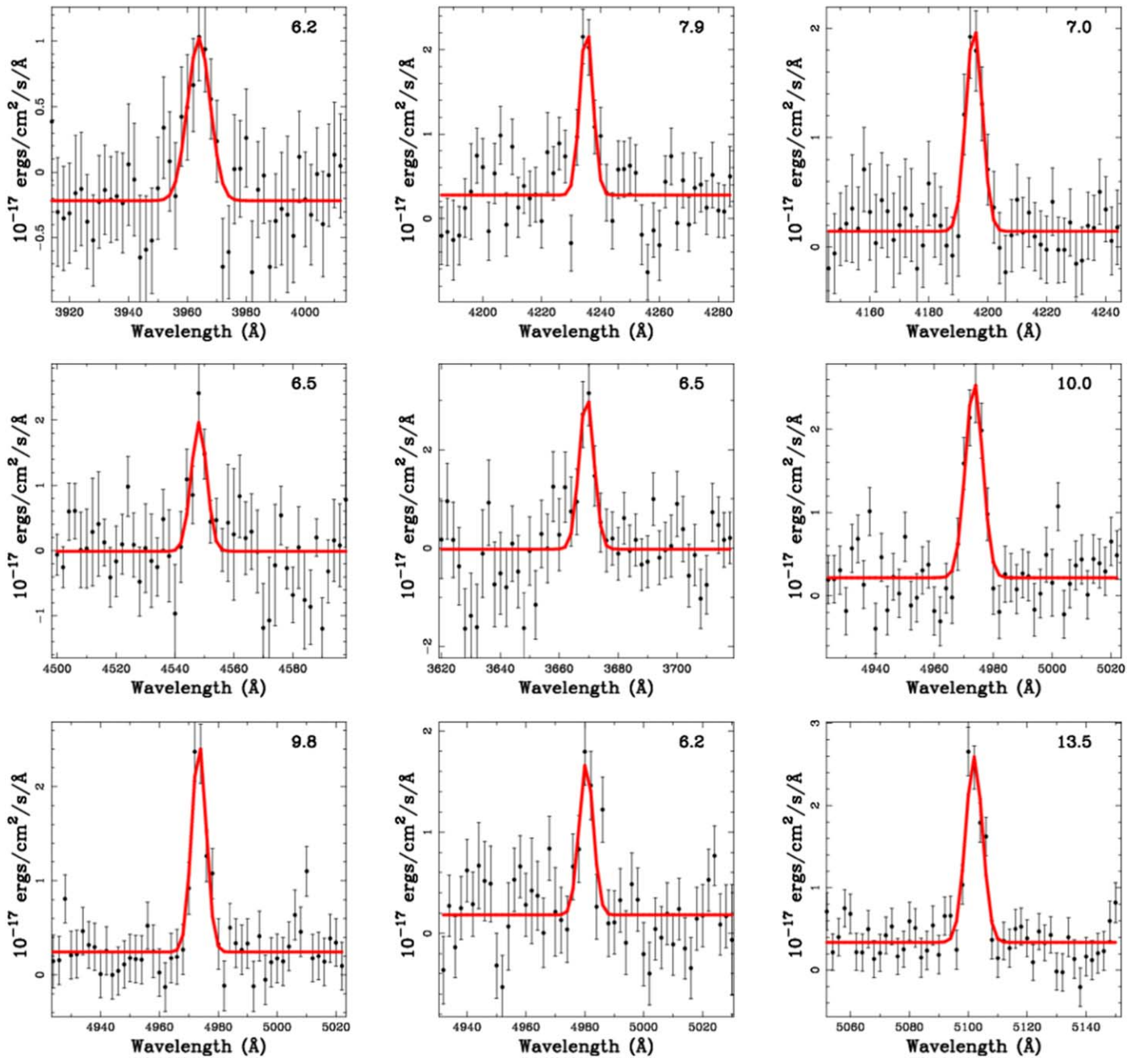
Figure 20. Typical spectra and their fits from one exposure set taken on 2020 April 25. The signal-to-noise ratio is shown on the top right of each figure. These spectra were chosen at random from over 400 emission lines present in the observation.

7.5. Detection Rate

There are three science requirements that are essential to track, as these are the most important for reaching the cosmological specifications of measuring H and D_A to better than 1% accuracy. These are (1) the number of Lyα emitters detected, (2) the number of false detections, and (3) the number of interloping galaxies, such as $z < 0.5$ [O II] emitters that are mistaken for LAEs.

For the total number of detected LAEs, the best metric to use is the number of emission-line detections per IFU. From the science requirements, our goal for this number is 4.1. Based on the LAE and [O II] galaxy luminosity functions (Gronwall et al. 2007; Ouchi et al. 2008; Ciardullo et al. 2012, 2013) we expect

that, on average, 2.5 sources will be LAEs and 1.6 will be [O II] emitters. (There are sources detected from other lines, but these comprise less than 10% of the full sample.) Figure 21 compares the results of our HDR2 analysis to these goals. From the figure, is it apparent that to reach specification, we need to work down to a signal-to-noise ratio of ~4.8. We also include an estimate of the number of LAEs per IFU based on P(LAE)/P([O II]), i.e., the probability ratio that a detected emission line is Lyα as opposed to [O II] $\lambda3727$. This likelihood is derived primarily by comparing the object's emission-line flux to its broadband flux density, either from the HETDEX spectra itself, or from imaging catalogs (Leung et al. 2017). However, there are multiple additional factors that enter into the probability calculation, including the presence of other emission lines, the
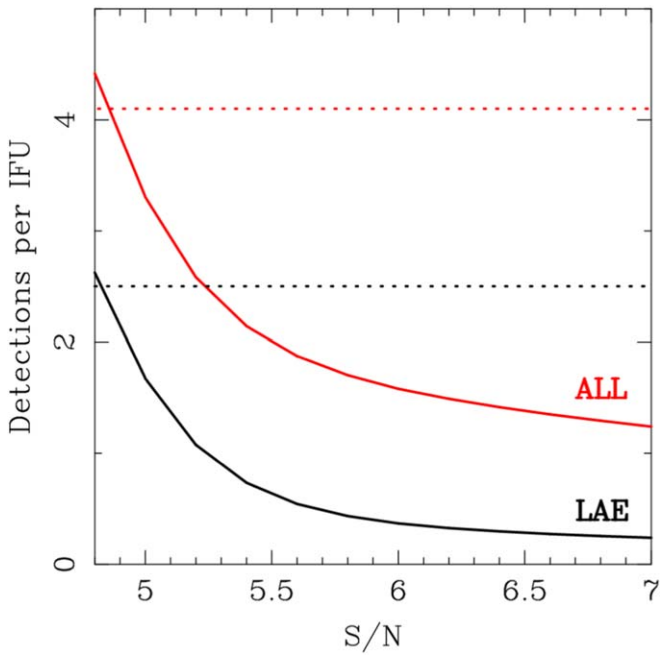
Figure 21. Number of emission line sources per IFU for HDR2 versus the signal-to-noise ratio. The red solid line is for the total number of emission lines in the catalog, and the red dotted line is the required number for the sum of LAEs and [O II] emitters. The black solid line is the current number of LAEs using our automated classification software. The dotted black line is the required number of LAEs. We expect to increase the number of LAEs in the catalog as we study the various cuts that we have applied and refine the classification algorithm. Currently, we are meeting the specification for signal-to-noise ratios >4.8, but small changes to our procedures may result in significant improvements.

size of the source, and the astrometric offset between the emission line and the continuum. These will be discussed in detail in D. Davis et al. (2021, in preparation). In Figure 21 LAEs are identified as having an P(LAE)/P([O II]) >5, but we have not yet performed a statistical analysis for how this ratio affects the rate of contamination. The science requirements specify that no more than 2% of the LAE sample be misclassified [O II] emitters. We will be evaluating the probability ratio with this 2% limit in mind (D. Davis et al. 2021, in preparation).

7.6. False Detections

The target false-positive rate from the science requirements is below 10%. This number is chosen so that the effect of this white noise would be well below the statistical noise. This false-positive rate is included in all cosmological analyses.

There are several ways to estimate the false-positive rate produced by our detection algorithms, including the analysis of inverse frames, use of classical statistical techniques, and repeat observations. The optimal way to quantify false detections is to empirically measure the rate from fields observed more than once, either through serendipitous overlapped regions on sky or through intentionally repeated observations at either nominal depth or deeper. We discuss each method below and find that the overlapping regions provide the best measures since there are a large number of them and they cover different fibers.

The inverse frame technique, which involves multiplying a sky-subtracted frame by −1 and then running the data through the detection algorithms, requires even more attention to the

pixel level defects that have been discussed. Our focus for the pixel-level defects has been on the normal frames, where we maintain a list of pixels or regions that trigger detections at a much higher rate than expected. We have not spent the time listing pixels or regions that trigger detections in the inverse frames. An analysis on the inverse frames is still worthwhile, and we will readdress this issue once the pox issues are mitigated at a better level. The other sources of detections in the inverse frames that differs from the normal frames are faint continuum objects. On an inverse frame, absorption lines turn into emission lines so one must carefully remove all the faint continuum sources before conducting the analysis. We do automatically generate the inverse frames and run detections on these negative spectra. In fact, after going through individual inverse detections and removing those features associated with continuum sources and obvious pixel-level defects, we find a rather small amount of contamination, below the 10% specification. However, we cannot trust this estimate until we refine the inverse frame technique with more recent data. Given the importance of the false-positive rate for cosmology measurements, this technique will be revisited in future analyses.

The classical approach of counting resolution elements across the field and using Gaussian statistics to estimate the number of false positives is also inadequate. The issue here is that the appropriate noise model does not have Gaussian tails, and since many of our objects have a signal-to-noise ratio of 5, understanding the shape of these tails is critical. While this technique does not work globally (that is, defining the noise per resolution over the full field), we do use the noise shape for individual detections in order to help identify real sources. As discussed below, for an individual source with repeat measurements, we generate the signals from all fibers within an aperture and compare the distribution of negative fluxes to that of positive fluxes. Real sources stand out clearly with this local statistical estimator.

At present, our false-positive rate relies primarily on sources with repeat measurements. If a signal-to-noise ratio = 5 source in one field also has a significant detection in another observation in a different part of the array, it is clearly real. While most HETDEX fields are observed only once, there are three sets of fields where we intentionally repeated observations (up to 26 times!). Moreover, there are also several regions of the sky that have overlapping fibers caused by our field placements when the array of IFUs was still incomplete. By crossmatching the one million sources in the HDR2 catalog with the 0.5 billion spectra taken to date, we can identify a significant set of overlapping observations.

There are several fields dedicated to helping determine the false-positive rate. These include deep fields and dedicated repeat fields. In addition, as mentioned above, we also have a set of observations where some of the IFUs overlapped the positions of previous observations. These serendipitous fields dominate in terms of numbers used for the false-positive rate, but all data (i.e., dedicated repeats, deep fields, and serendipitous overlaps) are considered in the analysis.

The deep fields used for the calculation of the false-position rate are in Great Observatories Origins Deep Survey North (GOODS-N). In these two fields, the exposure times are 4 times longer than normal. One of those fields was taken in moderately good conditions (image quality and transparency) and has flux limit 1.8× deeper than most HETDEX data. The

other field had poorer conditions, with a flux limit that is ~1.4× fainter than normal. The number of sources that overlap with these two deep fields is small, around 100, limiting its use for the false-positive rate analysis.

In addition to the deep fields, we targeted several regions of the sky for repeat observations. One of these fields is in the Spitzer/HETDEX Exploratory Large-area survey (Papovich et al. 2016) and has twelve 3×6 min dither sets at the same location (to within a few arcseconds). In addition, four dedicated repeat fields are in Cosmological Evolution Survey (COSMOS; Scoville et al. 2007), each with five 3×6 min dither sets. Each of these visits consists of an observation with the standard HETDEX dither pattern and was reduced independently of the other observations. The result is a set of data that allows us to combine observations and compare a deep summed spectrum to its individual components.

A unique feature of the HETDEX survey is its use of data with large fiber diameters, large gaps between the fibers, moderately poor image quality, and the limited track time of an observation. These effects cause the signal-to-noise of a detection to vary by a factor up to 1.5 depending on the image quality and the exact location relative to the dither pattern. For example, a source observed under good conditions and centered on a fiber will be easier to detect than an object whose center is located between three fibers. As a result, an individual deep exposure is not as valuable as a set of repeat observations where the source position moves around relative to the fiber dither pattern. We test this through input sources simulations and real data, and find sources that can sometimes be detected using the nominal HETDEX exposure time but not on deeper exposure. (Because of the way the HET tracks, it is almost impossible to repeat the exact same observation, with all the fibers at their exact same positions on two different nights.) Thus, it is hard to generate a false-positive rate from a single deep exposure. Thus we do not rely on the deep exposures for our false-positive rates, but we do use these observations for confirmation when possible.

As stated, the areas of serendipitous overlap contribute most to the false-positive study. The main reason is the large numbers of overlapping fibers, allowing us to significantly increase the number of sources in our study. Also, there are advantages associated with the observing conditions, as many of the dedicated repeat fields had strongly varying image quality and sky transparency. In contrast, most of the fields with serendipitous overlap had observing conditions, which were more typical for the HETDEX survey as a whole.

There are about 3000 sources that have been covered by more than 10 dither sets and over 30,000 independent measures of these candidate galaxies. Still, even with this large database, there are difficulties, as a source with signal-to-noise ratio of 4.8 might only appear in a single observation. Thus, we use four methods to confirm whether a source is real. The first method stacks all of the exposures to create a single, extremely deep (up to 9 hr) observation. Line fitting procedures can then be applied to the stacked frame to confirm the existence of an emission line. For about 70% of the 3000 sources, we were able to use this technique to test the reality of a feature flagged by our detection software.

Unfortunately, stacking does not work for all sources due to variations in the observing conditions, the fact that many of the overlapping fibers are near the edges of an IFU, and the issue of luck with the source position (i.e., whether an object is centered
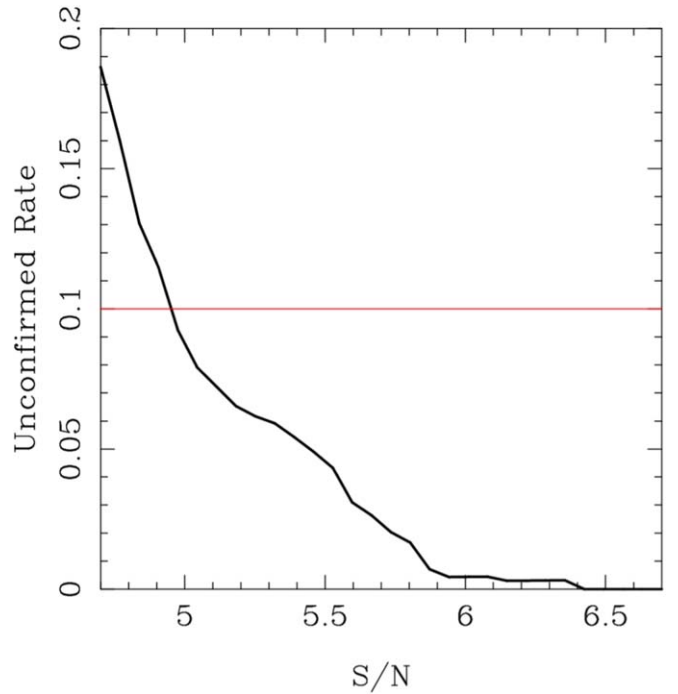
Figure 22. Rate of unconfirmed detections, which is an upper limit to the false-positive rate, from the study of overlapped pointings. Emission-line objects detected with a signal-to-noise ratio above 5.5 are almost always real; it is only when the signal-to-noise ratio drops below 5 that the rate of false positives becomes problematic.

on a fiber or at its edge). Moreover, there are some cases where an emission line is very obvious in one subset of observations but not in the overall stacked spectrum. Thus, a second approach is to confirm the existence of an object by visually inspecting all the spectra of a given source. This method implicitly takes observing conditions into account and uses a reviewer's experience with data to help with the classification.

The third approach, which we performed for about 10% of the sample, is to visually inspect not only the 1D spectrum of each object but also the spectrum produced by each individual fiber in each individual observation. This time-consuming procedure allowed us to better understand how the counts of an emission line are built up and to better judge the noise inherent in the observation.

The fourth approach to the false-detection problem is to compare our HETDEX detections with the results of other surveys that offer higher signal-to-noise ratios and/or extended wavelength coverage. A subset of our data can be investigated in this way, as a number of observations were taken in the COSMOS (Scoville et al. 2007) and GOODS-N (Giavalisco et al. 2004) fields, where a substantial amount of archival data exists. Unfortunately, most of the spectroscopy in this region targeted objects selected by their broadband magnitudes, so the utility of these data for confirming the emission lines of continuum-faint sources is limited. We are in the process of using follow-up observations to help with the confirmation rate.

At the end of our four-tiered process, we created a list of real emission lines from the set of 3000 original detections and plotted the rate of confirmed detections as a function of the signal-to-noise ratio of the detection. This curve is shown in Figure 22 and provides our current best estimate for the false-positive rate. We note that these tests can only confirm whether a source is real; it is extremely difficult to prove that a source is
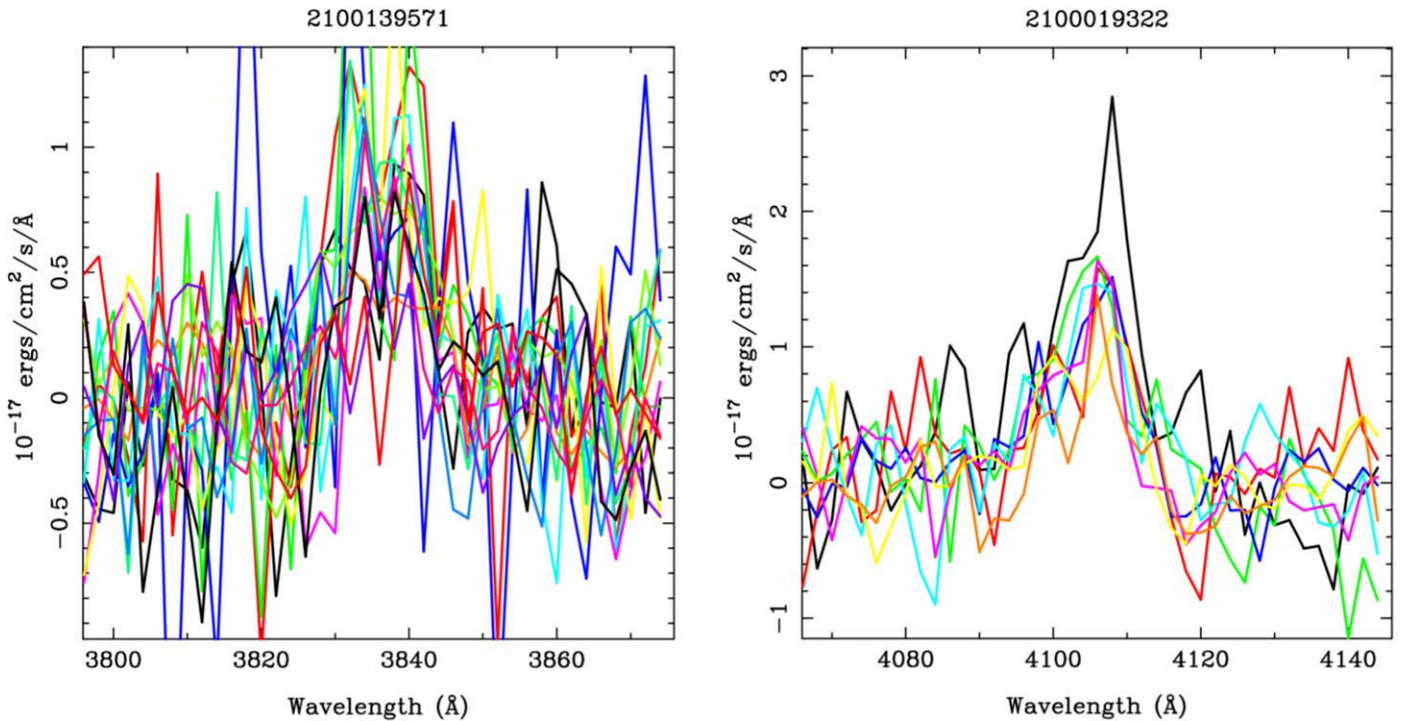
Figure 23. Flux versus wavelength for example detections seen in repeat exposures for two different emission lines. Each color represents a different exposure set. All observing sets are plotted, but only a subset makes it into the catalog.

false. Thus, the curve presented in the figure represents an upper limit to the rate of false positives; we call this measure the rate for unconfirmed sources. As pointed out in Section 3, the HETDEX science requirements state that the fraction of false detections must be less than 10%; as Figure 22 shows, we currently reach this level at a signal-to-noise ratio of around 5.

Figure 23 shows two examples of sources that have repeat observations. The lines with color represent the spectra from the individual observations. All overlapping observations are considered, but only a subset of these make it into the final catalog, since some have too low signal-to-noise ratios. These sources are flagged as LAEs.

We can further examine these data using the spectra of individual fibers. Figure 24 shows an example of a source observed multiple times, where we plot the spectrum of every fiber within the extraction radius. Many of the repeat sources have a distribution similar to that displayed. We use the distribution of central fluxes over all fibers to help determine the statistical significance of the source being real. For this case, the line is obviously real.

There are two pieces of information we obtain from these multiple exposures. First, the summed spectrum provides the signal-to-noise ratio that is necessary to differentiate against false detections. In the case of Figure 24, the source is considered real. Second, the detection probability, compared to that determined from the simulations, confirms the applicability of the simulations for computing our flux limits. For example, for this source shown in the figure, the simulations give a probability of detection of 30%, based on the measured counts in one dither set. Thus we expect to detect this source on about one-third of the 12 dither sets. Note that for the source shown in Figure 24, the emission line is resolved.

As the survey continues, there will be additional dedicated repeat observations and a significant number of additional fibers with overlapping spectra. We will continue to explore the false-positive rate with these on-sky studies.

8. Simulations, Completeness, and Flux Limits

An essential requirement for our cosmological measurements is to understand the completeness rate of a source of a given flux over each spectral and spatial resolution element in the survey. We measure the completeness using input source simulations combined with an approximation of the noise properties for each element.

Ideally, we would run a simulation for each spectral and spatial element in the survey, using a range of input source fluxes. However, the amount of CPU time needed for this type of study is prohibitive and in the end not necessary. Instead, our procedure is to select a set of observations that span a range of observing conditions, run an extensive suite of input source simulations on these data, and measure an approximate recovery rate using the appropriate noise models. This approximation requires a trivial amount of CPU and can be applied to all spectral and spatial elements within an observation.

As our focus is on high-redshift LAEs, our source simulations use point sources; numerous studies have shown that, although Lyα emission from high-redshift sources can be extended, in the vast majority of objects, virtually all the emission fits within the 3″ extraction aperture of the HETDEX observations (e.g., Guaita et al. 2015; Momose et al. 2016; Wisotzki et al. 2016; Ouchi et al. 2020). Our choice of spectral line widths is empirical; we use the average of our detections. Simulations with a range of line widths show that the use of a fixed line width for object detection does not bias our recovery fraction, as long as the true widths are less than ∼20 Å FWHM.

We have identified 20 observations that span the range of observing conditions applicable to HETDEX. These observations form the basis for the simulations and subsequent calibration of
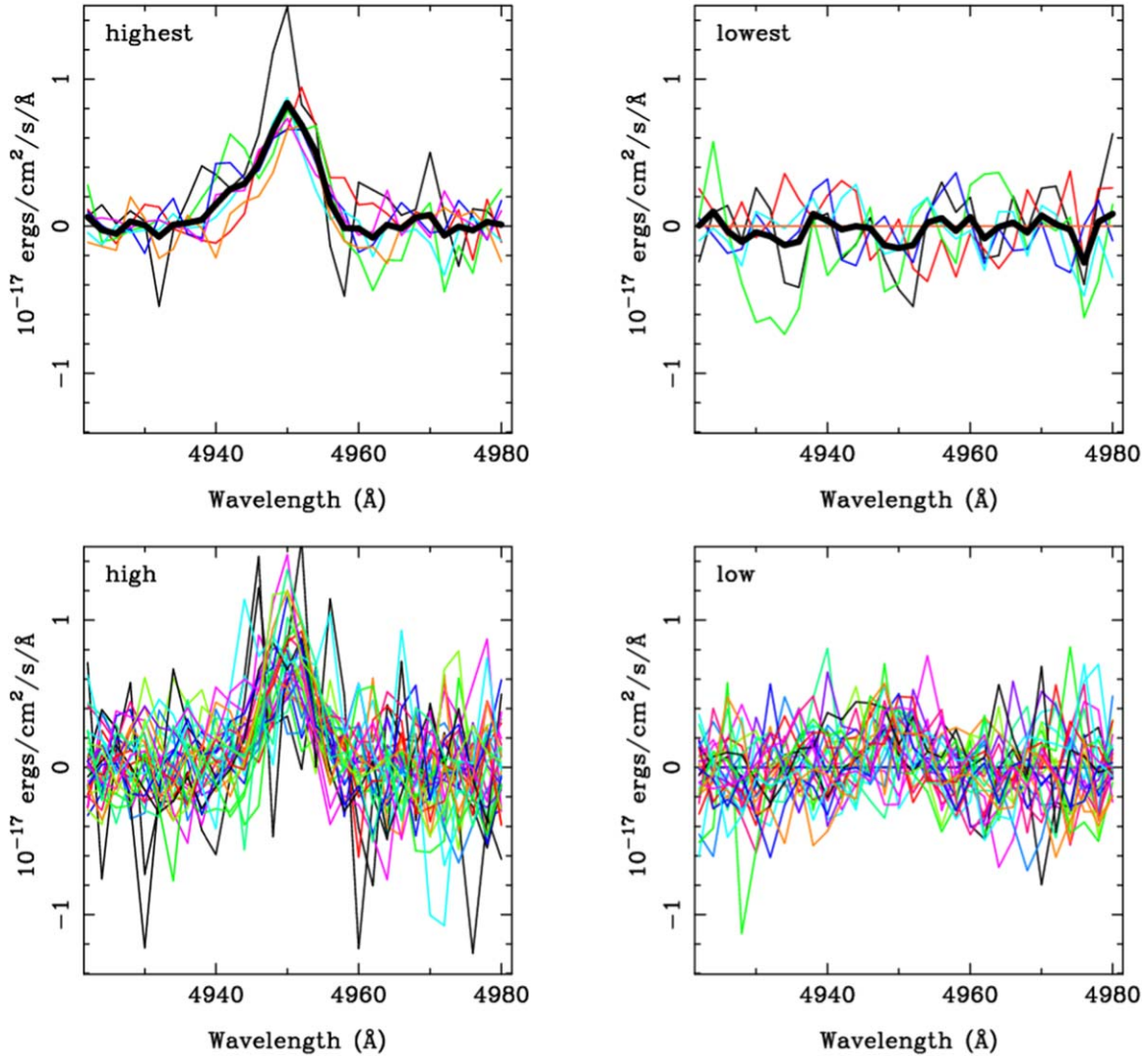
2100080662

Figure 24. Flux versus wavelength for the individual fiber spectra of an LAE with repeat HETDEX observations. Each line in all four panels is a different fiber spectrum. For this source, there are 12 repeated frames, with a total of 78 individual fibers. We sort the fibers by summed flux in the central four bins, with the top left panel containing the seven fibers with the largest flux, the top right panel displaying the seven fibers with the lowest flux, and the bottom panels showing the intermediate cases, with the larger fluxes on the left and the lower fluxes on the right. There are some fibers that are flagged and set to zero, and these are ignored. The thick black lines in the top two panels represent the averaged spectra for those fibers.

the HETDEX flux limits. The image quality of these data ranges from $1\rlap{.}{''}21$ to $2\rlap{.}{''}46$, and the throughput as measured at $4540\,\text{Å}$ varies between 0.105 and 0.184. These ranges span the majority of the data.

8.1. Simulations

We simulate Lyα emitters by adding flux into the actual fiber-spectra database. We then use our HETDEX detection algorithms to generate a mock catalog of the sources. (Of course, since we are adding the flux to the actual data, this catalog will also contain the list of real sources.) Since there are three dithers, each with its own normalization, we take care to match both the PSF and spatial locations for the fibers.

For each IFU, we generate a set of artificial emission lines that linearly span the flux range between 0 and 100×10^{-17} erg cm^{-2} s^{-1} and the wavelength range between 3500 and $5500\,\text{Å}$. To avoid losing objects due to source crowding, we

only add 200 objects at a time to a single frame. (For comparison, a frame will typically contain ∼5 real sources.) We then run the simulation 66 times to produce a sample of 13,200 artificial emission lines per frame. Since HETDEX employs over 70 active IFUs, the simulation of one observation therefore contains close to one million artificial sources.

Once the simulated frames are created, we run the same line-detection algorithms that are used on the real HETDEX data. Any source in the output catalog that is within $2''$ and $3\,\text{Å}$ of an input position is considered to be a match. For each match, we track the difference in the flux, wavelength, spatial location, and signal-to-noise ratio; we do not explore these trends on an IFU by IFU basis but instead analyze the integrated results over all IFUs.

The simulations determine the maximum precision attainable by our HETDEX observations in terms of the wavelength, flux, and completeness. As expected, the wavelength accuracy is a function of the signal-to-noise ratio; for emission lines
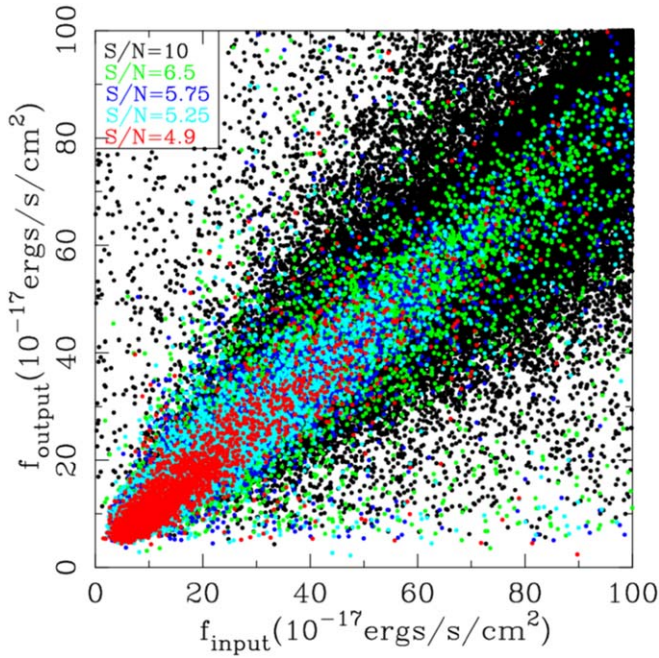
Figure 25. Input versus output fluxes for the million simulated sources in one of the HETDEX science verification fields (20200428v020). The different colors represent the signal-to-noise ratio of the detected source, with central signal-to-noise ratios of 4.9, 5.25, 5.75, 6.5, and 10 in red, light blue, dark blue, green, and black, respectively.

Figure 26. Distribution of input fluxes (solid lines) for a given measured flux (vertical dotted lines). These distributions come from the 20 million simulated sources used in the science verification fields and are horizontal cuts through Figure 25. The curves are all for the same signal-to-noise ratio range of 5.5–6.0; the different colors represent the different measured fluxes, with the dotted lines representing the measured (output) flux. Black represents a measured line flux of 4.8–5.0, red is 6.0–7.0, green is 7.0–9.0, and blue is 11–15, in units of 10^{-17} erg s^{-1} cm^{-2}.

measured with signal-to-noise ratio bins of 4.8–5.2, 5.2–5.5, 5.5–6.5, and >6.5, the recovered wavelengths agree with the input values to within 0.63 Å, 0.56 Å, 0.49 Å, and 0.25 Å, respectively. All are well within the required HETDEX specifications.

Figures 25 and 26 show how our measured emission-line fluxes compare with the input values of the simulation. Once again, we break the data down by the signal-to-noise ratio measured by our detection algorithms. Both figures demonstrate that there is a bias and a scatter in the relation. Figure 26 provides a horizontal cut in Figure 25 at various output fluxes. The vertical dotted lines show the output flux, and the corresponding color shows the distribution of input fluxes that generate that output value. We use these distribution directly when estimating the expected input flux values.

These trends are better seen in Figure 27; the left-hand panel shows the systematic bias in the measured fluxes, while the right-hand figure shows the scatter about these lines, both as a function of the measured flux and measured signal-to-noise ratio. The observed scatter is greater than the statistical error due primarily to uncertainties in the object centroids—with prior knowledge of the centroid, the scatter decreases considerably. We typically measure an input flux to about 27% accuracy.

8.2. Completeness and Flux Limits

The input sources that are not found in the output catalogs define the recovery rate. We track these sources as a function of the input source flux and as a function of the wavelength.

Figure 28 shows the completeness corrections for all 20 of the science verification fields. Each curve has been scaled according to its 50% completeness value. The recovery fractions are remarkably consistent; while each field may have a different depth, and the shapes of the curves vary with

wavelength, at a given wavelength, the curves for all 20 fields are virtually identical. We can therefore average the data to provide the final shape for the completeness function. This is given as the thick black line in the figure. The completeness function at any location can be found from these curves, along with the region's 50% completeness limit.

We estimate the 50% recovery fractions from the noise properties of each spatial and spectral element. It is not feasible to simulate this limit for every spatial and spectral element in our survey, as the observations contain over 10^{11} such elements. Instead, we generate an extracted spectrum based on an observation's PSF and sum the noise using the central four spectral pixels (i.e., a top-hat function). This sum provides the 1σ noise model at a given spatial and spectral position. This approximation is then calibrated against the actual detection algorithm, where we use the source simulations to determine the scaling between the two methods. For example, if we want to determine the recovery curve for a sample using a signal-to-noise ratio cut of 5.0, the 50% completeness value is given by $4.6 \times$ the 1σ noise. The value of 4.6 is determined by comparing the 1σ noise of the four summed spectral pixels to that found from the source simulation's 50% completeness limit. We obtain a smaller value than a simple multiplication of 5.0 since emission lines are fitted to the instrumental spectral profile, as opposed to the top-hat function of the 1σ noise model.

For each IFU, we define a grid from $-26''$ to $+26''$ on $2''$ spacing and measure the flux limit as a function of the wavelength and position, as described above. We call these models flux limit cubes. Edge effects are naturally included since the number of fibers decreases near the IFU edges and,
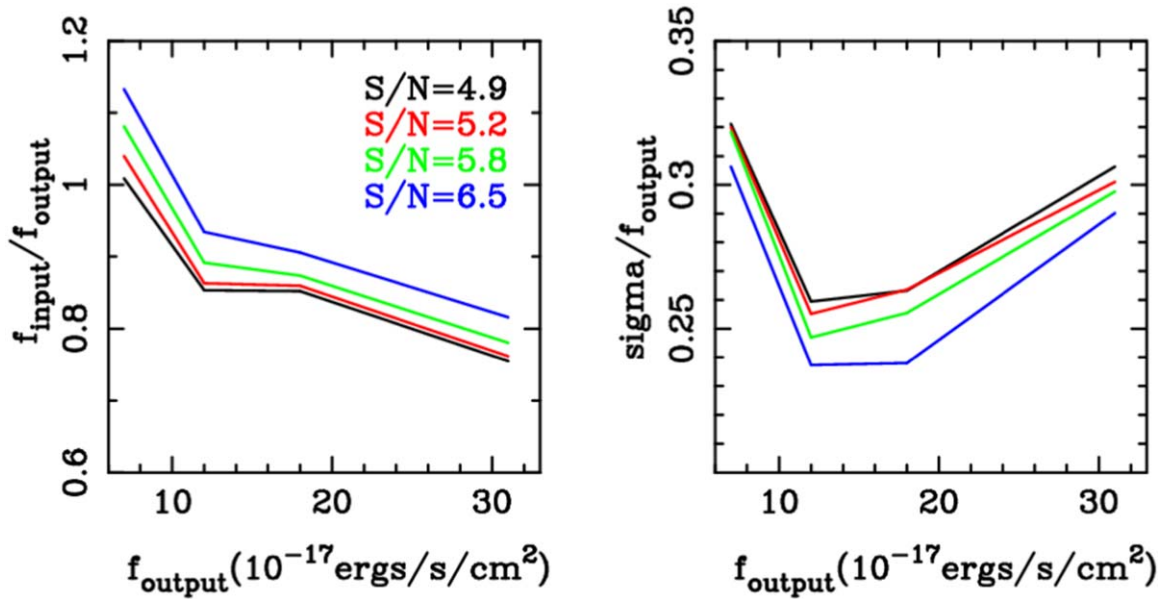
Figure 27. Left: The mean ratio of input flux to the measured flux for our artificial emission-line simulations. The curves represent detections with a signal-to-noise ratio of 4.9 (black), 5.2 (red), 5.8 (green), and 6.5 (blue). We use these curves to correct the output fluxes. Right: The relative scatter about the mean relations shown in the left panel with the same color coding. The dispersion is higher than the statistical uncertainty of the measurements caused (primarily) by uncertainties in the source positions.

hence, the noise there is larger. This technique basically makes a fuzzy IFU, which is primarily determined by the image quality of that observation.

We provide a grid of (R.A., decl.) for each IFU, each with a flux limit curve defined in 2 Å wavelength bins. These grids, combined with the completeness curves, are inputs to the window function estimates used for the cosmological analysis. In addition to the flux limits derived from these grids, we can also measure flux limits at any specific spatial and spectral position. Depending on the accuracy and computing efficiency, we use one or the other.

9. Line Identification

This paper is focused on describing the pipeline used to generate the HETDEX emission line catalog. An important next step is to determine the line identification. Leung et al. (2017) provide our initial basis for line discrimination using a Bayesian analysis with the equivalent width distributions of Lyα and [O II] emitting galaxies. D. Davis et al. (2021, in preparation) and M. Cooper et al. (2021, in preparation) describe the details of the line discrimination and various tests. Line identification relies on multiple pieces of information, including defining an imaging counterpart, the presence (or absence) of additional features in the spectrum, the measurement of the continuum flux density in the HETDEX spectrum or ancillary imaging data, and the use of line expectations based on known distributions of equivalent widths (e.g., Gronwall et al. 2007; Ciardullo et al. 2013).

The primary method for classifying a HETDEX emission line as either Lyα or [O II] λ3727 is through the identification of an imaging counterpart and estimating the line's equivalent width. As discussed in Leung et al. (2017) and Ouchi et al. (2020), a rest-frame equivalent width around 20 Å generally works well as a discriminant, though the optimum cut does change with redshift due to the evolution of both lines' equivalent width distributions (Ciardullo et al. 2012, 2013).

Other lines that can cause confusion in our redshift determinations include Mg II λ2799, [O III] λ5007, C IV λ1550, and C III] λ1909, and these features are considered and included in our line identification software.

Determination of the imaging counterpart also has uncertainties, due to the physical offset of an emission from the position of the host galaxy, astrometric uncertainties for the spatial position of the line, and most importantly, crowding. We study all imaging counterparts within 3″ and apply appropriate weighting given the object size, image quality of the observation, line flux, and wavelength. Fortunately the HETDEX continuum level for a 20 min exposure is around $g = 24$; while this does not allow us to detect the continuum emission from most LAEs, it does enable us to measure the continuum flux density of most nearby galaxies. Thus, HETDEX continuum measurements serve as a strong check on the use of imaging catalogs for determining the counterparts of emission lines.

Our required specification is to have no more than 2% of objects classified as Lyα emitters to be [O II] galaxies; though the other lines have similar criteria, they are not significant contaminants in our survey. Our tests for line identification are presented in D. Davis et al. (2021, in preparation) and M. Cooper et al. (2021, in preparation).

The most secure method of confirming our line identifications is to compare the redshifts of other surveys to HETDEX redshifts. We gather these redshifts from many published catalogs, and COSMOS (Scoville et al. 2007) is particularly useful. These comparisons are discussed in detail in M. Cooper et al. (2021, in preparation). As a summary, we currently are using over 3800 sources with published redshifts. From this initial comparison, the contamination rate of LAEs by [O II] galaxies is 4%. Our target goal is 2%, and this is easily achieved via slight modifications to our algorithm. However, before we apply any changes to our algorithm, we must vet the published spectroscopic catalogs, since those data have their own misclassifications. This work is time consuming and will
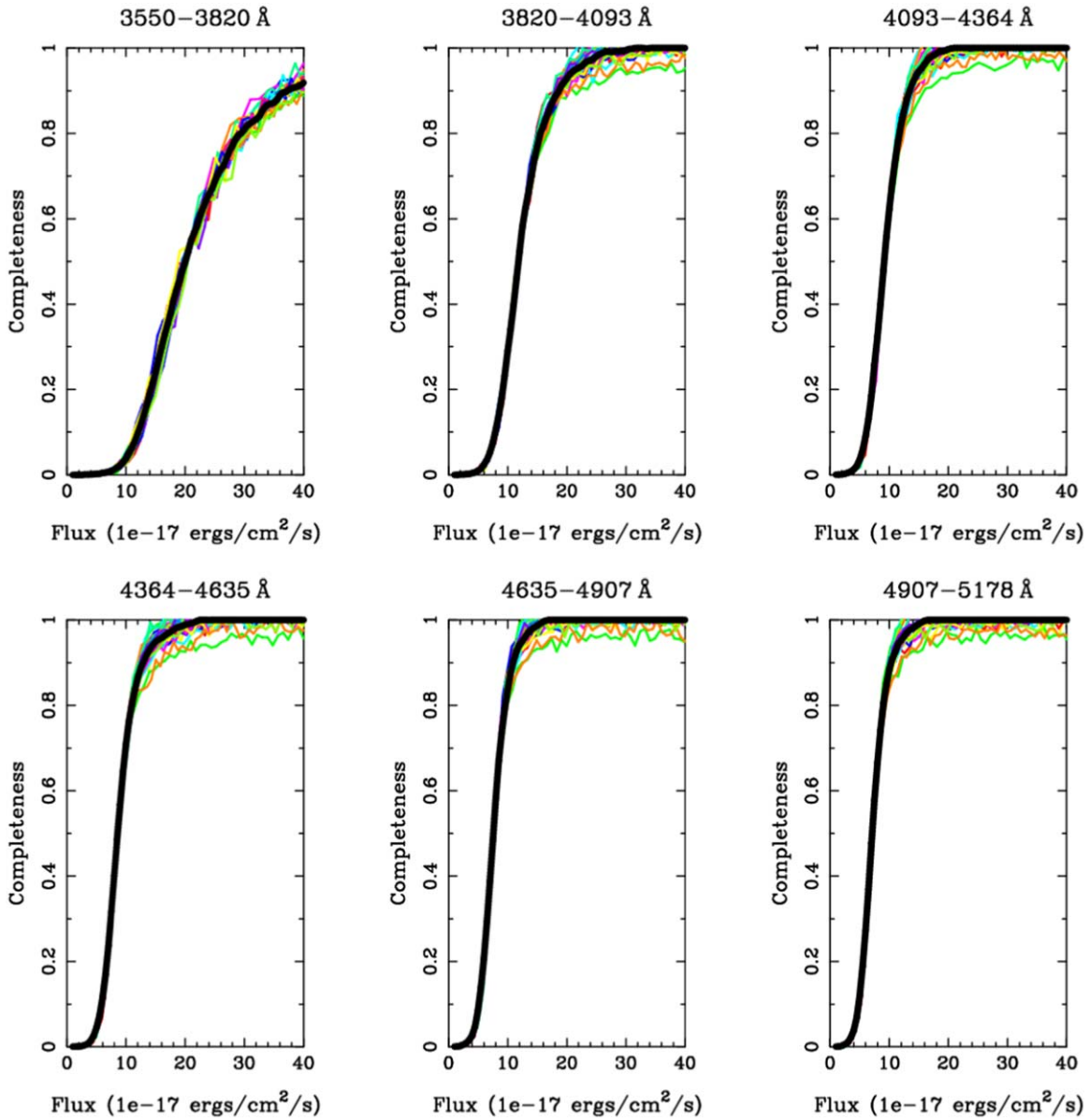
Figure 28. Completeness curves from the simulations, where each panel is for a different wavelength region. The lines with color are from the individual calibration fields, and the thick black line is the average curve used to represent the completeness corrections. In the extreme blue (top left panel), we do not reach 100% completeness until the flux is greater than 40×10^{-17} erg cm^2 s^{-1}.

be published separately (M. Cooper et al. 2021, in preparation). For those sources in agreement (i.e., the 96%) the redshift accuracy is excellent and within our specification of under 180 km s^{-1}.

10. Summary

A primary goal of HETDEX is to study the large-scale clustering of LAEs at $1.88 < z < 3.52$ in order to measure the effect of dark energy. This paper describes the reduction and analysis procedures that will allow us to measure the Hubble expansion rate and angular diameter distance to better than 1% accuracy. This level of precision is comparable with that expected from the ongoing measurements of the late-time expansion rate. Thus, the combination of all these experiments will provide a full picture of any potential evolution on dark energy, with HETDEX providing the high-redshift anchor.

Figure 29 shows redshift slices for the current sample described in this paper (and discussed in M. Cooper et al. 2021, in preparation). The scale is in megaparsec, with the left panel displaying [O II] emitting galaxies out to $z = 0.5$, and the right panel showing the full radial extent of the survey using both the distant LAEs and the nearby [O II] emitters. The (few) galaxies in the redshift gap are objects identified via other emission lines. There are a number of features seen in these redshift slices that fold back into our measurement of clustering. For example, the radial features (the spokes) seen in both panels are due to nonuniform coverage of our observations to date; at the end of survey, we expect these spokes to be filled in. Similarly, the wedge diagrams also contain some regions that are missing data due the presence of large nearby galaxies (such as M101) and the scattered light from very bright stars, such as Mizar, Alcor, and three other second-magnitude stars of the Ursa Major constellation. Obviously, these holes will remain, even
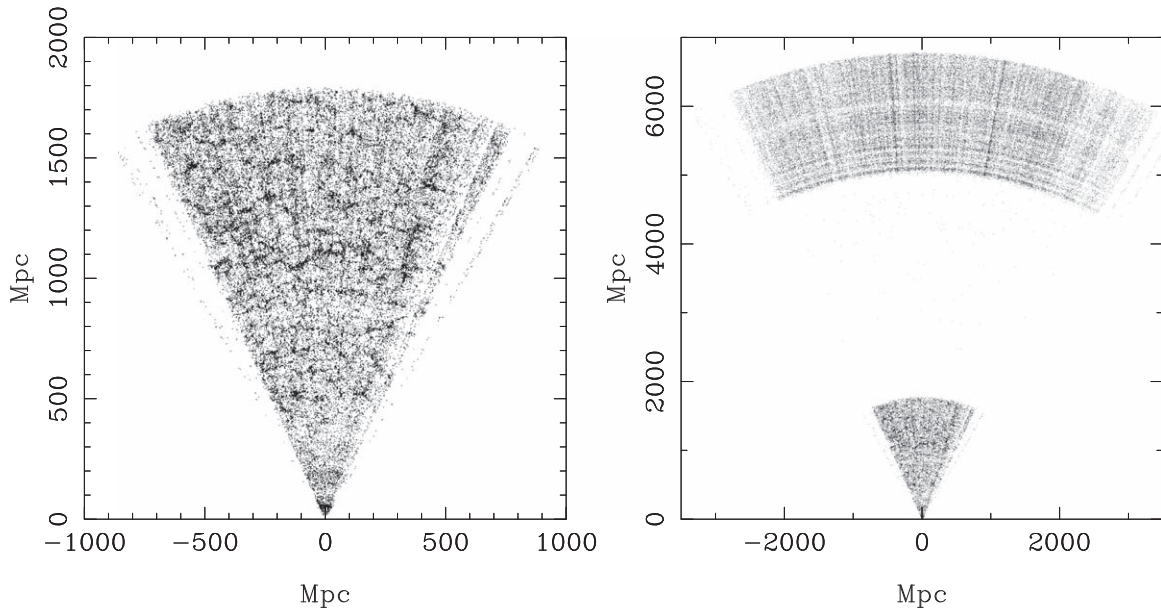
Figure 29. Redshift wedge diagrams for galaxies recorded in HDR2. The left panel displays the distribution of $z < 0.5$ galaxies between $50° < \delta < 52°$, detected principally via their emission at [O II] $\lambda 3727$; the right panel covers the decl. range from $45° < \delta < 55°$ and plots both [O II] emitters with $z < 0.5$ and the LAEs between $1.88 < z < 3.52$. The (few) points in the redshift gap come from the identification of other emission lines. The range in R.A. runs from $155°$ to $245°$. The radial features (over and under densities) are due to the incomplete spatial coverage of HDR2; the angular features (seen easily in the distant galaxies in the right panel) are due to variations in the survey depth due to sky lines. Both of these issues are understood and included in the flux limit cubes.

when the survey is complete. Finally, the angular arc-like features seen in the wedge diagram of the distant universe are due to variations in the survey depth caused by the presence of weak emission lines in the sky. Since the line fluxes of LAEs are generally fainter than those of the [O II] emitters, their detection is more sensitive to the existence of these features. These variations are tracked and calibrated through flux-limit analysis but are not removed in this visualization.

The large-scale clustering of nearby ($z < 0.5$) sources is easily seen in the left panel of Figure 29. The clustering at $z > 1.9$ is less obvious, due to the greater dependence on the observing conditions, the lower number density of LAEs, and the fact that high-z galaxies are less clustered than local objects. Nevertheless, the example demonstrates that by using the flux-limit analysis techniques presented in this paper, we can quantify the clustering of $1.88 < z < 3.52$ LAEs to high precision.

The science requirements to reach our goal are outlined in Section 3. All design elements in the instrumentation and in the data analysis have been coordinated to reach these specifications, and this paper describes the reduction steps. The most difficult aspect is obtaining an adequate number of LAEs while keeping the number of false positives and line misidentifications to a minimum. We show in Figure 21 that the current catalog of emission-line sources reaches our specification. We obviously need to further demonstrate that these sources are indeed LAEs, and that we are finding them at a rate consistent with other studies. Zhang et al. (2021) provides our first LAE luminosity function as measured from HETDEX. While the focus of that paper is on AGN, they include analysis of a large sample of LAEs and demonstrate that the LAE luminosity function is consistent with previously published results. That sample contained 18,000 galaxies, a small fraction of the current catalog, and additional luminosity functions will be published.

For the false-positive rate, the noise model is particularly important and must be well calibrated. We use a combination of repeat observations and simulations for this work, and while additional work is required, Section 7.6 shows that we are close to specification. The line misidentification rate is best studied using known sources or follow-up with other spectrographs. As discussed in Section 9, we are already at 4% contamination, and this number is likely to improve. Additionally, Grasshorn Gebhardt et al. (2019); Awan & Gawiser (2020), and Farrow et al. (2021) outline techniques to measure and/or include the contamination from [O II] from the galaxy power spectrum itself. Our target requirement remains a 3σ detection of the dark energy density at $z = 2.4$, assuming a cosmological constant. There are physical and systematic unknowns in both the LAE properties (namely, the bias value and nonlinear effects) and these could increase (or decrease) the uncertainties on the cosmological parameters. Our primary mitigation strategy against increased uncertainties is observing time. We designed the spring field to allow for a 30% increase in LAE number density, if required. As we obtain more data and improve the analysis, and if a compelling theoretical model for dark energy emerges, we will determine whether additional time is warranted.

ORCID iDs

Karl Gebhardt https://orcid.org/0000-0002-8433-8185
Erin Mentuch Cooper https://orcid.org/0000-0002-2307-0146
Robin Ciardullo https://orcid.org/0000-0002-1328-0211
Viviana Acquaviva https://orcid.org/0000-0002-6788-6315
Ralf Bender https://orcid.org/0000-0001-7179-0626
William P. Bowman https://orcid.org/0000-0003-4381-5245
Dustin Davis https://orcid.org/0000-0002-8925-9769
Roelof S. de Jong https://orcid.org/0000-0001-6982-4081
Niv Drory https://orcid.org/0000-0002-7339-3170
Daniel J. Farrow https://orcid.org/0000-0003-2575-0652
John Feldmeier https://orcid.org/0000-0003-2908-2620
Steven L. Finkelstein https://orcid.org/0000-0001-8519-1130
Cynthia S. Froning https://orcid.org/0000-0001-8499-2892

Eric Gawiser https://orcid.org/0000-0003-1530-8713
Caryl Gronwall https://orcid.org/0000-0001-6842-2371
Gary J. Hill https://orcid.org/0000-0001-6717-7685
Ulrich Hopp https://orcid.org/0000-0003-1008-225X
Steven Janowiecki https://orcid.org/0000-0001-9165-8905
Donghui Jeong https://orcid.org/0000-0002-8434-979X
Eiichiro Komatsu https://orcid.org/0000-0002-0136-2404
Martin Landriau https://orcid.org/0000-0003-1838-8528
Jennifer Marshall https://orcid.org/0000-0003-0710-9474
Ken Mawatari https://orcid.org/0000-0003-4985-0201
Shiro Mukae https://orcid.org/0000-0003-3823-8279
Donald P. Schneider https://orcid.org/0000-0001-7240-7449
Andreas Schulze https://orcid.org/0000-0002-6660-6131
Matthew Shetrone https://orcid.org/0000-0003-0509-2656
Chris Sneden https://orcid.org/0000-0002-3456-5929
Jan Snigula https://orcid.org/0000-0003-4044-5357
Matthias Steinmetz https://orcid.org/0000-0001-6516-7459
Sarah Tuttle https://orcid.org/0000-0002-7327-565X
Tanya Urrutia https://orcid.org/0000-0001-6746-9936
Isak Wold https://orcid.org/0000-0002-0784-1852
Gregory Zeimann https://orcid.org/0000-0003-2307-0629
Yechi Zhang https://orcid.org/0000-0003-3817-8739

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Adachi, S., Aguilar Faúndez, M. A. O., Arnold, K., et al. 2020, ApJ, 904, 65
Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, ApJS, 192, 5
Addison, G. E., Bennett, C. L., Jeong, D., Komatsu, E., & Weiland, J. L. 2019, ApJ, 879, 15
Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
Aiola, S., Calabrese, E., Maurin, L., et al. 2020, JCAP, 2020, 047
Alam, S., Aubert, M., Avila, S., et al. 2021, PhRvD, 103, 083533
Allende Prieto, C., Hubeny, I., & Smith, J. A. 2009, MNRAS, 396, 759
Awan, H., & Gawiser, E. 2020, ApJ, 890, 78
Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2020, MNRAS, 500, 736
Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Bhatti, W. A., Richmond, M. W., Ford, H. C., & Petro, L. D. 2010, ApJS, 186, 233
Bianchi, L., Shiao, B., & Thilker, D. 2017, ApJS, 230, 24
Bohlin, R. C., Gordon, K. D., & Tremblay, P. E. 2014, PASP, 126, 711
Bond, N. A., Gawiser, E., Gronwall, C., et al. 2009, ApJ, 705, 639
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Byrohl, C., Saito, S., & Behrens, C. 2019, MNRAS, 489, 3472
Capitanio, L., Lallement, R., Vergely, J. L., Elyajouri, M., & Monreal-Ibero, A. 2017, A&A, 606, A65
Castelli, F., & Kurucz, R. L. 2003, in IAU Symp. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray (Cambridge: Cambridge Univ. Press), A20
Cenarro, A. J., Peletier, R. F., Sánchez-Blázquez, P., et al. 2007, MNRAS, 374, 664
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chiang, C.-T., Wullstein, P., Jeong, D., et al. 2013, JCAP, 2013, 030
Ciardullo, R., Gronwall, C., Adams, J. J., et al. 2013, ApJ, 769, 83
Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, ApJ, 744, 110
Coupon, J., Ilbert, O., Kilbinger, M., et al. 2009, A&A, 500, 981
Cowie, L. L., & Hu, E. M. 1998, AJ, 115, 1319
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
DES Collaboration, Abbott, T. M. C., Aguena, M., et al. 2021, arXiv:2105.13549
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
de Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2020, MNRAS, 501, 5616

Di Valentino, E. 2021, MNRAS, 502, 2065
Di Valentino, E., Mena, O., Pan, S., et al. 2021, arXiv:2103.01183
Doi, M., Tanaka, M., Fukugita, M., et al. 2010, AJ, 139, 1628
du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, ApJ, 901, 153
Dutcher, D., Balkenhol, L., Ade, P. A. R., et al. 2021, PhRvD, 104, 022003
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Farrow, D. J., Sánchez, A. G., Ciardullo, R., et al. 2021, arXiv:2104.04613
Filippenko, A. V. 1982, PASP, 94, 715
Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, ApJS, 251, 7
Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34
Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gawiser, E., Francke, H., Lai, K., et al. 2007, ApJ, 671, 278
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
Gil-Marín, H., Bautista, J. E., Paviot, R., et al. 2020, MNRAS, 498, 2492
Grasshorn Gebhardt, H. S., Jeong, D., Awan, H., et al. 2019, ApJ, 876, 32
Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2014, ApJ, 783, 114
Gronwall, C., Ciardullo, R., Hickey, T., et al. 2007, ApJ, 667, 79
Guaita, L., Gawiser, E., Padilla, N., et al. 2010, ApJ, 714, 255
Guaita, L., Melinder, J., Hayes, M., et al. 2015, A&A, 576, A51
Gurung-López, S., Saito, S., Baugh, C. M., et al. 2021, MNRAS, 500, 603
Hawkins, K., Zeimann, G., Sneden, C., et al. 2021, ApJ, 911, 108
Hayes, D. S., & Latham, D. W. 1975, ApJ, 197, 593
Herenz, E. C., & Wisotzki, L. 2017, A&A, 602, A111
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hill, G. J. & HETDEX Consortium 2016, in ASP Conf. Ser. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields, ed. I. Skillen, M. Balcells, & S. Trager (San Francisco, CA: ASP), 393
Horne, K. 1986, PASP, 98, 609
Indahl, B. L., Hill, G. J., Drory, N., et al. 2016, Proc. SPIE, 9908, 990880
Jeong, D., & Komatsu, E. 2006, ApJ, 651, 619
Jeong, D., & Komatsu, E. 2009, ApJ, 691, 569
Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
Kelz, A., Jahn, T., Haynes, D., et al. 2014, Proc. SPIE, 9147, 914775
Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2019, MNRAS, 489, 555
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Komatsu, E., Bennett, C. L., Barnes, C., et al. 2014, PTEP, 2014, 06B102
Kurucz, R. L. 1970, SAOSR, 309
Kurucz, R. L. 1979, ApJS, 40, 1
Kurucz, R. L., Furenlid, I., Brault, J., & Testerman, L. 1984, Solar Flux Atlas from 296 to 1300 nm (Sunspot, NM: National Solar Observatory)
Kusakabe, H., Shimasaku, K., Ouchi, M., et al. 2018, PASJ, 70, 4
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Lemaux, B. C., Fuller, S., Bradač, M., et al. 2021, MNRAS, 504, 3662
Leung, A. S., Acquaviva, V., Gawiser, E., et al. 2017, ApJ, 843, 130

Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Massey, P., Strobel, K., Barnes, J. V., & Anderson, E. 1988, ApJ, 328, 315
McCullagh, N., Jeong, D., & Szalay, A. S. 2016, MNRAS, 455, 2945
Momose, R., Ouchi, M., Nakajima, K., et al. 2016, MNRAS, 457, 2318
Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682
Muzahid, S., Schaye, J., Marino, R. A., et al. 2020, MNRAS, 496, 1013
Narayan, G., Matheson, T., Saha, A., et al. 2019, ApJS, 241, 20
Oke, J. B. 1990, AJ, 99, 1621
Ouchi, M., Ono, Y., & Shibuya, T. 2020, ARA&A, 58, 617
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301
Padmanabhan, N., Schlegel, D. J., Finkbeiner, D. P., et al. 2008, ApJ, 674, 1217
Papovich, C., Shipley, H. V., Mehrtens, N., et al. 2016, ApJS, 224, 28
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Planck Collaboration VI 2020, A&A, 641, A6
Pullen, A. R., Hirata, C. M., Doré, O., & Raccanelli, A. 2016, PASJ, 68, 12
Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, ApJL, 545, L85
Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Roddier, F. 1981, PrOpt, 19, 281
Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440
Runnholm, A., Gronke, M., & Hayes, M. 2021, PASP, 133, 034507
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, ApJ, 788, 74
Shoji, M., Jeong, D., & Komatsu, E. 2009, ApJ, 693, 1404
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Stetson, P. B. 1987, PASP, 99, 191
Stetson, P. B. 1990, PASP, 102, 932
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
Tegmark, M. 1996, ApJL, 470, L81
Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, ApJ, 809, 89
Vargas-Magana, M., Brooks, D. D., Levi, M. M., & Tarle, G. G. 2019, arXiv:1901.01581
Wahba, G. 1990, in CBMS-NSF Regional Conf. Ser. in Applied Mathematics (Philadelphia, PA: Society for Industrial and Applied Mathematics)
Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98
Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2017, MNRAS, 467, 1890
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
York, D. G., Adelman, J., Anderson, J. E., et al. 2000, AJ, 120, 1579
Zhang, Y., Ouchi, M., Gebhardt, K., et al. 2021, arXiv:2105.11497