

Fermilab

Clustering analysis of medium-band selected high-redshift galaxies

FERMILAB-PUB-25-0724-PPD

arXiv:2509.26467

This manuscript has been authored by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Clustering analysis of medium-band selected high-redshift galaxies

H. Ebina^{1,2,3} M. White^{1,2,3} A. Raichoor³ Arjun Dey⁴
D. Schlegel³ D. Lang⁵ Y. Luo⁶ J. Aguilar³ S. Ahlen⁷
A. Anand³ D. Bianchi^{8,9} D. Brooks¹⁰ F. J. Castander^{11,12}
T. Claybaugh³ A. Cuceu³ K. S. Dawson¹³ A. de la
Macorra¹⁴ Biprateep Dey^{15,16} P. Doel¹⁰ S. Ferraro^{3,1}
A. Font-Ribera¹⁷ J. E. Forero-Romero^{18,19}
E. Gaztañaga^{11,20,12} S. Gontcho A Gontcho^{3,21} G. Gutierrez²²
H. K. Herrera-Alcantar^{23,24} C. Howlett²⁵ M. Ishak²⁶
R. Joyce⁴ R. Kehoe²⁷ D. Kirkby²⁸ T. Kisner³ A. Kremin³
O. Lahav¹⁰ A. Lambert³ M. Landriau³ L. Le Guillou²⁹
C. Magneville²⁴ M. Manera^{30,17} P. Martini^{31,32,33} A. Meisner⁴
R. Miquel^{34,17} J. Moustakas³⁵ E. Mueller³⁶ S. Nadathur²⁰
N. Palanque-Delabrouille^{24,3} W. J. Percival^{37,5,38}
C. Poppett^{3,39,1} F. Prada⁴⁰ I. Pérez-Ràfols⁴¹ G. Rossi⁴²
E. Sanchez⁴³ M. Schubnell^{44,45} J. Silber³ D. Sprayberry⁴
G. Tarlé⁴⁵ B. A. Weaver⁴ C. Yèche²⁴ R. Zhou³ H. Zou⁴⁶

¹University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, USA

²Department of Physics, University of California, Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, USA

³Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

⁴NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA

⁵Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada

⁶Department of Physics & Astronomy, University of Wyoming, 1000 E. University, Dept. 3905, Laramie, WY 82071, USA

⁷Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215 USA

⁸Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy

⁹INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy

¹⁰Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

- ¹¹Institut d'Estudis Espacials de Catalunya (IEEC), c/ Esteve Terradas 1, Edifici RDIT, Campus PMT-UPC, 08860 Castelldefels, Spain
- ¹²Institute of Space Sciences, ICE-CSIC, Campus UAB, Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain
- ¹³Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA
- ¹⁴Instituto de Física, Universidad Nacional Autónoma de México, Circuito de la Investigación Científica, Ciudad Universitaria, Cd. de México C. P. 04510, México
- ¹⁵Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
- ¹⁶Department of Physics & Astronomy and Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260, USA
- ¹⁷Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Edifici Cn, Campus UAB, 08193, Bellaterra (Barcelona), Spain
- ¹⁸Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia
- ¹⁹Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia
- ²⁰Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK
- ²¹University of Virginia, Department of Astronomy, Charlottesville, VA 22904, USA
- ²²Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA
- ²³Institut d'Astrophysique de Paris. 98 bis boulevard Arago. 75014 Paris, France
- ²⁴IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
- ²⁵School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia
- ²⁶Department of Physics, The University of Texas at Dallas, 800 W. Campbell Rd., Richardson, TX 75080, USA
- ²⁷Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, USA
- ²⁸Department of Physics and Astronomy, University of California, Irvine, 92697, USA
- ²⁹Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France
- ³⁰Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
- ³¹Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA
- ³²Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, USA
- ³³The Ohio State University, Columbus, 43210 OH, USA
- ³⁴Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, 08010 Barcelona, Spain
- ³⁵Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA

- ³⁶Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, U.K
- ³⁷Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
- ³⁸Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
- ³⁹Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, CA 94720, USA
- ⁴⁰Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain
- ⁴¹Departament de Física, EEBE, Universitat Politècnica de Catalunya, c/Eduard Maristany 10, 08930 Barcelona, Spain
- ⁴²Department of Physics and Astronomy, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea
- ⁴³CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain
- ⁴⁴Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109, USA
- ⁴⁵University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, USA
- ⁴⁶National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, P. R. China

E-mail: ebina@berkeley.edu

Abstract. Next-generation large-scale structure spectroscopic surveys will probe cosmology at high redshifts ($2.3 < z < 3.5$), relying on abundant galaxy tracers such as Ly α emitters (LAEs) and Lyman break galaxies (LBGs). Medium-band photometry has emerged as a potential technique for efficiently selecting these high-redshift galaxies. In this work, we present clustering analysis of medium-band selected galaxies at high redshift, utilizing photometric data from the Intermediate Band Imaging Survey (IBIS) and spectroscopic data from the Dark Energy Spectroscopic Instrument (DESI). We interpret the clustering of such samples using both Halo Occupation Distribution (HOD) modeling and a perturbation theory description of large-scale structure. Our modeling indicates that the current target sample is composed from an overlapping mixture of LAEs and LBGs with emission lines. Despite differences in target selection, we find that the clustering properties are consistent with previous studies, with correlation lengths $r_0 \simeq 3 - 4 h^{-1} \text{Mpc}$ and a linear bias of $b \sim 1.8 - 2.5$. Finally, we discuss the simulation requirements implied by these measurements and demonstrate that the properties of the samples would make them excellent targets to enhance our understanding of the high- z universe.

Contents

1	Introduction	1
2	Tracers at high redshift	3
2.1	Lyman alpha emitters	3
2.2	Lyman break galaxies	4
3	Data	5
3.1	Imaging data	5
3.2	Target selection	5
3.3	DESI spectroscopy	9
3.4	Overlap with traditional tracers	10
4	Clustering analysis	11
4.1	Angular clustering	11
4.2	HOD fit	14
5	Inferences from simulations	16
5.1	Mock Clustering	17
5.2	Forecast	21
6	Conclusion	23
7	Data Availability	24
A	Relevant scales	25
B	Fiber assignment	28
C	Angular clustering in medium-band selected samples	29
D	Extended models of HOD	31

1 Introduction

The large-scale structure (LSS) of the universe can be used as a probe of fundamental physics, cosmology and galaxy formation containing as it does information from inflation, cosmic expansion, gravitational evolution and the relationship between galaxies and the cosmic web. Spectroscopic LSS surveys, in particular, precisely map galaxies in the radial direction, allowing for 3D clustering as a function of redshift. This allows observers to collect information of more modes than in 2D and gain access to the dynamics of the universe at late times.

The current generation of (“Stage IV”) spectroscopic surveys measure 3D galaxy clustering over a significant fraction of the extragalactic sky approaching sample variance precision on quasi-linear scales up to redshift $\lesssim 2$ [1–3]. Measuring millions of galaxy and quasar spectra [4], these surveys enable a precise measurement of galaxy clustering and e.g. sub-percent

constraints on the expansion history of the Universe [5]. With rapidly improving detector technologies, the precision cosmology frontier can be advanced to even higher redshift, providing improvements in multiple directions simultaneously [6–10]. DESI-II, a proposed successor experiment to DESI probing galaxies with redshifts up to ~ 3.5 , will be the first step towards a next-generation spectroscopic instrument designed to efficiently map the high redshift Universe [8, 10, 11].

Moving beyond $z \sim 2$ requires us to trace LSS with new galaxy tracer populations. The LSS probes usually used at $z \lesssim 2$, namely Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs; including H α galaxies), and Quasars (QSOs) become sparse past $z \simeq 2$ [1, 2]. The number of bright QSOs that can be used to probe the intergalactic medium also declines. Rather, promising tracers at high redshifts are two sets of galaxies: Ly α emitters (LAE; [12]) and Lyman break galaxies (LBG; [13–17]). These galaxy tracers are defined by Ly α emission and the Lyman break feature in the galaxy spectra (respectively) and have been readily observed to $z \gtrsim 5$ [15, 18–21], supplying targets for the full redshift extent of next-generation (“Stage V”) LSS surveys.

Existing surveys have already demonstrated the feasibility of using these galaxies as cosmological tracers. Narrow-band imaging with Subaru Suprime-Cam [18], Hyper Suprime-Cam in the SILVERRUSH program [22] and, most recently, the One-hundred-square-degree DECam Imaging in Narrowbands (ODIN) survey [23] detects LAEs out to $z \sim 7$. Spectroscopic observations with the VIMOS VLT Deep Survey (VVDS) [24], Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) [25] and, over the ODIN fields with DESI [1, 26] has confirmed thousands of these sources. For LBGs, wide-field, broadband imaging that employs the “dropout” technique – e.g. CFHTLS-Archive-Research Survey (CARS) [27] and Subaru GOLDRUSH [15] – identifies sources to $z \sim 7$, while spectroscopic programs such as VIMOS Ultra-Deep Survey (VUDS) [28] and, more recently, DESI [29] have obtained secure redshifts for many of them at $z \lesssim 6$. Although these tracers have been defined through their distinct observational features, they are physically similar objects across a wide range of physics. LAEs and LBGs form a continuum of Ly α rest-frame equivalent width (REW), with LAEs populating the high-REW end of the continuum [12, 14].

In this work, we report observational results from the DESI ancillary campaign in 2024B, which provided spectroscopic follow-up of medium-band selected high-redshift galaxies. The spectroscopic data provide information on the interloper fraction and redshift distribution of galaxies that can be used to analyze the clustering of the photometrically selected ‘target’ galaxy samples. These findings not only establish direct constraints on clustering statistics and offer valuable observational insights, but also represent a key step in developing the forecasting and analytical infrastructure necessary for continued high-redshift investigations. This work is done concurrently with related work that together aims to characterize tracers relevant for future high- z surveys. Refs. [29, 30] have performed investigation of selection and clustering on broad-band selected LBGs, ref. [31] investigates the medium-band selection of high- z galaxies using Subaru Suprime-Cam, and there is ongoing work characterizing the astrophysical properties of LAEs using DESI data.

The present paper is organized as follows. First in §2 we will describe the high-redshift tracers. In §3, we will briefly describe the photometric and spectroscopic datasets used for the galaxy catalog in this work and use them to characterize where the sample stands with respect to traditional selections. In §4 we describe the clustering statistics that we use for analysis including an HOD-based analyses of these samples. The mock catalogs populated with representative HODs are then used to study the properties of these tracers (§5.1) and

to provide forecasts for future prospects (§5.2). Finally, we conclude in §6. A series of appendices provide supporting details. Throughout we shall use comoving $h^{-1}\text{Mpc}$ units, unless otherwise specified, assume a ΛCDM cosmology using the best-fit values from *Planck* [32], and quote all magnitudes in the AB system corrected for dust using the SFD98 [33] maps.

2 Tracers at high redshift

In this section we describe the galaxy tracers that are populous at high redshifts, of interest for next-generation experiments. The tracers come in two categories: Ly α emitters (LAE [34]) and Lyman break galaxies (LBG [35]). Both tracers have been extensively studied previously [6, 13–15, 17–21, 27, 36–39], however in the cosmology community they are relatively unfamiliar due to the expense of obtaining redshifts of faint galaxies over large volumes to pursue classical LSS studies. Recently, however, photometrically selected LBGs, in combination with CMB lensing, have been used to make a measurement of the growth of structure at high redshift [40].

2.1 Lyman alpha emitters

LAEs are galaxies characterized by their strong Ly α emission line (at rest-frame 1216Å) often with low continuum levels, with criteria frequently placed on the minimum rest-frame equivalent width (REW) of the Ly α line. Although this designation is primarily observational, LAEs have been physically identified to be young, low-mass ($M_{\star} \sim 10^{8-9}M_{\odot}$), actively star-forming ($\text{SFR} \sim 1 - 10 M_{\odot} \text{yr}^{-1}$) galaxies [12]. Physically LAEs are a subset of the LBG population, however due to the differences in selection and magnitude/mass ranges probed they are often considered to be a separate class.

LAEs are well motivated density tracers for next-generation large-scale structure experiments for multiple reasons. First, astronomy surveys have identified populations of LAEs reaching beyond the redshift range of Stage V spectroscopic experiments ($z \lesssim 5$) [41], with high enough density to enable precise clustering measurements over most of the linear and quasi-linear regimes [42]. Second, the spectra of LAEs, with a single emission line on an otherwise featureless spectrum, provide relatively simple targets to select photometrically and obtain spectroscopic redshifts. By selecting for a single big feature in the galaxy spectrum, it is relatively easy to make a simple, yet effective, photometric selection of LAEs, which can limit the effect of systematics in the clustering signal. In the astronomy context, most color cuts have been defined using a combination of deep narrow (or medium, although less popular) bands and broad-bands [21, 41]. These selections primarily focused on obtaining low interloper samples with enough sample size for a statistical analysis of galaxies. Recent work from the ODIN survey [23] has followed this approach, as a low interloper sample is important for angular clustering analyses [26, 43]. For future spectroscopic cosmology surveys, however, it will be important to cover more volume (i.e. wider redshift ranges) at the expense of some interlopers, which can be eliminated by spectroscopic follow-up. In addition to simply using medium bands in combination with broad bands, recent work using the Subaru Suprime dataset [31] have shown that some selections require only (a closely distributed set of) medium bands.

However, study of LAE in the context of cosmology introduce non-negligible complications as well. One straightforward issue is the non-trivial overlap between traditional LAEs using (deep) narrow bands and LAEs optimal for next-generation cosmology surveys, using

(shallow) medium bands. While both types of selections are designed to target similar-shaped spectra, there has been no direct indication that the different methods are selecting identical galaxies. In particular, this complicates interpretation of past LAE studies.

Another subtlety, perhaps more complicated, is the radiative transfer (RT) of Ly α photons through HI in the galaxy environment. Ref. [44] suggested that RT can significantly distort target selections depending on local overdensity and LOS velocity divergence, while later work [45] argued that the magnitude of this effect was largely an artifact of limited simulation resolution. Recent work [46] has also shown that RT can significantly suppress LAEs that are central galaxies in massive ($M_h \gtrsim 10^{12} M_\odot$) halos. While studies have not yet measured this RT effect, nor have they settled on a complete model, the impact of HI absorption on LAE clustering has been detected [47]. Simple Fisher forecasts suggest that the impact of a simple RT model can be limited by using a multi-tracer cosmology analysis [42]. Further investigation is necessary to fully assess the influence of RT on future cosmology surveys.

2.2 Lyman break galaxies

LBGs, also referred to as “dropout” galaxies, are a population of galaxies characterized by the Lyman break (rest-frame 912Å) in the galaxy UV continuum, which arises from the spectral absorption from neutral hydrogen in stellar atmospheres and the interstellar medium [13, 14, 16]. For galaxies with redshift $z \gtrsim 2$, the Lyman break feature is broadened towards 1216Å due to absorption by neutral Hydrogen in the IGM (i.e. ‘line blanketing’ from the Ly α forest) [6, 14]. One then searches for a large drop in flux between the ‘dropout’ band (containing $(1+z)1000\text{Å}$) and the ‘detection’ band (containing $(1+z)1500\text{Å}$) with an almost flat continuum to redder wavelengths. As with the LAEs, the definition is primarily observational but it effectively selects massive, actively star-forming galaxies at high redshift. We refer the reader to refs. [13, 14, 16] for a more in-depth and broad study of LBGs in an astronomical context.

As with the LAEs, there are strong observational reasons to consider LBGs as a central tracer candidate in future high-redshift cosmology experiments. First, LBGs have been confirmed to populate the entire redshift range for potential cosmology experiments in the next decade and beyond, with candidates up to $z \sim 9$ discovered with the *James Webb Space Telescope* [48]. At the redshifts of interest to next-generation experiments, LBGs will provide sufficient densities to conduct full-shape 3D power spectrum analyses to quasi-linear regimes [6]. LBGs have also been found to have significantly stronger clustering than LAEs, providing a potential opportunity to break degeneracies using the different parameter dependencies (“multi-tracer” cosmology) [6, 42]. For parameters that benefit from sample variance cancellation, such as local-type primordial non-Gaussianity, f_{NL}^{loc} , having a high bias tracer and a second tracer with lower bias significantly improves the constraints, by a factor of ≈ 2 or greater [30, 49].

Traditionally, these galaxies are selected in broadband color spaces using three adjacent broadbands with the “drop” in the first two bands and close-to-zero color in the second two bands (with refs. [15, 20, 29, 50] providing recent examples). However, using medium-band filters from the Suprime-Cam [51] on the Subaru Telescope [52, 53], it has recently been demonstrated that it is possible to select these galaxies in the absence of broadbands [31]. This has partly motivated our study.

3 Data

In this section we provide an overview of the high redshift galaxy sample that we employ in this work. The main drivers are the Intermediate Band Imaging Survey (IBIS) [54] and Dark Energy Spectroscopic Instrument (DESI) [1], providing respectively the bulk of the imaging and spectroscopic data. The imaging data is also supplemented by the public Subaru Hyper-Prime Cam (HSC) Strategic Survey Program (SSP) [55] dataset.

3.1 Imaging data

The imaging data used for target selections is mainly from the medium-band imaging survey IBIS (co-PIs: A. Dey & D. Schlegel, NOAO Proposal # 2023B-184194) on the Dark Energy Camera (DECam) [56] at the Blanco telescope in Chile, with assistance from broadband photometry in the HSC-wide survey. IBIS is a wide-field, medium-band imaging survey, that is planned to cover 3000 deg^2 [54]. The observations started in 2024A¹, with the two deep fields (XMM-LSS and COSMOS, each of $\sim 10 \text{ deg}^2$) completed in 2025A. Each deep field is obtained via 92 dithered exposures in each band, achieving homogeneous depth inside a disk of radius $\sim 1.6^\circ$. For this work, we consider the dataset from the XMM-LSS field, which was completed and spectroscopically followed up (§3.3) in 2024B.

As shown in Figure 1, the filter set of IBIS consists of 5 adjacent medium bands, spanning roughly $4000 < \lambda < 5300 \text{ \AA}$, each with a width of $\sim 260 \text{ \AA}$. In the XMM-LSS field, IBIS provides medium band imaging to a magnitude of 25.5 (5σ PSF depth). Of the total XMM-LSS field, we restrict the sample to a 1.4 deg radius to reduce systematics from lower depth in the photometry near the field edges. When accounting for area lost due to masked pixels, bad data or bright object avoidance, the field covers 6.11 deg^2 . We complement these medium-band data with forced-photometry broadband imaging from the HSC-SSP wide field, which reaches depths of $g = 26.5$, $r = 26.0$, $i = 26.5$, and $z = 25.5$ [55].

The data reduction and photometry is performed with the `legacypipe`² [57] pipeline and the photometry is performed with the `Tractor` software³ [58]. For each source the properties, such as position and flux, are obtained by χ^2 minimization of the model images over the relevant region. `Tractor` produces both “total” flux and a “fiber” flux for each source, with the latter being the predicted flux within a $1.5''$ diameter for $1''$ Gaussian seeing. The fiber fluxes approximate the typical amount of light that a DESI fiber would see, and thus are highly correlated with the SNR.

3.2 Target selection

As this is the first time we selected galaxies from the IBIS medium bands, the target selection in this work is preliminary, using a combination of several novel techniques using the consecutive configurations of medium bands. The goal is to have a selection optimized towards the next wide-field spectroscopic survey, likely DESI-II, with the current selections aiming towards exploring the parameter space rather than a final tuning for efficiency. Based on ongoing investigation, we believe that we can increase the size of the target sample through improved selection techniques. In the meantime, an analysis of these objects provides important information for future planning.

¹The observations are traditionally described in the ‘A’ and ‘B’ semesters, covering January to June and July to December, respectively.

²<https://github.com/legacysurvey/legacypipe>

³<https://github.com/dstndstn/tractor>

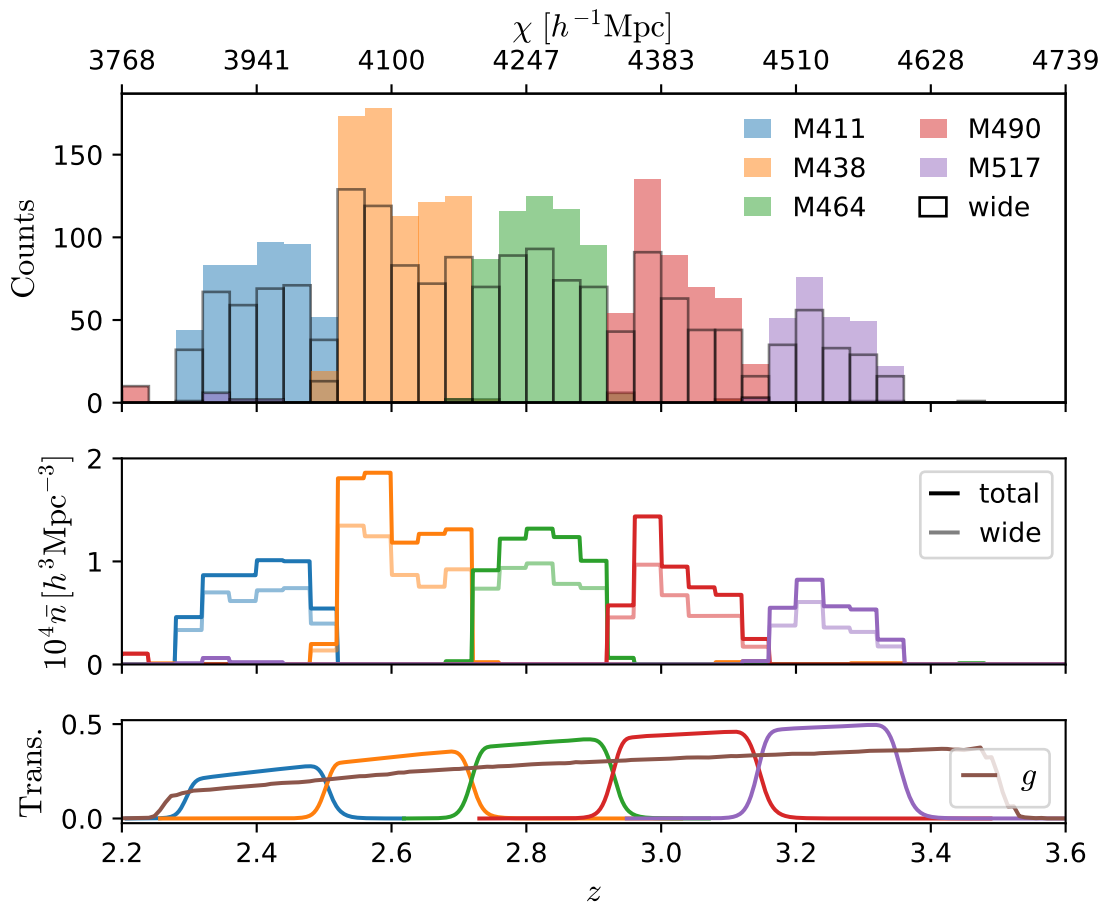


Figure 1: The spectroscopic redshift distribution of the high redshift samples in this work (both the total and wide samples) colored by the medium band used to select it (top), the inferred 3D number density (middle), and the corresponding filter curves (bottom), plotted against redshift assuming $\lambda = (1 + z)1216\text{\AA}$. The g band is added on the bottom panel for comparison.

The target selection in this work is a combination of a number of different, exploratory selections. While we defer details about the selections to companion papers [31, 59], we will describe them briefly here. We used three of the selections explored to date, each of which can be performed over multiple medium bands. The first category relies on the colors between three adjacent medium bands to select a slight peak in the center band, e.g. to select objects with an emission line in the M464 filter

$$M438 - M464 > 0.6, \quad M464 - M490 < 0, \quad M438 - M464 > 1 + 0.6(M464 - M490) \quad (3.1)$$

where M438, M464, and M490 are the magnitudes in their respective filters (Fig. 1). An additional ‘quality’ cut ($\text{FRACFLUX} < 0.15$) is applied for each medium band to reject objects with significant flux contributions from adjacent sources. The second category performs a traditional LAE color cut by using the five medium bands to construct a synthetic g band and placing a cut on the excess color between a medium band and the synthetic broad band,

Target sample	N_{targ}	N_{obs}	N_z	$N_{z\text{good}}$	f_{int}	z range	$10^2 \mathcal{L}^{-1}$
M411	1065	559	508	457	0.10-0.18	(2.26, 2.56)	0.55
M438	1938	991	824	736	0.11-0.26	(2.47, 2.77)	0.61
M464	1226	631	588	552	0.06-0.13	(2.68, 2.98)	0.69
M490	891	520	471	436	0.07-0.16	(2.89, 3.20)	0.78
M517	458	284	274	253	0.08-0.11	(3.10, 3.41)	0.86
M411 wide	707	395	385	338	0.12-0.14	(2.26, 2.56)	0.55
M438 wide	1213	672	588	509	0.13-0.24	(2.47, 2.77)	0.62
M464 wide	860	459	437	406	0.07-0.12	(2.68, 2.98)	0.67
M490 wide	606	364	338	303	0.10-0.17	(2.89, 3.20)	0.77
M517 wide	320	194	192	172	0.10-0.11	(3.10, 3.41)	0.87

Table 1: The observational properties of the high redshift galaxy sample considered for this work. The data are collected over an area of 6.11 deg^2 in the XMM-LSS field, with counts N being the total number of such galaxies (see text for definitions). Of the entire sample, 54% of the targets receive sufficient observation time ($\text{EFFTIME} > 2\text{HR}$) and 48% have secure spectroscopic redshifts ($\text{EFFTIME} > 2\text{HR} \ \& \ \Delta\chi^2 > 30$), providing information of the redshift distribution and interloper fraction for a ‘characteristic’ subset. f_{int} is the interloper fraction, estimated from DESI spectroscopic follow-up. The z -range is determined by the edges of the filter curves. The last column, \mathcal{L}^{-1} in $h \text{ Mpc}^{-1}$, indicates the impact of projection in diluting the angular clustering, as discussed in §4. A narrower redshift bin results a larger multiplier, implying a larger angular clustering given the same underlying distribution of galaxies.

e.g.

$$M464 - g_{\text{synth}} < -0.5 \quad , \quad M464 - g_{\text{synth}} < 0.5 - 2.8 \exp\{M464_{\text{fib}} - 26\} \quad (3.2)$$

where $M464_{\text{fib}}$ is the fiber magnitude in the M464 filter. The third category, similar to the second, uses the color between an individual medium band and a synthetic band, averaging the flux of the other four medium bands, e.g.

$$M464 - \text{synthetic} < -0.45 \quad , \quad M464 - \text{synthetic} < -0.5 - 0.7(M464 - 24.2) \quad (3.3)$$

We combine these three selections to construct the target sample used in this work.

By targeting the $\text{Ly}\alpha$ emission line at 1216\AA , each of these selections allow for an efficient selection of high- z galaxies in bins of $\Delta z \sim 0.2$. Together, they form a catalog spanning $2.26 < z < 3.41$. In this work we consider a subset⁴ of the initial target sample used for spectroscopic follow-up, in order to reduce and better constrain interloper fractions, which is important for the angular clustering that we will perform. For the latter three medium bands (that we analyze below as the $z = 3$ bin) we have applied tighter fiber magnitude cuts to each selection (the faintest at 25.3), as well as additional broadband color cuts using the g , r and i bands

$$g - r < 0 \quad | \quad (r - i < 0.3 \quad \& \quad g - r < 0.5) \quad . \quad (3.4)$$

At the expense of reducing the sample size by a factor ~ 2 , this effectively distinguishes true high- z galaxies from interlopers in the stellar locus, reducing the interloper fraction from

⁴There is a non-negligible fraction (22%) of our targets that were not part of the original target sample. This is due to a coding error in one of the initial target selections, which has been corrected by substitution with a similar, but nonidentical selection.

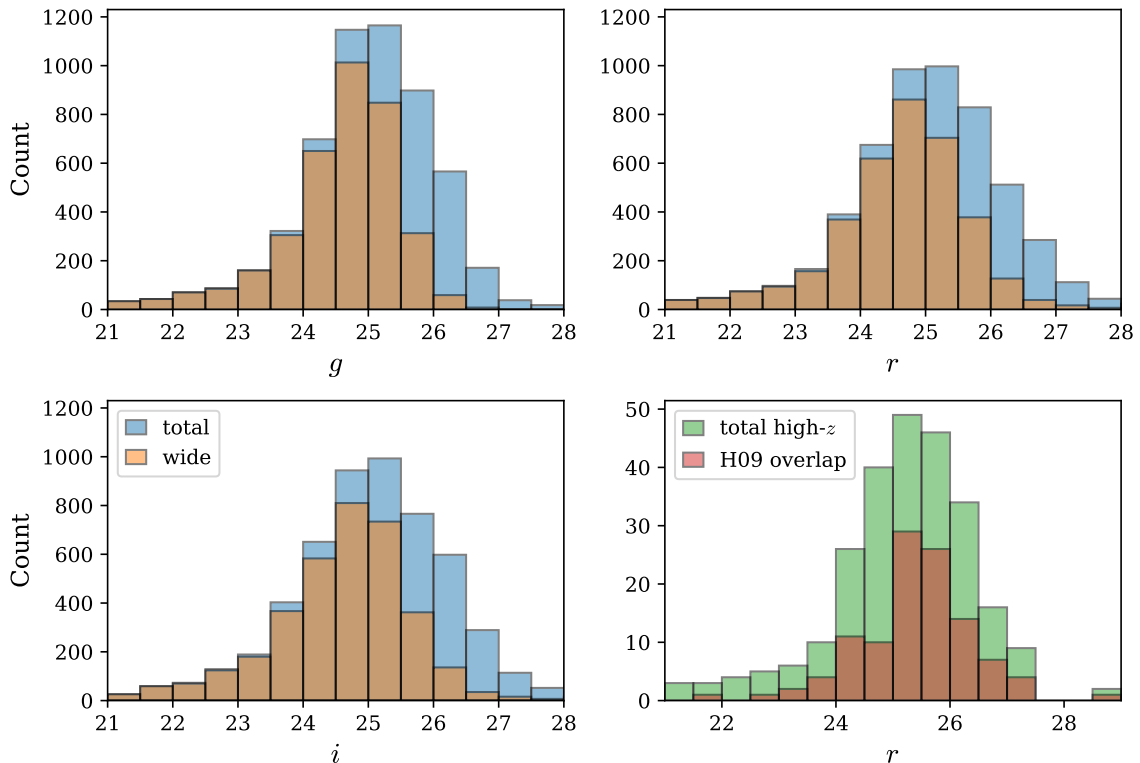


Figure 2: The broadband magnitude distributions of the medium-band selected samples. We see the g , r and i -band magnitudes are generally fainter than the MB flux cuts of 25 and 25.3 magnitudes, entirely as expected for samples with emission lines. In the lower-right panel we compare the r -band magnitude distribution of the galaxies that we check for LBG overlap in §3.4 (M490 and M517 galaxies that overlap the footprint of CARS [27]; green) to those that overlap with the LBGs selected by ref. [27] (red).

$f_{\text{int}} \lesssim 0.4$ to $\lesssim 0.15$ and reducing the uncertainty in the interloper fraction by a factor of ~ 3 . For the first two medium bands (that we will analyze as the $z = 2.5$ bin), the interloper fractions are significantly lower ($f_{\text{int}} \lesssim 0.25$) than the higher redshift medium bands, and hence we do not perform refinement in order to preserve the statistical power of the sample⁵.

The properties of our samples are given in Table 1. For each filter we define two subsets of the full sample, with the ‘total’ sample reaching a medium-band fiber magnitude of 25.5 ($z = 2.5$ bin) or 25.3 ($z = 3$ bin), and the ‘wide’ sample to 25. The ‘wide’ sample will provide a direct reference for the “wide-field” component of the IBIS survey, which will observe to medium-band depths of 25. Note that the intention here is not to simulate the target selection for the wide sample, as that will require degrading of the medium-band signals, but to provide reference to the luminosity limit of the brighter wide field. In Figure 2 we show the distribution of broadband magnitudes for the galaxy samples. Notice that they are considerably fainter than suggested by the medium-band magnitude limits, extending beyond $r \sim 26$. This is as expected for samples that contain significant $\text{Ly}\alpha$ line emission.

⁵When performing a high-purity refinement for the first two medium bands, we found that the detection of clustering was $\lesssim 1\sigma$ due to the large increase in shot noise.

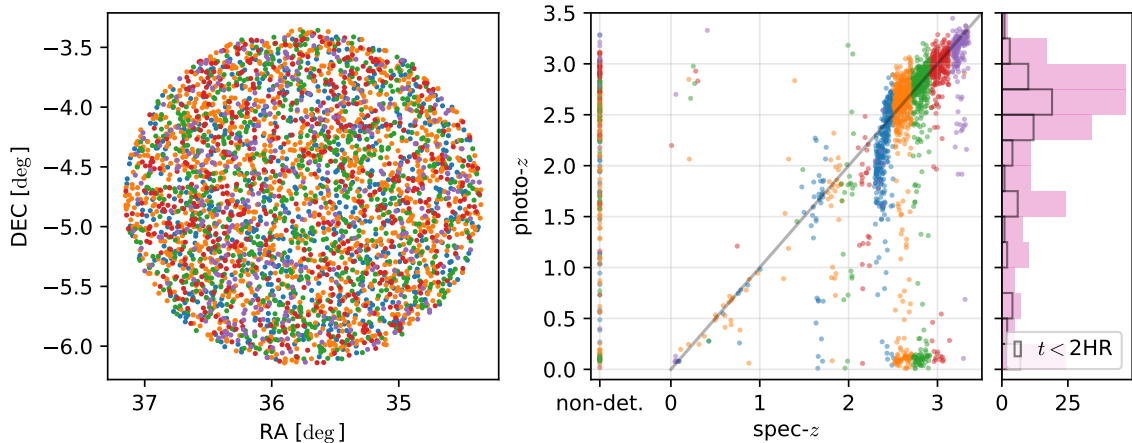


Figure 3: Left: The distribution of selected targets on the sky, with colors corresponding to the filters in Figure 1. Center: The performance comparison between CLAUDS SExtractor LePhare photometric redshifts, which were used to inform the target selection, and the DESI spectroscopic redshifts ($\Delta\chi^2 > 30$ in template fitting software RedRock [60]). Targets with $\text{EFFTIME} > 1\text{HR}$, but no spec- z , are shown on the left end of the figure. The targets are again colored according to their selection bands. Right: The photo- z distribution of objects with observation time $\text{EFFTIME} > 1\text{HR}$, but no reliable spec- z . We isolate the objects that do not reach our fiducial EFFTIME in black outlines and find that the duration of observation does not have a qualitative impact on the photo- z distribution of failures.

3.3 DESI spectroscopy

DESI is a Stage-IV cosmology survey designed to study dark energy [1, 61]. It employs a highly multiplexed spectroscopic instrument stationed on the 4-m Nicholas U. Mayall Telescope at the Kitt Peak National Observatory in Arizona, USA. The main survey started in 2021 and has collected over 40 million extragalactic redshifts to date. Using 5000 robotic fibers covering a 3.2 deg diameter field of view, it is able to collect 10^{4-5} spectra per night [62–67]. Aside from the main survey, DESI allocates time for ancillary programs to pursue a wide range of science goals with (relatively) cheaper fiber cost. We use the ancillary program `tertiary44` (TILEIDs=83519–83548) to follow-up the targets selected using IBIS medium bands. Here, we describe the details of this observation.

Observations are spread over a series of tiles, each assigning targets to the 5000 fibers (of which just over 4000 are for science targets). Targets are assigned to fibers using information about the targets’ positions, the current state of the focal plane and a set of priorities and required number of observations. The observations are ‘dithered’ so that each target obtains data from a number of different fibers to obtain the desired exposure time.

In order to maximize the observation efficiency in the ancillary program, the fiber assignment [68] is done over two iterations – a pre-assignment and final assignment. The pre-assignment runs a typical fiber assignment algorithm over the proposed targets and determines the targets that are likely to receive the required number of assigned fibers (for this sample, $\text{NASSIGN} = 8$, which translates to $\text{EFFTIME} \simeq 2\text{HR}$ where EFFTIME is a measure of exposure time accounting for observational conditions – see §4.14 of [69]). The targets that are deemed unlikely to reach the required exposure are demoted from the target class in order to assign their allocated fibers to other ancillary targets. While this maximizes the fiber

usage efficiency, it poses significant difficulty in terms of the clustering analysis, as discussed further in Appendices A and B.

The spectroscopic observations, reduced through DESI’s extensive pipeline [69] version 10a, provide template-based redshift measurements of the targets using a modified version of RedRock [60] including four high-redshift templates (courtesy of J. Moustakas as described in [31]). While the pipeline’s performance has been validated on earlier datasets [31], we did not carry out a dedicated visual-inspection (VI) campaign for the present sample, so further refinements remain possible. The redshift information allows us to determine the three-dimensional positions and quantify key information such as redshift distribution and interloper fractions, as shown in Figure 1 and Table 1. Here, we define the interlopers to be galaxies that do not lie in the redshift range of the respective medium-band filter (i.e. an M411 galaxy with $z = 2.6$ is an interloper). Note that even without the spectroscopic data, we are able to locate most galaxy redshifts to $\Delta z \simeq 0.2$ due to the medium-band selections targeting the Ly α emission line.

Figure 3 shows the on-sky positions of the objects making up our total sample, colored by the medium-band filter in which Ly α would lie. We also compare the photo- z from CLAUDS SExtractor LePhare [70], which were originally used in planning the target selection, and the DESI spectroscopic redshift where we have a secure⁶ redshift. Overall the photo- z and spec- z track each other quite well. The spectroscopic redshifts are concentrated in the ranges defined by the filter bands, as expected: we are more likely to get a secure spec- z when we have Ly α emission in the band. The distribution of photo- z s for the redshift failures suggests that many of them also lie at $z > 2$, but this is partly due to the fact that we used the photo- z s in designing our targeting cuts. The number of objects with $z_{\text{phot}} > 2$ with secure spectroscopic redshifts $z < 2$ is quite small (1%), illustrating the effectiveness of the medium-bands for selecting high-redshift targets.

Later we will measure the angular clustering of our targets, and use the dN/dz and interloper fractions (f_{int}) determined from spectroscopy to infer the underlying 3D clustering. Of our entire target sample, 48% of objects have secure ($\text{EFFTIME}>2\text{HR} \ \& \ \Delta\chi^2 > 30$) spectroscopic redshifts and hence will be able to provide reliable characterization of its redshift distribution. As we will see, restricting the sample to only objects for which we have a well characterized f_{int} means that the remaining uncertainties on dN/dz and f_{int} are much smaller than our statistical uncertainties. We shall henceforth neglect them.

3.4 Overlap with traditional tracers

As our selections are dissimilar to many previous high redshift target selections, we would like to understand the composition of the galaxy sample in this work viewed with respect to traditional descriptions of LAEs and LBGs. Note at the outset that our investigation must be preliminary as the data are prone to significant selection effects and incompleteness with pipelines that are still under development.

All three color cuts described in §3.2 are designed to pick out Ly α emission, so the resulting galaxies should, on average, exhibit large rest-frame equivalent widths (REWs) and qualify as LAEs under the definition of various literature [18, 37, 38, 71]. Even so, the catalog is not expected to consist of only LAEs, as LAEs and LBGs populate an overlapping spectrum of Ly α emission strength [12, 14]. To place this discussion into context in the previous literature, we cross-match with the u -dropout LBG sample from ref. [27] (called the

⁶Note that the DESI pipeline has not been optimized for redshifting galaxies such as these in this z -range so we anticipate that further work could increase the number of secure redshifts non-trivially.

H09 catalog from here on) to find the overlapping objects. For an accurate assessment of the overlapping galaxies, we restrict our cross-matching catalog to the M490 and M517 samples ($2.89 < z < 3.41$), which match the narrower redshift distribution ($2.9 \lesssim z \lesssim 3.6$) of the H09 sample⁷.

The cross-matching reveals that 46.1% of our M490 and M517 galaxies are LBGs, implying that a significant fraction of LBGs have Ly α emissions strong enough to be picked up by the medium bands used in the target selections. This, taking into account the sampling bias from our target selection, aligns with past study of LBGs, that LBG REWs are distributed approximately evenly between positive and negative Ly α emissions, with 25% having $\text{REW} > 20\text{\AA}$ [14, 72]. The broadband magnitude distribution of the overlapping galaxies, shown in the lower-right panel of Figure 2, reveals that the LBGs present in our catalog follows the distribution of the overall sample and peaks at a relatively faint continuum magnitude of $r \sim 25.5$. This can be understood as a consequence of the target selections detecting Ly α peaks through a color excess between Ly α band and adjacent bands. While this result sheds light on the mixture of LAEs and LBGs in the present work, extrapolating this result to the total galaxy sample requires care as incorporating the redshift dependence of each sample would be non-trivial.

4 Clustering analysis

4.1 Angular clustering

Although we obtain spectra for a significant fraction of our galaxy samples, we will not be using 3D clustering due to fiber assignment complications, as described in Appendix B. Instead, we will use the target sample to perform a pseudo-3D clustering through layers of angular clustering. This approach also has the advantage that it can be directly applied to the “wide” layer of IBIS without requiring extensive spectroscopic follow-up.

As our target samples naturally come subdivided into slices of width $\Delta z \sim 0.2$ ($\Delta\chi \sim 200 h^{-1}\text{Mpc}$), we can perform pseudo-3D clustering by collecting the clustering information in each redshift bin. Specifically we compute the angular auto-correlation function, $w(\theta)$, of the targets within each filter and then combine the resulting measurements into broader redshift bins. We choose to combine the 5 filters into two z bins at $z \sim 2.5$ (M411 and M438) and $z \sim 3$ (M464, M490, and M517). Each $w(\theta)$ corresponds to taking the (distribution-weighted⁸) radial integral over the 3D correlation function ξ

$$w(\theta) = \int d\chi_1 d\chi_2 f(\chi_1) f(\chi_2) \xi \left(\sqrt{\chi_1^2 + \chi_2^2 - 2\chi_1\chi_2 \cos \theta} \right) \quad (4.1)$$

$$\approx \left[\int d\chi f^2(\chi) \right] \int d\Delta \xi \left(\sqrt{\chi_0^2 \tilde{\omega}^2 + \Delta^2} \right) = \mathcal{L}^{-1} w_p(R = \chi_0 \tilde{\omega}) \quad (4.2)$$

where $\tilde{\omega} = 2 \sin(\theta/2) \simeq \theta$, $\Delta = \chi_2 - \chi_1$ and $f(\chi)$ is the radial (redshift) distribution normalized such that $\int f(\chi) d\chi = 1$. We have defined $\mathcal{L}^{-1} = \int d\chi f^2(\chi)$ and note that for a tophat $f(\chi)$ of full-width $\Delta\chi$ we have $\mathcal{L} = \Delta\chi$. The \mathcal{L} values for each redshift slice are listed in Table 1. In the second line above we have used that our filters select narrow f peaked around comoving distance $\chi_0 \gg \Delta\chi$ and converted from angle to transverse separation using the

⁷The mismatch in the redshift range $3.41 < z \lesssim 3.6$ does not concern us, as we are only interested in the overlap of our sample with the H09 catalog.

⁸The impact of lensing is expected to be highly subdominant to our observational errors.

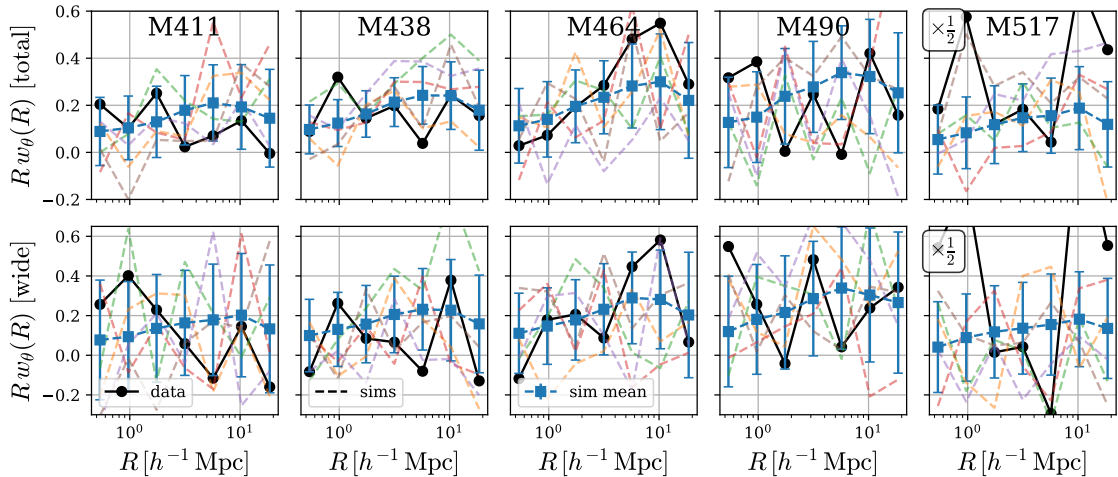


Figure 4: The angular clustering measurements $w_\theta(R) = w(\theta = R/\chi_0)$ of each medium band, with the total and wide sample shown on the top and bottom panels, respectively. The data measurements, in black circles, are shown together with the predictions from the best-fit HOD model (blue squares) and 5 random realizations of the best-fit HOD (faint dashed lines). Note that these measurements and errorbars are not used to directly infer goodness-of-fit to the mock and determine the best-fit HOD. That is done using $w_p(R)$, shown in Figure 5. The errors are calculated using 256 pseudo-independent realizations of the HOD, by offsetting the observer and observing the simulation through the survey mask and redshift selection function. For display purposes the black points have been scaled by 1/2 in the final column, since they exhibit large scatter.

comoving distance to each slice center (see Appendix C or refs [73, 74]). We recognize the integral of ξ over $d\Delta$ as $w_p(R)$ evaluated at $R = \chi_0\theta$ and with the limits of the line-of-sight integration defined by the width of the redshift slice (i.e. approximately $\pm 100 h^{-1}\text{Mpc}$) rather than a fixed maximum separation. Note that this projection justifies the use of the real-space correlation function in Eq. 4.1. We discuss this approximation further, and compare it to the full calculation, in Appendix C.

We compute $w(\theta)$ using the Landy-Szalay estimator [75] in each bin

$$w(\theta) = \frac{DD - 2DR - RR}{RR} \quad (4.3)$$

where DD , DR and RR represent (normalized) counts of pairs of data and randoms, with the random points that define the survey footprint and masks generated using the `desitarget` pipeline within DESI [76].

In order to combine information from five medium-band redshift slices into two z -bins, we will bin the correlation function in transverse distances rather than angles ($w_\theta(R) = w(\theta = R/\chi_0)$) to estimate the projected correlation function $w_p^{(i)}$ in each bin (Eqn. 4.2) and produce their inverse variance-weighted sum

$$w_p(R) = C \sum_i C_i^{-1} \left[\mathcal{L}^{(i)} w_\theta^{(i)}(R) \right] \quad (4.4)$$

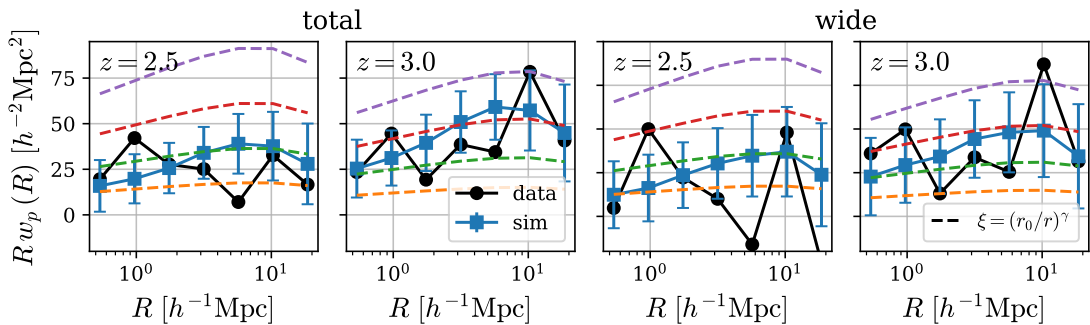


Figure 5: The pseudo-projected clustering, $w_p(R)$, for the $z = 2.5$ and 3.0 z -bins for each clustering sample. The data (black circles) are compared against the best-fit HOD model (blue, Table 2) yielding fits within $\approx 1\sigma$. We also overlay the predictions from a power-law correlation function $\xi = (r_0/r)^\gamma$, where the orange, green, red, and purple lines correspond to $r_0 = 2, 3, 4,$ and $5 h^{-1} \text{Mpc}$. The errors are calculated using 256 pseudo-independent realizations, as in Figure 4.

where the summation index i runs over the combined redshift slices, $\mathcal{L}^{(i)} w_\theta^{(i)}(R)$ and C_i are the projected correlation function and covariance matrix for slice i , and $C^{-1} = \sum_i C_i^{-1}$ is the inverse covariance of w_p . The \mathcal{L} applied in the calculation of $w_p^{(i)}(R)$ are listed in Table 1.

Although this is not a pure 3D clustering measurement, it includes information about every galaxy pair within each redshift slice and to the extent that these slices are much larger than the clustering distances that we are concerned with ($\Delta\chi \gg r \sim \mathcal{O}(10) h^{-1} \text{Mpc}$) it retains most of the information present in a real-space 3D correlation function. By integrating in the line-of-sight direction we avoid the need to model redshift-space distortions and by limiting the line-of-sight integration to the width of the medium band we avoid excessive signal loss due to projection. Importantly, by using the target sample directly we avoid needing to accurately model the impacts of fiber assignment. The angular clustering for each redshift bin $w_\theta(R)$ are shown in Figure 4, along with clustering from the best-fit mock catalog discussed in §4.2. The combined measurements w_p , for both (large) redshift bins are shown in Figure 5, again along with the mock catalog fit. For both figures, we have corrected the ‘raw’ measurements up by a factor of $(1 - f_{\text{int}})^{-2}$ to account for interlopers diluting the signal, under the (reasonable) assumption that the interlopers are spatially uncorrelated with the signal and of lesser clustering amplitude⁹. We use the middle of the f_{int} range quoted in Table 1 for each sample to minimize (the already negligible) uncertainty.

As is apparent from Figure 5, we have detected clustering in both $z = 3$ samples and the total sample at $z = 2.5$, but the detection is marginal for the wide sample at $z = 2.5$. However even this marginal detection is of interest because it provides an upper limit to the clustering amplitude and hence provides information about the host halos of these galaxies that is valuable for planning purposes. To get a feel for the implied clustering, Figure 5 shows w_p for four different correlation lengths ($r_0 = 2, 3, 4$ and $5 h^{-1} \text{Mpc}$) assuming the real-space correlation function $\xi = (r_0/r)^\gamma$ with fixed slope $\gamma = 1.8$. The array of curves fit the data successfully relative to our generous errors, indicating that we already have a basic

⁹Due to this correction, to linear theory order, an error in f_{int} translates into an error in linear bias b .

understanding of the galaxies at hand¹⁰ and suggest $r_0 \approx 2 - 4 h^{-1} \text{Mpc}$. We will refine this initial estimate using N-body simulations next, but we note that this is in line with previous LAE studies despite differences in sample definition [26, 37, 77–79]. The comparison with previous work on LBGs is less trivial as the correlation length is a clear function of continuum magnitude m_{UV} . With this in mind, our rough inference agrees with previous LBG samples of faint ($m_{UV} < 25, 25.5$) continuum magnitudes [27, 80, 81].

4.2 HOD fit

In this section, we will use mock catalogs from Halo Occupation Distribution (HOD) models to compute covariance matrices and interpret our clustering measurements. A HOD model assigns (mock) galaxies to dark matter halos through a probability distribution function that depends (primarily) upon halo mass. The form of the halo occupation for galaxies such as those in our sample is not well understood, nor are there strong observational constraints on the dependence upon secondary properties of halos, beyond mass. On the other hand, our data are currently unable to provide precise constraints over a wide range of scales and thus many of our inferences will be robust to model (Appendix D) and cosmology misspecification.

In this work, we will primarily use the functional form introduced by ref. [82] as implemented in the AbacusUtils¹¹ software accompanying the ABACUSSUMMIT N-body suite [83], which have been generated using the ABACUS N-body software [84]. Comparison to a different HOD form can be found in Appendix D. For each halo, the ‘standard’ model assigns central galaxies with a binomial distribution and a Poisson-distributed number of satellites, with means

$$\langle N_{\text{cen}}(M) \rangle = \frac{1}{2} \text{erfc} \left(\frac{\ln M_{\text{cut}}/M}{\sqrt{2}\sigma} \right) \quad (4.5)$$

$$\langle N_{\text{sat}}(M) \rangle = \langle N_{\text{cen}}(M) \rangle \left(\frac{M - \kappa M_{\text{cut}}}{M_1} \right)^\alpha \quad \text{for } M > \kappa M_{\text{cut}} \quad (4.6)$$

where M is the halo mass and $\{M_{\text{cut}}, M_1, \sigma, \kappa, \alpha\}$ are model parameters. Although this functional form, and HOD models more generally, were originally designed to model magnitude limited samples at $z \lesssim 0.5$, due to the level of flexibility of the model we expect moderate success even for high-redshift galaxies. Indeed, N-body studies and observations have each shown that the halo distribution and stellar mass/halo mass relations are roughly consistent across a wide range of redshifts, indicating that the functional form of the HOD model will be at least roughly applicable at $z \sim 3$ [85]. Works modeling the high-redshift universe with semi-analytic models or hydrodynamical simulations have also shown that these HOD models are moderately successful at describing high- z galaxies [36, 38, 39, 46, 86, 87]. While this clearly does not guarantee that these models will correctly mimic the galaxy samples at hand, at the precision level of this study we deem it sufficient. It will be valuable to follow this up in the future with more constraining data, as detailed simulation-based study of medium-band selected high- z galaxies will provide important insight into both future cosmology survey planning and the evolution of galaxies within large-scale structure.

Within the standard cosmological model, the amplitude of clustering we measure indicates that our galaxies are hosted by halos only slightly more massive than the characteristic

¹⁰Note that while we do not depend explicitly on the best-fit model here, we adopt the covariance matrices of the best-fit HOD models to provide the statistical comparison to data.

¹¹<https://abacusutils.readthedocs.io/en/latest/>

z	Sample	$\log M_{\text{cut}}$	$\log M_1$	σ	κ	α	$\chi^2 (n\sigma)$	f_{sat}	r_0
2.5	Total	11.25	11.95	0.66	1.00	0.33	7.36 (0.86 σ)	0.12	3.28 ± 0.70
	Wide	11.00	12.00	0.66	1.00	0.50	12.10 (1.66 σ)	0.03	3.08 ± 0.80
3.0	Total	12.00	12.70	0.66	1.00	0.50	7.22 (0.83 σ)	0.10	4.13 ± 0.24
	Wide	11.75	12.45	0.66	1.00	0.33	6.65 (0.73 σ)	0.05	3.85 ± 0.68

Table 2: The HOD parameters for the best-fit projected correlation functions against data in Figure 5, along with the goodness-of-fit χ^2 , satellite fraction f_{sat} , and correlation length r_0 of the best-fit model. f_{sat} and r_0 are parameters derived from simulation (see text). The uncertainties in r_0 are derived from simulations satisfying $\Delta\chi^2 < 2$. Due to the limited mass resolution, we quote the error corresponding to the upper end of the r_0 range (Figure 7). Masses are in $h^{-1}M_{\odot}$ and distances in $h^{-1}\text{Mpc}$.

mass in the halo mass function. Since at these redshifts the characteristic mass is quite low, we will construct mock catalogs using the high-resolution simulation of the ABACUSUMMIT suite. This simulation assumes a Planck ΛCDM cosmology [32, 88] and uses 6300^3 particles in a $1 h^{-1}\text{Gpc}$ (periodic) box. This implies a particle resolution of $m_{\text{part}} = 3.5 \times 10^8 h^{-1} M_{\odot}$ [83], which we will see later provides just sufficient resolution for the halos of interest. Given the wavelength coverage of our medium-bands, we will use the $z = 2.5$ output to analyze the combination of the first two filters and the $z = 3$ output for the combination of the latter three.

The measurement errors can be estimated by observing 256 pseudo-independent realizations of each model by randomly offsetting the simulation along each axes of the box and “observing” them by applying 2D masks and redshift distributions. The masks are constructed from the random catalog using fine (`nside=8192`) `healpy` pixels [89] – any pixel that has a random in it is considered observable. The redshift distributions are inferred from the data obtained through DESI spectroscopic follow-up (see Figure 1). As the observation volume is significantly smaller than the simulation volume (by a factor of ~ 80 and ~ 60 for the $z = 2.5$ and 3.0 bins, respectively), each random offset provides nearly independent realizations of the density field¹².

Given an estimate of the covariance matrices, we can test any given HOD model for consistency with the data. Since the data at present have limited constraining power we have not attempted a formal fit. Instead we scan over a range of HOD parameters¹³, testing each to find the best-fit model (in our set) and the range of models that provide acceptable fits. This allows us to constrain both clustering and galaxy properties. The HOD parameters for the best-fit mock catalog from our set are shown in Table 2, along with associated goodness-of-fit values, satellite fractions f_{sat} , and correlation length r_0 . The corresponding clustering predictions are shown in Figure 4 against the data, with five random realizations of the mock catalog to give a rough intuition for individual fluctuations from sample variance. We observe that the highest redshift bin (M517) has the poorest fits in each sample. The reason for this is not clear, but we believe that it owes to the significantly smaller sample size of those bins compared to other bins, as shown in Table 1.

¹²While the number of mock observations is significantly larger than the ratio of simulation volume to survey volume, for the covariance estimation purposes it is unnecessary to have fully independent realizations.

¹³Specifically we fix $\kappa = 1$ and choose $\sigma = \{0.33, 0.5, 0.66\}$, $\alpha = \{0.33, 0.50, 0.66\}$ while sampling $\log_{10} M_{\text{cut}}/(h^{-1}M_{\odot})$ in steps of 0.25 between 11.0 and 12.5 and M_1/M_{cut} in $\{5, 10\}$ inspired by the HOD parameters explored in ref. [26].

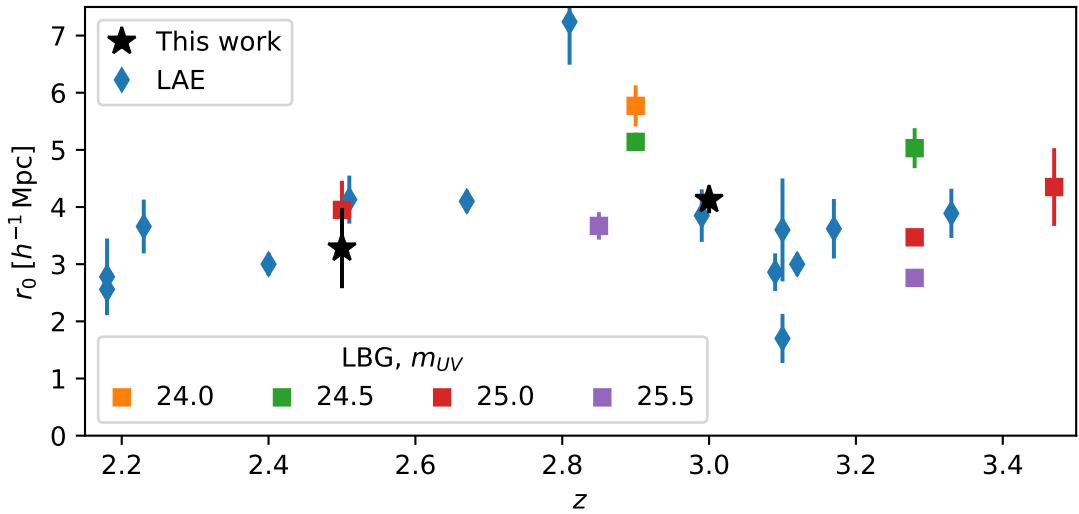


Figure 6: The best-fit r_0 inferred from our HOD-based mock fits to the data, compared to previous literature on LAEs [26, 37, 77–79, 90–93] and LBGs [27, 80, 81, 94]. We only plot the r_0 of the ‘total’ sample here, for visualization.

Figure 7 shows the χ^2 for each of the HOD models in our grid, plotted against the r_0 for that model (as determined from $\xi(r)$ in the N-body simulation). The absolute value of the χ^2 determines whether the model is statistically consistent with the data, with $\chi^2 \simeq 14$ being 2σ exclusion for 7 degrees of freedom (d.o.f.). Further, we see (most clearly in the total sample at $z = 3$) that the data follow an approximately parabolic shape with a moderate amount of scatter. This illustrates the extent to which r_0 ‘predicts’ the χ^2 value and hence the degree to which there is additional information beyond this amplitude in the data for the model class described by our set of HODs. That the $z = 2.5$ points show no clear minimum in the χ^2 vs. r_0 is a combination of limited numerical resolution (preventing us from examining lower r_0) and low SNR for that redshift bin. The r_0 of the best-fit model inferred for each redshift (as shown in Table 2) is consistent with that from an array of past work [26, 27, 37, 77–81, 90–94], as shown in Figure 6. Note that each r_0 estimate varies in its definition of galaxy samples, including this work, contributing to the large scatter among data points. Finally, the best-fit model of the wide sample displays clustering that is slightly ($< 10\%$) stronger than that of the total sample, a result that is expected given its brighter luminosity. However, this difference falls well within the uncertainties of our analysis and is therefore inadequate to derive a definitive conclusion.

5 Inferences from simulations

In this section, we will use the HODs that are consistent with the abundance and clustering of our samples to investigate the implications for galaxy-halo connection, scale-dependent bias and modeling. We do not have the statistical precision to see significant differences in the clustering as a function of luminosity or redshift, though we do see differences in just the best-fitting models. At $z = 2.5$ we barely detect clustering for the wide sample, although the mock halo masses that match the level of clustering in our data are consistent with past

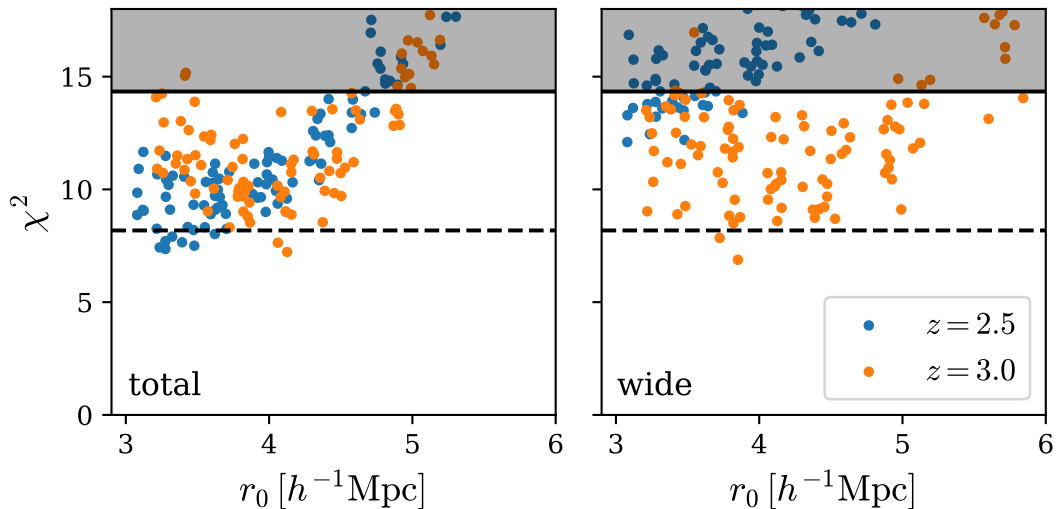


Figure 7: The χ^2 distribution of HOD models plotted against predicted correlation length, r_0 , with the total and wide samples shown in the left and right panels, respectively. The 1σ and 2σ CL for 7 d.o.f. are shown in dashed and solid lines, respectively, with the region beyond 2σ shaded. The distribution follows a parabolic curve with moderate levels of scatter. Lower scatter indicates that r_0 serves as a good summary statistic for the clustering, containing most of the information (for models following these HOD forms at the current level of precision). That the $z = 2.5$ points show no clear minimum in the χ^2 vs. r_0 is a combination of limited numerical resolution (preventing us from examining lower r_0) and low SNR for that redshift bin.

inferences of halo masses ($\sim 10^{10-11} M_\odot$ [12]) for LAEs. At $z = 3$ we have a stronger detection of clustering in both samples and our data prefer higher mass halos than the ‘typical’ LAE in the literature, which are more consistent with traditional LBG halo masses $\sim 10^{12} h^{-1} M_\odot$ [6, 13, 14]. We discuss the halo occupation further in Appendix D.

5.1 Mock Clustering

Our mock catalogs match the clustering and abundance of the samples selected in this work, and populate dark matter halos within a simulation based on the CDM paradigm. To the extent that they provide reasonable proxies for the real galaxies it is useful to study their clustering properties. Since we have access to much more information in the simulation than in observations, we calculate results in both real and redshift space and include cross-correlation with the dark matter field, to provide further insights into these high- z galaxies which is of particular interest for their usefulness as cosmological tracers.

For a cosmological tracer, the large-scale bias and its scale-dependence are key features important to cosmology. The results are shown in Figure 8, for both the auto-spectrum-derived bias $b_a = \sqrt{P_{gg}/P_{mm}}$ and cross-spectrum-derived bias $b_\times = P_{gm}/P_{mm}$, where P_{gg} , P_{gm} , and P_{mm} are the real-space auto-, cross-, and nonlinear matter spectra and the ratio of these biases b_\times/b_a is the correlation coefficient between the galaxy and matter fields. We subtract Poisson shot noise from the auto-spectra to show the maximum galaxy-matter correlation – thus we would expect any observed sample to be less correlated with the matter

z	Sample	b_1	b_2	b_s	α_0	α_2	N_0	N_2
2.5	Total	1.84	1.02	0.75	-6.08	2.11	5.6×10^1	-1.2×10^2
	Wide	1.76	0.69	0.50	-3.24	-0.64	3.2×10^1	2.5×10^1
3.0	Total	2.56	-0.77	-2.12	-6.08	5.99	1.5×10^2	-3.4×10^2
	Wide	2.42	-0.40	-2.78	-5.19	2.46	8.9×10^1	-1.1×10^2

Table 3: The one-loop, Eulerian PT parameters for each clustering sample at $z = 2.5$ and 3.0 , in order of biases b_i , counterterms α_{2n} , and stochastic terms N_{2n} . The parameters are inferred by a minimizer to the measurement of galaxy auto-spectrum in redshift-space, using ZCV [95–97], and galaxy-matter cross-spectrum in real-space. By using ZCV, the auto-spectrum can be measured precisely up to large scales. To improve the performance of the control variates technique, we do not randomly downsample the galaxies (to match \bar{n}) prior to measuring the clustering. Instead, we add Poisson shot noise to the original fit, using the number density from the target sample (Table 1). The counterterms α_{2n} are in units $h^{-5}\text{Mpc}^5$ and stochastic terms N_{2n} are in units $h^{-3-2n}\text{Mpc}^{3+2n}$.

field. We observe that both biases exhibit very modest scale dependence, with a $\sim 5\%$ evolution from large to small scales in the $z = 2.5$ bin and $\sim 15\%$ evolution in the $z = 3$ bin. The fact that $b_a \approx b_\times$ for all samples indicates that galaxy and matter fields remain well correlated (in the absence of Poisson shot noise and RSD). Given sufficient number density, these objects thus make excellent samples for probing large-scale structure and for cross-correlation studies. Further notice how the analytical predictions discussed below (Table 3), shown in solid lines, agree well with the N-body predictions indicating that the samples can be accurately modeled.

In redshift space, we will model the mock clustering using the one-loop perturbative power spectrum, specifically with the publicly available software `velocileptors` [98]. This modeling is identical to that adopted for the key results from two-point clustering in DESI DR1 [99]. The recent simulation-based study on high-redshift galaxies [87] also uses a variant of the theory, and it has been shown that the analysis results are consistent between variants of these perturbative models [100]. In addition to the inclusion of non-linear dynamics and redshift space distortions the model has parameters for the non-linear bias relation, $\{b_i\}$, stochastic terms associated with shot noise and fingers of god, $\{N_{2n}\}$, and a set of “counter terms”, $\{\alpha_{2n}\}$, that encapsulate small-scale physics not included in the model and that are dictated entirely by symmetry.

The biases, $\{b_i\}$, describe the galaxy overdensity field perturbatively

$$\delta_g \approx b_1 \delta + \frac{b_2}{2} \delta^2 + b_s s^2 \quad (5.1)$$

where we omit the third-degree bias b_3 as discussed in detail in refs. [100, 101]. The counterterms and stochastic terms each contribute to the power spectrum linearly

$$P(k, \mu) \supset \alpha_{2n,a} \mu^{2n} (k/k_\star)^2 P_L, \quad P(k, \mu) \supset N_{2n} (k\mu)^{2n} \quad (5.2)$$

with $k_\star = 1 h \text{Mpc}^{-1}$. We refer the reader to refs. [98, 102] for detailed discussion.

One set of nuisance parameters is able to consistently model any clustering statistic of the galaxy, both in real- and redshift-space. We will take advantage of this to simultaneously model the multipoles ($\ell = 0, 2$) of the redshift-space galaxy auto-spectrum, $P_{\ell,gg}$,

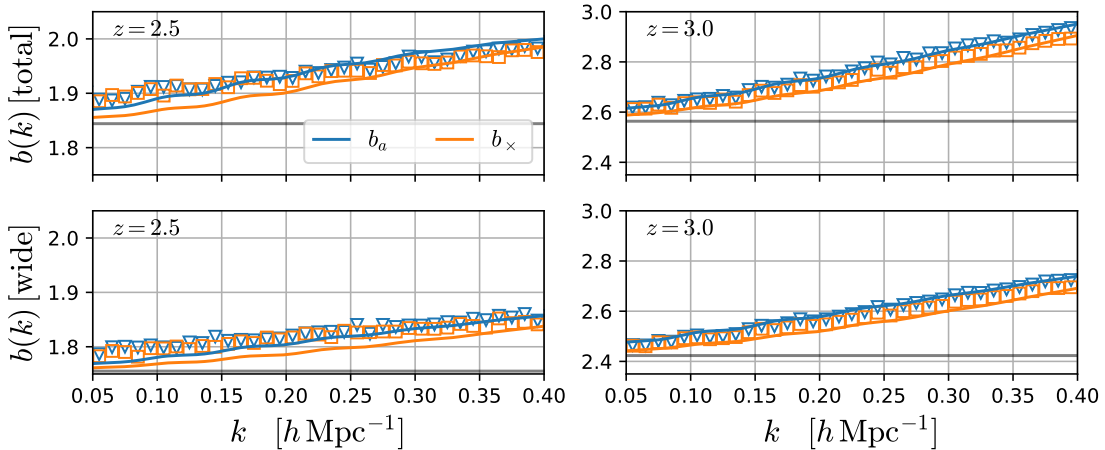


Figure 8: The real-space, scale-dependent bias in Fourier space for the mock clustering of two redshift bins (left: $z = 2.5$, right: $z = 3$) and two samples (top: total, bottom: wide). In blue we show the auto-spectrum bias $b_a(k) = \sqrt{P_{gg}/P_{mm}}$ and in orange the cross-spectrum bias $b_x(k) = P_{gm}/P_{mm}$, where we use the nonlinear matter spectrum for P_{mm} and subtract Poisson shot noise from P_{gg} to show maximum galaxy field-matter field correlation. The solid lines are the PT predictions for the joint fit to the auto- and cross-spectrum in real and redshift space, where we see good agreement consistent within the limited statistics of the simulation. At large scales, all biases asymptote to the linear bias b_1 (grey solid lines), as expected.

the real-space galaxy auto-spectrum, P_{gg} , galaxy-matter cross-spectrum, P_{gm} , and matter auto-spectrum, P_{mm} . In addition, although we are constrained to a $1 h^{-1}\text{Gpc}$ box, we are able to provide precise¹⁴ measurements of the mock power spectrum P_{mock} up to large scales using Zeldovich control variates [95–97] and working at fixed cosmology. The bias fits at different redshifts, found through the minimizer implemented in the *Cobaya* code [103, 104], are shown in Figure 9, with associated bias parameters shown in Table 3. We find that the linear biases b_1 at $z = 3$ fall in the approximate range of b_1 's derived in the LAE simulations ($1.89 < b_1 < 2.68$) of ref. [87], while falling significantly short of the b_1 's in LBG simulations ($3.92 < b_1 < 4.24$). Note, however, that the disparity with the LBG simulations are not surprising, as the LBGs in this work have significantly fainter continuum than that modeled in ref. [87] and LBG bias correlates strongly with continuum magnitude [6].

The nonlinearity of the high- z galaxy power spectra is a key quantity that we seek to characterize in these initial clustering analyses. The high- z universe is known to host highly biased tracers, such as broadband-selected high- z LBGs [6, 29], and high bias usually implies strongly non-linear bias [105]. To what extent the medium-band galaxies in this study exhibit such nonlinearity will be important for future modeling efforts.

The nonlinearity of the power spectrum can be captured as a combination of two effects. One is the nonlinearity of the underlying density field. The decorrelation of the nonlinear matter field with the initial (Gaussian) density field reduces the information about the latter

¹⁴An alternative means of overcoming sample variance is to do a field-level fit of the model to the mock density field [87]. Our approach achieves similar performance, and has the agreeable side effect of generating a low-noise measure of P_{mock} .

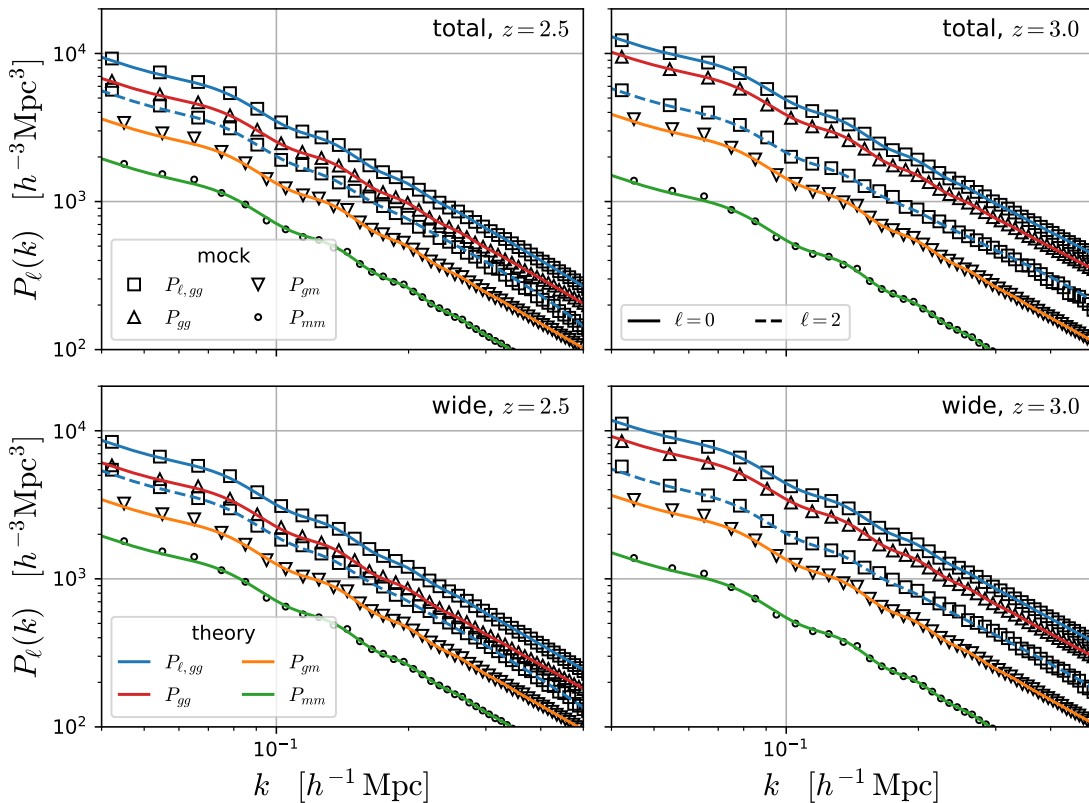


Figure 9: One-loop power spectrum fits to the mock catalog clustering, where the theory predictions are shown in solid lines and the mock measurements in markers. For each sample and redshift bin we simultaneously fit the redshift-space galaxy auto-spectrum (theory: blue, mock: squares) and the real-space galaxy auto- (red, upright triangle), galaxy-matter cross- (orange, reverse triangle), and matter auto-spectrum (green, circle). The fitted parameters for each sample and z -bin are shown in Table 3.

that can be learned from the former. At higher redshifts the matter field remains linear to smaller scales (Appendix A). This can be observed from deviations between the linear matter spectrum P_L and nonlinear spectrum P_{mm} . The second effect is the nonlinearity due to scale-dependent bias, which indicates a decorrelation between the galaxy and matter density fields¹⁵. This effect can be isolated by comparing P_{gg} , bP_{gm} , and b^2P_{mm} , which would be identical in the absence of scale-dependent bias. We show both of these comparisons in Figure 10, with the Poisson shotnoise subtracted for auto-spectra to emphasize scale dependence of clustering. While at $z = 2.5$ both the non-linearity and scale-dependent bias are quite modest over the range of scales plotted, we see this is not true for the $z = 3$ sample. There the non-linearity in the matter field is smaller than at $z = 2.5$, indicating the lower amplitude of large-scale structure growth, but the scale-dependent bias is much more pronounced. For

¹⁵The visually apparent difference between the spectra also rules out separable halo bias in the (analytical) halo model, which would imply that $P_{gm}/\sqrt{P_{gg}P_{mm}}$ is unity beyond the virial radius of halos. This agrees with recent work from ref. [106]. For further connections between PT and the halo model, see Appendix C of ref. [101].

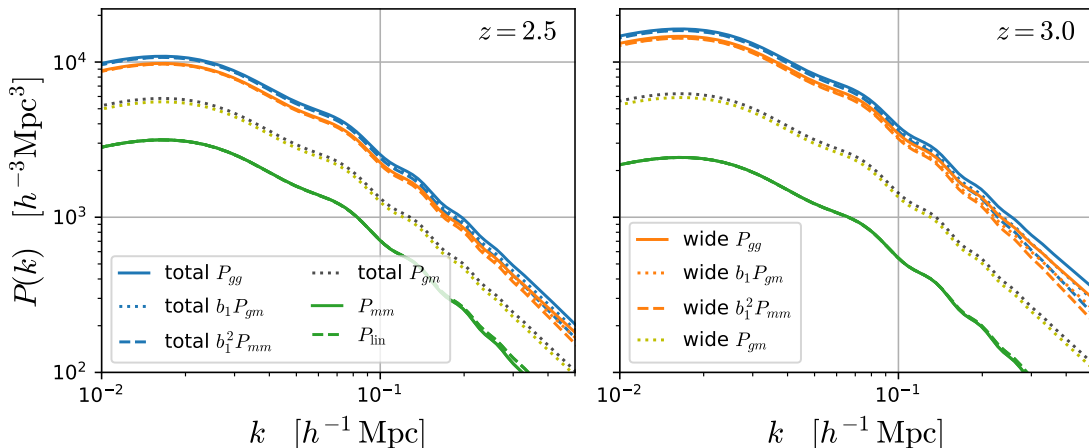


Figure 10: In blue and orange are comparisons between the real-space galaxy auto-spectrum, galaxy-matter cross-spectrum, and matter auto-spectrum (as computed using the best-fit PT model of Table 3; compared to N-body in Fig. 9). In green is the comparison between the linear and non-linear matter auto-spectra, illustrating the nonlinearity of the matter field. The blue and orange solid, dashed and dotted lines highlight the scale-dependence of the bias. The left and right panels correspond to the $z = 2.5$ and $z = 3$ results (Table 3), respectively (and the orange and blue lines almost overlap for $z = 2.5$).

this sample, scale-dependent bias drives the decorrelation between the matter field and linear theory, requiring increased complexity (and more parameters) in the modeling.

The FoG of these galaxies is another important modeling question that must be addressed. While there are no theoretical indications that the galaxy modeling of FoG will break down at high redshift, galaxies at high redshift tend to have higher bias and recent work using hydrodynamical simulations indicate that high- z galaxies may exhibit FoG that range from smaller than that of low- z galaxies [86, 87] to stronger effects requiring higher-order perturbative terms [87]. We observe that our PT models provide decent fits to the FoG of the best-fit models (Figure 9) which is sufficient for our purposes. The mock clustering also shows that the zero-point crossing of the power spectrum quadrupole, which can be used as a proxy of FoG strength [107], lies at $k > 0.5 h \text{ Mpc}^{-1}$, beyond our estimate of the non-linear scale, $k_{\text{nl}} \approx \Sigma_{\text{Zel}}^{-1}$ at $z \sim 3$ (Appendix A; Fig. 12). A small FoG term is also consistent with the measurements of line-of-sight correlations between LAEs and the Ly α forest measured in the spectra of LBGs [108].

5.2 Forecast

With the IBIS survey continuing to observe the sky, including deep coverage in the COSMOS field in 2025A (with spectroscopic follow-up with DESI) and wide fields in the future, we expect the observational constraining power to increase steadily in both 2D and 3D clustering. Hence, it is worthwhile considering what progress we may expect in the near future.

In the wide survey, measurement of angular clustering over a sky area of 3000 deg^2 will yield one of the most precise high-redshift clustering measurements to date. While the galaxy auto-correlation alone will show complete degeneracy between bias and clustering σ_8 (on linear scales), the introduction of galaxy-CMB lensing cross-correlation will effectively break

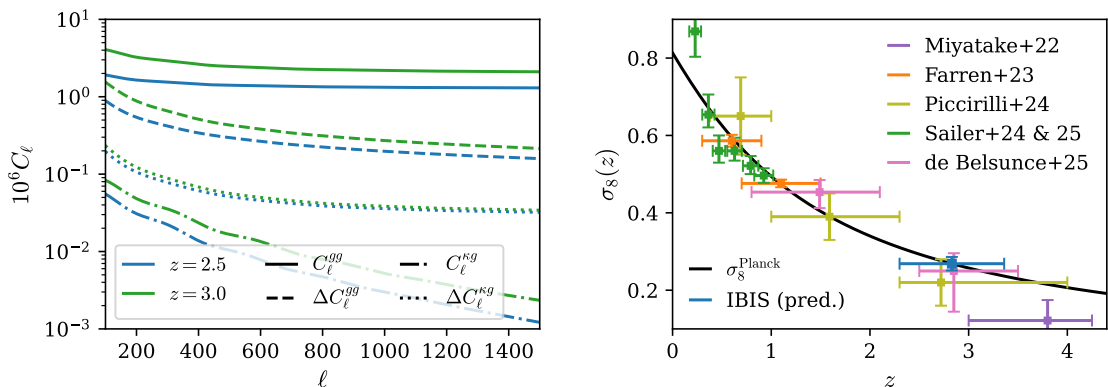


Figure 11: Left: Angular power spectra and their respective errors for the galaxy auto-spectrum and galaxy-lensing cross-spectrum for the third redshift bin (at $z \sim 2.83$; corresponding to M464), assuming the galaxy characteristics of the wide sample and forecasts for SO lensing noise curves. The errors are calculated assuming a Gaussian covariance and 3000 deg^2 sky area. Right: Fixed-shape σ_8 Fisher forecasts assuming the total sample over 3000 deg^2 using `FishLSS` [7], in comparison with recent structure growth measurements obtained from galaxy-CMB lensing cross-correlations [40, 111–115]. The 6.5% constraints on $\sigma_8(z)$ (blue) will be the tightest constraint on the clustering at such high redshift. This will improve to 5.1% if the number density is doubled due to improved target selection.

this degeneracy [6, 40, 42, 109]. The expected power spectra and errors (for one redshift bin) are shown in the left panel of Figure 11, where the errors are calculated assuming Gaussian covariance.

Using Fisher forecast software `FishLSS` [7], we are able to translate the observation expectations into projected constraints¹⁶. Assuming the statistics of the wide subsample (which has the same luminosity limit as the wide survey) and ‘goal’ sensitivity for Simons Observatory (SO) [110] lensing noise curves¹⁷, we find that IBIS will yield the tightest σ_8 constraint at high redshift after 3000 deg^2 , as shown in the right panel of Figure 11.¹⁸ We refer the reader to ref. [7] for a detailed description of fixed-shape σ_8 forecasts.

We want to emphasize that the clustering analysis in this work deliberately refines the sample to have well-characterized f_{int} and dN/dz , at the expense of increased statistical noise. Due to this refinement, we can reliably measure σ_8 from cross-correlation, even without extensive spectroscopic follow-up. However, if we can improve our selections, we can dramatically increase the constraining power of the sample. Undoing the refinement will quickly increase the sample size by a factor of ~ 2 and ongoing work in target selection indicates that further improvement of another factor of ~ 2 is likely. We intend to report these investigations in a future publication.

¹⁶For forecasting purposes, we use the z -range derived from the FWHM of the filter curves instead of that in Table 1 to avoid overlap between the z -bins, which is assumed when adding their respective Fisher matrices.

¹⁷The noise curves are generated by the `SO Noise Calculator`.

¹⁸If the footprint is altered from 3000 deg^2 , constraints in Fisher forecasts will reduce by a factor of $\sqrt{f_{\text{sky,new}}/f_{\text{sky,old}}}$. This indicates that a footprint of 300 deg^2 , which would degrade the constraints in Figure 11 by a factor of $\sqrt{10}$, is sufficient to deliver the tightest constraints on σ_8 at $z \sim 3$ at the Fisher forecast level.

6 Conclusion

Over the last several decades, the study of large-scale structure has become an integral part of establishing and testing our cosmological model, with the latest generation of surveys pushing to ever greater statistical precision. For example, the latest generation of BAO measurements has provided constraints on the distance scale with an aggregate precision well under a percent [5]. However the fraction of the available cosmological volume that these surveys have mapped is still small, and attention has been shifting to surveys of higher redshifts [8, 10]. Such surveys promise transformational gains in cosmological constraining power and redshift lever-arm. However in order to properly plan for them we need to better understand the population of galaxies that would be used as targets. In this paper we have begun such an investigation using observations from the IBIS and DESI surveys, with the goal of characterizing their clustering and halo occupations. These, in turn, have direct implications for the design, data analysis and simulation requirements for future surveys.

We have shown that the medium-band information from IBIS naturally allows us to define populations lying in well-delineated redshift shells of depth $\approx 200 h^{-1}\text{Mpc}$ spanning $z \simeq 2.3 - 3.4$. These galaxies are typically faint in broad band magnitudes (e.g. median $r \simeq 25$, Figure 2) but the presence of Ly α emission allows us to measure redshifts with 1 – 2 hr integrations on DESI [31]. Our color cuts select objects that span the traditional divide between LBGs and LAEs. We estimate that despite the color cuts selecting for detectable Ly α peaks, almost half of the high- z ($2.9 < z < 3.41$) are LBGs that would be selected by e.g. a u -dropout technique (§3.4).

For our analysis, we have refined our sample to have a well-characterized interloper fraction, at the cost of increased shot noise. While this allows us to provide preliminary estimates of the clustering based on the angular data, for a spectroscopic survey we will likely choose to increase the target sample size at the expense of some uncertainty, as a dedicated spectroscopic survey can well characterize these uncertainties, at the expense of losing some observation efficiency. Further work on target selection optimized for a future spectroscopic campaign is ongoing.

The DESI observations for this program were designed to maximize the efficiency for obtaining redshifts while leaving fibers for other ancillary programs, and thus do not enable a direct 3D analysis (Appendix B). However, the natural decomposition of the sample into nearly disjoint shells of $\approx 200 h^{-1}\text{Mpc}$ depth allows us to make a detection of clustering through angular correlations (§4.1, Appendix C). We find a marginal detection for one of the samples at $z = 2.5$, though we are able to set interesting upper limits on the amplitude, and a convincing detection of clustering of the other sample at $z = 2.5$, and both samples at $z = 3$. The angular clustering, along with dN/dz and f_{int} from spectroscopic follow-up, then allows us to infer the real-space clustering. An introductory, model-agnostic analysis of these samples indicate correlation lengths of $r_0 \sim 3 - 4 h^{-1}\text{Mpc}$ and linear bias $b \sim 1.8 - 2.5$, roughly in agreement with previous work on LAEs and LBGs.

Further insight can be obtained by modeling our samples using HODs (§4.2). While the appropriate form of the HOD is not well known, we adopt the standard, five-parameter HOD model in anticipation that it provides sufficient flexibility at the precision of this study. Since our target selection is complex and galaxies with ‘strong’ Ly α emission lines comprise only a fraction of the total galaxies at any epoch, we cannot simply use the abundance to determine the properties of the dark matter halos which host our targets. Instead we use the clustering, in our case the projected correlation function on “two halo scales”. This provides us with a

determination of the bias, which by assuming a particular functional form for the HOD, we can turn into an estimate of the characteristic halo mass. We find that the halo masses are roughly consistent with previous estimates for LAE halo masses, with the exception of the wide sample at $z = 3$ being closer to LBG halo mass estimates.

The simulations that we employ had barely enough resolution to model the galaxy populations in this work. This has direct implications for simulation requirements for future surveys. Assuming the standard form of the HOD (Eq. 4.5), to include the halos hosting the majority of the galaxies we need halos to be resolved down to $\approx 0.2 M_{\text{cut}}$, suggesting particle masses of two orders of magnitude lower than M_{cut} . Using the numbers for the $z = 3$ sample, for example, we would need to resolve $\geq 10^{10} h^{-1} M_{\odot}$ halos (Fig. 16) implying a particle mass less than $\sim 10^9 h^{-1} M_{\odot}$. A simulation box subtending 30° (for a survey area just under 1000 deg^2) at $z = 3$ would have a side length of $> 2 h^{-1} \text{ Gpc}$ (comoving) and require $\sim 10^{12}$ particles in order to resolve such halos. Covering more area, or more cleanly resolving the host halo population, would increase the demands even further. These requirements are relaxed by assuming a different form of the HOD (Appendix D), but remain very demanding even in this case. This suggests that it will be necessary to carefully plan what simulations are really required for what tasks.

To the extent that our sample is characteristic of samples that will be utilized by further surveys, and that our mock catalogs approximate the true halo occupation of such samples, we can use clustering statistics measured from the mocks to inform us about the promise and modeling challenges ahead. Our samples have a lower bias than dropout-selected LBGs at similar redshifts, have lower scale-dependent bias, are more correlated with the linear density field and have smaller fingers-of-god than those objects. In the long term these properties make our targets well suited for 3D clustering to constrain cosmology. In the more immediate term we forecast that 3000 deg^2 of IBIS data, in combination with CMB lensing, would allow a very strong measurement of the clustering amplitude of fluctuations at $z = 2 - 4$.

7 Data Availability

Software used for the analysis in this work are publicly available at <https://github.com/HarukiEbina/IBIS-HOD>.

Acknowledgments

HE and MW are supported by the DOE Office of Science under DE-SC0025523. We thank Hendrik Hildebrandt for sharing the LBG galaxy catalog for cross-matching. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF’s National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science, Innovation

and Universities of Spain (MICIU/AEI/10.13039/501100011033), and by the DESI Member Institutions: <https://www.desi.lbl.gov/collaborating-institutions>. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies. The authors are honored to be permitted to conduct scientific research on I’oligam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

This research draws upon DECam data as distributed by the Astro Data Archive at NSF NOIRLab. NOIRLab is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the U.S. National Science Foundation.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the US Department of Energy, the US National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute for Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas–Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF NOIRLab, the University of Nottingham, the Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

Based on observations at NSF Cerro Tololo Inter-American Observatory, a program of NOIRLab (NOIRLab Prop. ID 2023B-184194; PI: A. Dey and D. Schlegel), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the U.S. National Science Foundation.

A Relevant scales

In this appendix we briefly summarize some of the key physical scales entering into our calculations since at higher redshift different scales can dominate than at lower redshift where much of our intuition has been built. We show these scales in Fig. 12, as their evolution has important implications for the way in which survey design and survey non-ideality inform our analyses.

The non-linear scale can be estimated from the (comoving) Zeldovich displacement

$$\Sigma_{\text{Zel}}^2(z) = \int \frac{dk}{6\pi^2} P_{\text{lin}}(k, z) \simeq [5.87 D(z) h^{-1} \text{Mpc}]^2 \quad (\text{A.1})$$

where $D(z)$ is the linear growth rate, normalized to $D(z=0) = 1$ and the numerical value is for the Planck cosmology [88]. We expect this scale to show the limits of perturbative models, and also to describe the characteristic scale at which the matter field decorrelates from the linear field (as described in Fig. 10 the galaxy field can decorrelate at larger scales due to scale-dependent bias).

In a spatially flat Universe the comoving distance subtended by angle θ is simply $\chi\theta$ (with χ the comoving distance). This scale is important for selection effects that operate in angle, such as fiber assignment in a multi-object, fiber-fed spectrograph. Constraints on fiber placement depend upon several factors, but lead to an artificial suppression of pairs below a particular angular scale, which for DESI is around 0.05° (e.g. Fig. 2 of ref. [116]). At $z > 2$ the distance subtended exceeds Σ_{Zel} , our estimate of the non-linear scale.

Expressing redshift uncertainties as $\Delta v/c = \Delta z/(1+z)$ and converting to comoving distance uncertainty assuming Hubble's law we have

$$\sigma_z = \frac{c \Delta z}{H(z)} = \frac{c(1+z)}{H(z)} \frac{\Delta v}{c} \quad (\text{A.2})$$

If the distribution of redshift errors were Gaussian, this would lead to a damping of power along the line of sight by $\exp[-k_{\parallel}^2 \sigma_z^2]$. Our galaxies are much fainter than a typical low- z target and, in general, it is harder to provide precise redshifts for broad emission lines that are affected by radiative transfer and possible bulk flow of the emitting gas. The characteristic redshift error is thus a larger contributor to the total error budget than is common for low- z galaxies in spectroscopic surveys [6, 7]. From repeat observations of galaxies and comparison to external data, DESI has estimated characteristic redshift errors of $\approx 200 \text{ km s}^{-1}$ [29, 31]. If this continues to hold, line-of-sight power damping due to redshift errors will remain subdominant to the impacts from non-linearity (scaling as $(1+f)\Sigma_{\text{Zel}} \simeq 2\Sigma_{\text{Zel}}$) at the redshifts of interest.

One of the largest sources of non-linearity in the redshift-space clustering of galaxies is stochastic terms, often referred to as fingers of god. The size of this term is driven by the satellite fraction and the virial velocity of a halo that hosts on average one satellite galaxy [101], but we can get a feel for the scales involved by considering the virial velocity of an M_{\star} halo (i.e. a halo of characteristic mass in the halo mass function). A standard definition has $M_{\star} = (4\pi/3)\bar{\rho}_m R_{\star}^3$ where R_{\star} is the comoving radius such that $\sigma_{\text{lin}}(R_{\star}, z) = \delta_c \simeq 1.686$. If we measure our halo masses such that r_{vir} encloses a mean density of $200\times$ the background density then $R_{\star} = (200)^{1/3} r_{\text{vir}}$ for M_{\star} halos. The virial velocity associated with such a halo is $\sqrt{GM_{\star}/r_{\text{vir,phys}}} = \sqrt{1+z}\sqrt{GM_{\star}/r_{\text{vir,com}}}$, where we have distinguished ‘physical’ and ‘comoving’ radii. Using the Friedman equation to rewrite the $(8\pi G/3)\bar{\rho}_m$ in the GM_{\star} term and converting from velocity to distance units we have¹⁹

$$\sigma_{\star} = \sqrt{\Omega_m(z)} 10 r_{\text{vir,com}} = \frac{10 \sqrt{\Omega_m(z)}}{(200)^{1/3}} R_{\star} \simeq 1.71 R_{\star} \quad (\text{A.3})$$

¹⁹If instead we measured our halo masses using the critical density definition the $\sqrt{\Omega_m(z)}$ would not appear.

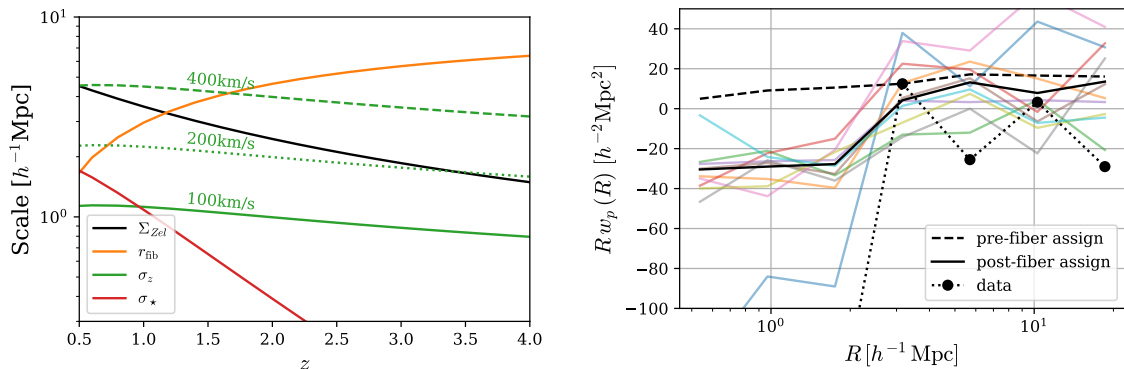


Figure 12: Left: Comparison of different ‘characteristic’ scales as a function of redshift. The black solid line is an estimate of the non-linear scale, Σ_{Zel} . The orange line is the transverse distance subtended by the DESI ‘fiber exclusion’ scale, 0.05° . The green lines are the limitations on line-of-sight modes due to redshift error, with dashed, dotted, and solid corresponding to 400, 200, and 100 km s^{-1} errors, respectively. The red line is the virial velocity of the characteristic halo mass. Right: Projected clustering measurement w_p computed from a mock fiber assignment. We fiber assign using dithering, inducing a complex selection effect on the clustering and causing the post-assignment measurement (faint solid lines with mean shown as the black solid line) to deviate significantly from the pre-assignment measurement (mean shown as the black dashed line). The post-assignment clustering qualitatively reproduces the large-scale anti-correlation in the data (black circles) including the sign change at large R .

where the last step assumes the observations are at high enough z that $\Omega_m(z) \simeq 1$. By $z \simeq 3$ this characteristic scale has shifted to very low masses, and shifts very rapidly to even higher redshift due to the flatness of $k^3 P_{\text{lin}}$ at high k , so the inferred σ_* is quite low.

When considering σ_* we should bear in mind that our galaxies are quite biased and thus inhabit halos more massive than M_* [117]. Indeed observations suggest that the most likely host halo mass does not evolve strongly with redshift [85], unlike M_* . On the other hand, should the satellite fraction be low and the central galaxies moving with the center of mass of their host halo the virial dispersion is much reduced. Thus σ_* should be taken as an ‘order of magnitude’ estimate. To the extent that our HOD fits reliably predict the halo occupancy of our sample a more refined estimate can be obtained from α_2 and N_2 .

Finally, we have omitted the virial radius of the halos hosting our galaxies as (a) it is a property of the samples themselves and not the underlying cosmology and (b) it is constant with redshift. However a $10^{12} h^{-1} M_\odot$ halo has $r_{\text{vir,com}} \simeq 200 h^{-1} \text{kpc}$ which suggests that the 1-halo term [85] remains very far below the non-linear scale and well below the reach of any of the clustering measurements described in this paper. The Lagrangian radius of such halos (at $200^{1/3} r_{\text{vir,com}}$) is however more comparable to the non-linear scale (Σ_{Zel}) at high z , indicating that the complexities of scale-dependent bias are expected to become more important than gravitational non-linearity in the early Universe [6, 118].

B Fiber assignment

The DESI spectroscopy provides redshifts for a fraction of the galaxy sample, in principle allowing us to perform a 3D clustering analysis (in redshift space). However, the effects of fiber assignment make this difficult. Although fiber assignment complications are not new to modern spectroscopy experiments, the situation here is unique due to two distinct reasons. The first is the high redshift of our sample. As fiber assignment, being a property of the focal plane, excludes pairs in angular coordinates, moving to higher redshift directly implies that the physical scale for fiber assignment subtends a larger transverse distance. When combined with the smaller non-linear scales at high redshift, fiber assignment may potentially be a much more difficult problem at high redshift than at low redshift. The second is the particular fiber assignment scheme adopted for the spectroscopic follow-up in this study (§3.3). The fiber assignment was optimized to maximize the limited observation time, rather than being tailored to clustering studies, and in the process introduces density-dependent selections.

DESI spectra are collected by assigning fibers to individual targets on the sky and the fibers are inherently limited in their proximity by the physical constraints of the focal plane. In any single observation we cannot place fibers on two objects within $\sim 0.05^\circ$ of each other. At low redshifts, this translates to relatively small scales ($r \lesssim 1 h^{-1}\text{Mpc}$), whereas at $z \sim 3$, it expands to around $r \sim 4 h^{-1}\text{Mpc}$, shifting the information lost due to fiber assignment from halo/sub-halo scales to larger, perturbative scales. This redshift-dependent change in scale is illustrated in Figure 12.

Such effects are not uncommon and many techniques have been proposed to deal with them (see e.g. §4 of [119]). One common method is to use weights for each object (sometimes referred to as independent inverse probability – IIP – weights) or each pair of objects (pairwise inverse probability – PIP – weights). A typical weight could e.g. be calculated from the frequency of observation across many iterations of a stochastic fiber assignment process [120]. An alternative, data-driven approach simply assigns redshifts of unobserved galaxies to that of their closest neighbor with a successful redshift measurement or increments the weight of that object in the clustering [121]. Other methods upweight close pairs based on the angular clustering [122] or use alternative clustering statistics that exclude pairs that are close in angle [116, 119].

Another approach is to give up on the full 3D information, and attempt to measure the projected correlation function, $w_p(R) = \int dr_{\parallel} \xi(R, r_{\parallel})$, via cross-correlation between the target (imaging) sample and spectroscopic sample. By using the target sample for one member of the pair, one is immune to pairwise spectroscopic incompleteness. The projected correlation function can be expressed as a weighted sum over angular cross-correlation measurements [73, 74]. However, a critical assumption in this computation is that the subsample of objects for which redshifts were obtained ‘fairly’ samples the full distribution. We will see that in our case this assumption is invalid.

All of these methods to mitigate fiber assignment effects are ineffective for the samples at hand due to the unique fiber assignment scheme that was adopted for these observations. In particular, the dependence of the final fiber assignment on the first round of fiber assignment introduces a density-dependent sampling, as shown in Figure 13 where we show the distribution of the local density measure δ defined by the normalized count of mock galaxies within a fiber assignment radius.

These difficulties even appear when attempting to correct for fiber assignment using the spectroscopic-photometric cross-correlation introduced above. We run the dithering fiber

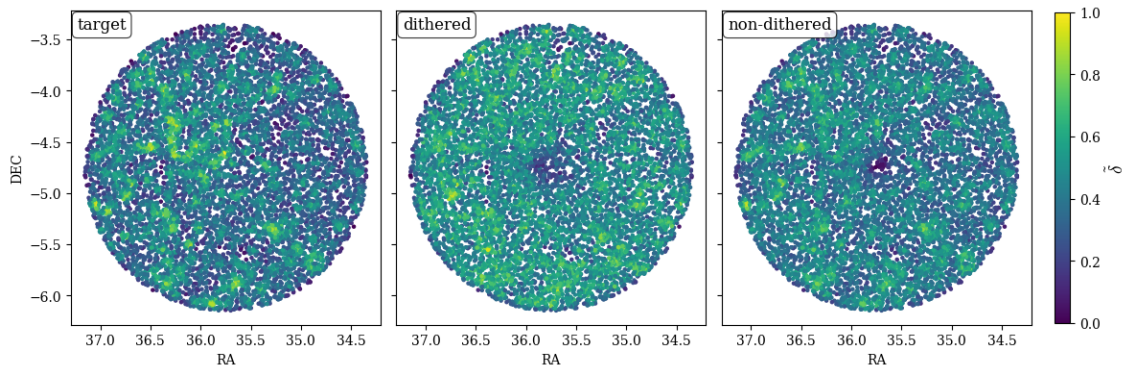


Figure 13: The impact of fiber assignment. We show, using a mock catalog, an estimate of the projected density, $\tilde{\delta}$, across the field for: the original target sample (left), objects assigned a fiber with dithered fiber assignment (this work, center), and with non-dithered fiber assignment (typical, right). The projected density estimate, $\tilde{\delta}$, is defined by the number of galaxies of the given type in a circle of radius 0.05° (the fiber collision scale), normalized to run from 0 to 1. Note that the dithered fiber assignment scheme produces notably smoother $\tilde{\delta}$ distributions than inherent in the target sample.

assignment over a mock catalog²⁰ and compare the pre- and post-fiber assignment w_p from cross-correlation. As shown in the right panel of Figure 12, the data show not only “small-scale” pairwise incompleteness up to $R \sim 4h^{-1}\text{Mpc}$, but also significant anti-correlation at large scales. Both effects are unphysical, and the latter, in particular, cannot be corrected by any method that upweights close pairs. We can qualitatively reproduce these effects using mock catalogs, although we do not expect precise agreement as the complex sampling effect likely introduces dependencies on aspects of the mocks that are not tuned to be correct (e.g. higher order moments of the clustering or close-pair statistics like counts in cells). Given these results, we deem that a correct modeling of this complex sampling at high redshift requires far more precise and accurate mocks than we are able to produce at present. Instead we use the angular clustering within the medium-band-selected shells (Appendix C), whose interpretation depends only upon knowing dN/dz and f_{int} and not on an understanding of the effects of fiber assignment.

C Angular clustering in medium-band selected samples

Recall that if $f(\chi)$ defines the (normalized) redshift distribution of an individual shell, then the auto-correlation can be written as

$$w(\theta) = \int d\chi_1 d\chi_2 f(\chi_1) f(\chi_2) \xi\left(\sqrt{\Delta^2 + \chi_1 \chi_2 \tilde{\omega}}\right) \quad (\text{C.1})$$

where $\Delta = \chi_2 - \chi_1$, $\tilde{\omega} = 2 \sin(\theta/2)$, we have assumed that the correlation function does not evolve over the shell and that projection effects will ‘wash out’ redshift-space distortions. We can approximate $\chi_1 \chi_2 \simeq \bar{\chi}^2$, with $\bar{\chi} \equiv (1/2)(\chi_1 + \chi_2)$, up to corrections of order $(\Delta/\bar{\chi})^2$. Since $\Delta < 0.1 \bar{\chi}$ in our case this approximation introduces a sub-percent error.

²⁰Note that this mock isn’t one of the best-fit catalogs introduced in the main text, as the fiber assignment effect must be modeled over the observed catalog, not the polished clustering catalog in this work.

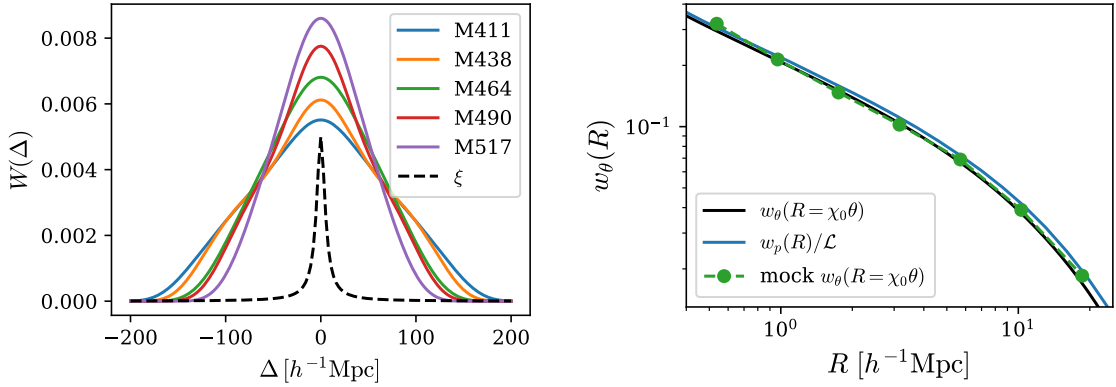


Figure 14: (Left) The line-of-sight window function, $W(\Delta)$, for the samples selected in our 5 filters. For comparison, the dashed black line shows $\xi(r) \propto (R^2 + \Delta^2)^{-1.8/2}$ as a function of Δ for $R = 5 h^{-1}\text{Mpc}$ and arbitrary normalization. (Right) Comparison between our medium-band estimator, $w_\theta(R)$, averaged over our mock fields and the projected correlation function, $w_p(R)/\mathcal{L}$, computed from the full periodic $1h^{-1}\text{Gpc}$ box. The HOD model used is the best-fit model for the total sample at $z = 3$. The solid black line is the result of Eq. (C.1) using ξ measured from the periodic box. The green dots show the average result of the mock pipeline for the M490 filter (with error bars suppressed), while the blue line then shows the approximation of Eq. (C.6). With respect to our errors, the estimator is comfortably consistent with the true $w_p(R)$. Since we use this approximation only to motivate our data combination, and not to actually perform fits to the data, we believe the degree of agreement is totally adequate.

The Jacobian for the change of variables $(\chi_1, \chi_2) \rightarrow (\bar{\chi}, \Delta)$ is unity, so we immediately obtain

$$w_\theta(R) \simeq \int d\bar{\chi} d\Delta f(\bar{\chi} - \Delta/2) f(\bar{\chi} + \Delta/2) \xi(\bar{\chi} \tilde{\omega}, \Delta) \quad (\text{C.2})$$

$$\approx \int d\bar{\chi} d\Delta f(\bar{\chi} - \Delta/2) f(\bar{\chi} + \Delta/2) \xi(\chi_0 \theta, \Delta) \quad (\text{C.3})$$

$$= \int_{-\infty}^{\infty} d\Delta W(\Delta) \xi(R, \Delta) \quad (\text{C.4})$$

where in the second line we have converted angles to (transverse) distances with $R = \chi_0 \theta$ rather than $\bar{\chi} \tilde{\omega}$ and we have defined

$$W(\Delta) = \int_0^{\infty} d\bar{\chi} f(\bar{\chi} - \Delta/2) f(\bar{\chi} + \Delta/2) \quad . \quad (\text{C.5})$$

In the above we have used the small-angle approximation to clarify that the arguments of ξ are r_\perp and r_\parallel with the integral over r_\parallel . Corrections to the small-angle approximation are expected to be $\mathcal{O}(\theta^2)$, and are entirely negligible in our case. Not negligible is the replacement $\bar{\chi} \rightarrow \chi_0$, i.e. that we can convert angles to transverse distances using the shell center (χ_0) rather than $\bar{\chi}$. This approximation is subdominant to our errors for our situation, where we are primarily determining how to combine disjoint slices, but introduces an error of $\mathcal{O}(\Delta\chi/\chi)$ that also depends upon the slope of ξ .

The line-of-sight window function, $W(\Delta)$, is the convolution of the radial selection function with itself. For a tophat f this would be a triangle function of twice the width (but the same overall area). We show examples of $W(\Delta)$ for our slices in Fig. 14. Note the width of W is sufficiently large that we can approximate

$$w_\theta(R) = \int_{-\infty}^{\infty} d\Delta W(\Delta) \xi(R, \Delta) \approx W(0) \int d\Delta \xi(R, \Delta) = \mathcal{L}^{-1} w_p(R) \quad (\text{C.6})$$

as in the main text. For transverse scales much smaller than the characteristic scale of variation of W the impact of redshift-space distortions is minimal.

Figure 14, right panel, shows how closely our mock $w_\theta(R)$ matches $\mathcal{L}^{-1} w_p(R)$ for the M490 sample. The solid black line shows the result of Eq. (C.1) using ξ measured from the periodic box and can be seen to be in good agreement with the points that are estimated from the average of many mock observations. The impact of the approximations made in going from Eq. (C.1) to Eq. (C.6) can be seen in comparison to the blue line. While the approximations are noticeably imperfect, they are well within the observational error bars (suppressed in Fig. 14 but the same as in Fig. 4). It is important to note that in computing the HOD/theory predictions we follow the same procedure for the data and the mocks, so the relevant comparison is the green points to the black line. The approximation in Eq. (C.6) is used only to motivate the manner in which we combine the samples adjacent in redshift (i.e. going from Fig. 4 to Fig. 5). This may lead to non-optimal errors, but should introduce no bias.

D Extended models of HOD

In this section we consider the High Mass Quenched (HMQ) HOD model in addition to the standard HOD model explored in the main text (§4.2). The HMQ model was initially proposed to model BOSS/eBOSS and DESI Emission Line Galaxies (ELGs) [123]. These ELGs are actively star forming galaxies with prominent OII. As star formation is inefficient near the center of massive halos (at lower z) the HMQ model, which inhibits (very) massive halos from hosting central galaxies, were designed to better match our empirical understanding of ELG occupancy. The LAEs and LBGs that we consider in this work are also known to be actively star forming galaxies and hence merit analysis under this extended model. In addition, recent studies independently report results supporting this picture. Ref. [87] has found that hydrodynamical simulations that fit to existing LAE data show a decline in central galaxy occupation past a halo mass of $\sim 10^{11} h^{-1} M_\odot$, while ref. [46] has found that RT of Lyman alpha photons can significantly suppress the observation of central LAEs in massive halos.

In the HMQ model, the halo occupation of central and satellite galaxies are defined as

$$\langle N_{\text{cen}}(M) \rangle = 2A\phi(M)\Phi(\gamma M) + \frac{1}{2Q} \left[1 + \text{erf} \left(\frac{\log_{10} M/M_{\text{cut}}}{0.01} \right) \right] \quad (\text{D.1})$$

$$\langle N_{\text{sat}}(M) \rangle = \left(\frac{M - \kappa M_{\text{cut}}}{M_1} \right)^\alpha \quad \text{for } M > \kappa M_{\text{cut}} \quad (\text{D.2})$$

where M is the halo mass, ϕ and Φ are normal distribution and cumulative distribution

$$\phi(x) = \mathcal{N}(\log_{10} M_{\text{cut}}, \sigma_M) \quad , \quad \Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (\text{D.3})$$

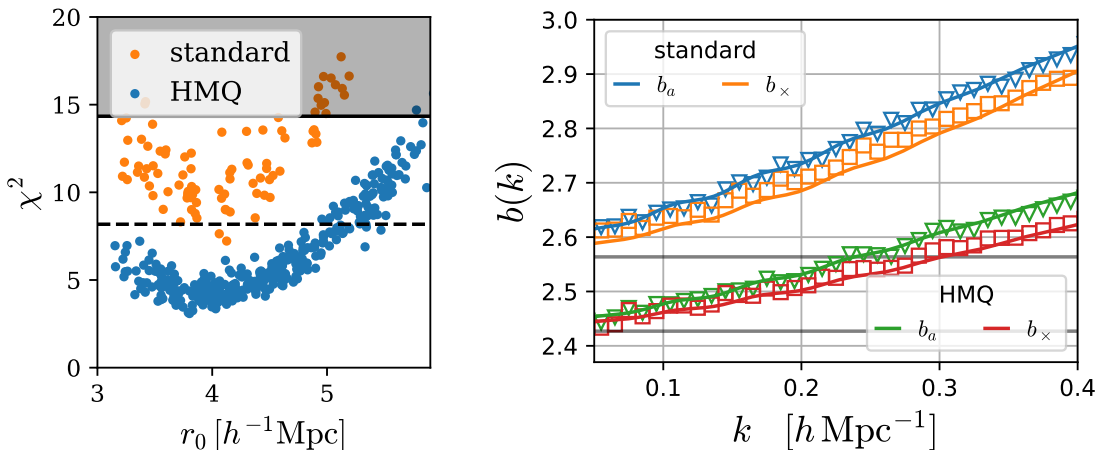


Figure 15: Left: The correlation length vs. goodness-of-fit for the ‘total’ sample at $z = 3$, using both the standard HOD (§4.2) and HMQ HOD model. The 1σ and 2σ CL for 7 d.o.f. are shown in dashed and solid lines, respectively, with the region above 2σ shaded in grey. We find that in both cases r_0 serves as an effective compressed statistic encoding the complexity of the HOD models, but the HMQ model is able to probe better fits, at lower r_0 . Right: The scale-dependent bias of the standard and HMQ HOD models. Both the auto- ($b_a = \sqrt{P_{gg}/P_{mm}}$) and cross-bias ($b_\times = P_{gm}/P_{mm}$) are shown, with their ratio b_a/b_\times showing the correlation between the galaxy and non-linear matter fields.

while $A = [p_{\max} - 1/Q]/\max[2\phi(x)\Phi(\gamma x)]$ and $\{\sigma_M, p_{\max}, \gamma, Q\}$ are the new model parameters, with σ_M substituting σ . The amplitude and width of the central HOD are controlled by p_{\max} and σ_M while γ controls the low- M turn-on and Q the limit of $\langle N_{\text{cen}} \rangle$ as $M \rightarrow \infty$.

While a detailed comparison between different flavors of HOD models to investigate which models align better with high- z galaxies is beyond the scope and precision of this work, we can compare the mocks made with the HMQ and ‘standard’ HODs to gain some insights into how sensitive our inferences are to the HOD form. For this we use the ‘total’ $z = 3$ sample, where the standard HOD results indicate that our simulation is most reliable, in terms of mass resolution.

The left panel of Figure 15 shows χ^2 vs. r_0 for a grid²¹ of HMQ models, compared to the grid for the standard form discussed in the text. We see the same general trends, with the HMQ models providing better fits than the standard HOD with less scatter; aligning with the results of refs. [46, 87]. These results indicate that a single-parameter compression via r_0 is insufficient to capture the differences between the two HOD models. In the right panel we compare the auto- and cross-spectrum biases for the best fits from the two grids of models. The behavior is very similar, with approximately the same degree of scale-dependence and the same level of agreement that $b_a \approx b_\times$, though the best-fitting HMQ model has a slightly ($< 10\%$) lower large-scale bias and scale-dependence than the standard form. These differences are well within the error bars of our observations, and reflect (at least in part) the limitations of the data and our relatively coarse model grid. To the extent that the large-scale bias, the

²¹We choose to vary between $\sigma_M = \{0.33, 0.66\}$, $\alpha = \{0.33, 0.66\}$, $Q = \{20, 100\}$, $\gamma = \{1, 5\}$, and $p_{\max} = \{0.33, 0.66\}$, while sampling $\log_{10} M_{\text{cut}}/(h^{-1}M_\odot)$ in steps of 0.25 between 11.0 and 12.0 and M_1/M_{cut} in $\{10, 30, 100\}$.

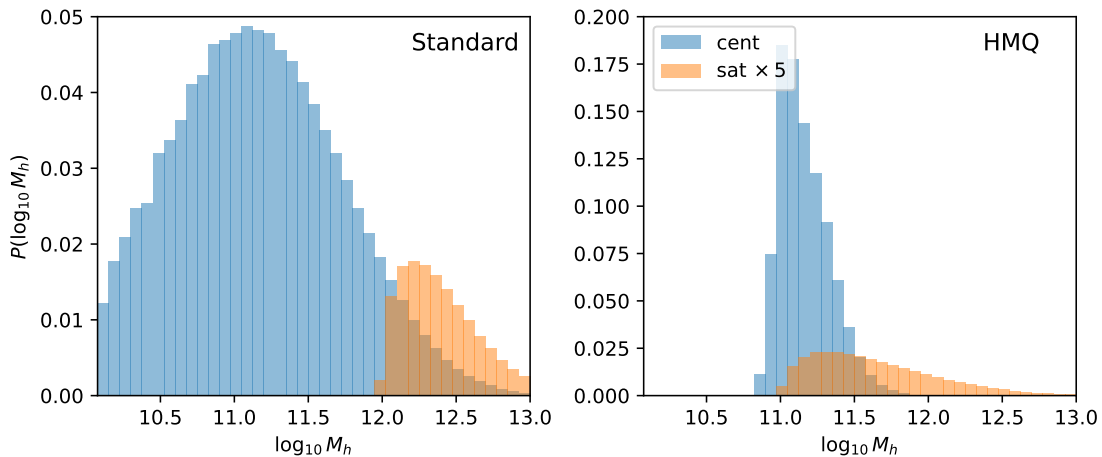


Figure 16: The host halo mass of central and satellite galaxies in the best-fit mock catalogs for the total sample at $z = 3.0$, with the standard HOD considered in the main text (left) and the HMQ HOD considered in this appendix (right). The total (central + satellite) distribution is normalized to integrate to unity and the satellite distribution is multiplied by 5 for visualization. The HMQ model predicts that the satellite galaxies are primarily hosted in low-mass halos, although with a significantly broad distribution extending to high-mass halos.

scale dependence of the bias and the degree of decorrelation are among the principal factors impacting the use of tracers for cosmological inference it appears that our main conclusions are robust to changes in the functional form of the HOD.

Figure 16 shows the distribution of host halo masses for our mock galaxies for the two HOD forms. In each case we have broken the distribution into that for the central and the satellite galaxies. Modeling the galaxies within the standard form requires us to resolve significantly lower mass halos than for the HMQ form. The satellites in the HMQ model are less numerous ($f_{\text{sat}} = 0.06$) than the standard model ($f_{\text{sat}} = 0.10$, Table 2) and reside in a much wider range of halo masses, suppressing the FoG effect. Further comparison between HOD models, semi-analytic prescriptions and with hydrodynamic simulations will be valuable for future simulation based studies and mock catalog construction of high- z galaxies [39, 46, 86, 87].

References

- [1] DESI Collaboration, A. Aghamousa, J. Aguilar, et al., *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *ArXiv e-prints* (Oct., 2016) [[arXiv:1611.00036](#)].
- [2] Euclid Collaboration, Y. Mellier, Abdurro’uf, et al., *Euclid. I. Overview of the Euclid mission*, *arXiv e-prints* (May, 2024) arXiv:2405.13491, [[arXiv:2405.13491](#)].
- [3] M. Takada, R. S. Ellis, M. Chiba, et al., *Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph*, *Publ. Astron. Soc. Japan* **66** (Feb., 2014) R1, [[arXiv:1206.0737](#)].
- [4] DESI Collaboration, M. Abdul-Karim, A. G. Adame, et al., *Data Release 1 of the Dark*

- Energy Spectroscopic Instrument*, *arXiv e-prints* (Mar., 2025) arXiv:2503.14745, [[arXiv:2503.14745](#)].
- [5] DESI Collaboration, M. Abdul-Karim, J. Aguilar, et al., *DESI DR2 Results II: Measurements of Baryon Acoustic Oscillations and Cosmological Constraints*, *arXiv e-prints* (Mar., 2025) arXiv:2503.14738, [[arXiv:2503.14738](#)].
- [6] M. J. Wilson and M. White, *Cosmology with dropout selection: straw-man surveys & CMB lensing*, *Journal of Cosmology and Astro-Particle Physics* **2019** (Oct., 2019) 015, [[arXiv:1904.13378](#)].
- [7] N. Sailer, E. Castorina, S. Ferraro, and M. White, *Cosmology at high redshift — a probe of fundamental physics*, *JCAP* **12** (2021), no. 12 049, [[arXiv:2106.09713](#)].
- [8] D. J. Schlegel, S. Ferraro, G. Aldering, et al., *A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5*, *arXiv e-prints* (Sept., 2022) arXiv:2209.03585, [[arXiv:2209.03585](#)].
- [9] S. Ferraro, N. Sailer, A. Slosar, and M. White, *Snowmass2021 Cosmic Frontier White Paper: Cosmology and Fundamental Physics from the three-dimensional Large Scale Structure*, *arXiv e-prints* (Mar., 2022) arXiv:2203.07506, [[arXiv:2203.07506](#)].
- [10] R. Besuner, A. Dey, A. Drlica-Wagner, et al., *The Spectroscopic Stage-5 Experiment*, *arXiv e-prints* (Mar., 2025) arXiv:2503.07923, [[arXiv:2503.07923](#)].
- [11] R. Bacon, V. Mainieri, S. Randich, et al., *WST – Widefield Spectroscopic Telescope: Motivation, science drivers and top-level requirements for a new dedicated facility*, *arXiv e-prints* (May, 2024) arXiv:2405.12518, [[arXiv:2405.12518](#)].
- [12] M. Ouchi, Y. Ono, and T. Shibuya, *Observations of the Lyman- α Universe*, *Ann. Rev. Astron. & Astrophys.* **58** (Aug., 2020) 617–659, [[arXiv:2012.07960](#)].
- [13] M. Giavalisco, *Lyman-Break Galaxies*, *Ann. Rev. Astron. & Astrophys.* **40** (2002) 579–641.
- [14] A. E. Shapley, *Physical Properties of Galaxies from $z = 2-4$* , *Ann. Rev. Astron. & Astrophys.* **49** (Sept., 2011) 525–580, [[arXiv:1107.5060](#)].
- [15] Y. Ono, M. Ouchi, Y. Harikane, et al., *Great Optically Luminous Dropout Research Using Subaru HSC (GOLDRUSH). I. UV luminosity functions at $z \sim 4-7$ derived with the half-million dropouts on the 100 deg² sky*, *Publ. Astron. Soc. Japan* **70** (Jan., 2018) S10, [[arXiv:1704.06004](#)].
- [16] N. M. Förster Schreiber and S. Wuyts, *Star-Forming Galaxies at Cosmic Noon*, *Ann. Rev. Astron. & Astrophys.* **58** (Aug., 2020) 661–725, [[arXiv:2010.10171](#)].
- [17] Y. Harikane, Y. Ono, M. Ouchi, et al., *GOLDRUSH. IV. Luminosity Functions and Clustering Revealed with 4,000,000 Galaxies at $z = 2-7$: Galaxy-AGN Transition, Star Formation Efficiency, and Implication for Evolution at $z \lesssim 10$* , *Astrophys. J. Suppl.* **259** (Mar., 2022) 20, [[arXiv:2108.01090](#)].
- [18] M. Ouchi, K. Shimasaku, M. Akiyama, et al., *The Subaru/XMM-Newton Deep Survey (SXDS). IV. Evolution of Ly α Emitters from $z = 3.1$ to 5.7 in the 1 deg² Field: Luminosity Functions and AGN*, *Astrophys. J. Suppl.* **176** (June, 2008) 301–330, [[arXiv:0707.3161](#)].
- [19] A. Konno, M. Ouchi, K. Nakajima, et al., *Bright and faint ends of ly α luminosity functions at $z = 2$ determined by the subaru survey: Implications for agns, magnification bias, and ism h i evolution*, *The Astrophysical Journal* **823** (may, 2016) 20.
- [20] M. A. Malkan, D. P. Cohen, M. Maruyama, et al., *Lyman-break galaxies at $z \sim 3$ in the subaru deep field: Luminosity function, clustering, and [o iii] emission*, *The Astrophysical Journal* **850** (nov, 2017) 5, [[arXiv:1711.04787](#)].
- [21] D. Sobral, S. Santos, J. Matthee, et al., *Slicing COSMOS with SC4k: the evolution of typical*

ly α emitters and the ly α escape fraction from $z \sim 2$ to 6, *Monthly Notices of the Royal Astronomical Society* **476** (feb, 2018) 4725–4752.

- [22] M. Ouchi, Y. Harikane, T. Shibuya, et al., *Systematic Identification of LAEs for Visible Exploration and Reionization Research Using Subaru HSC (SILVERRUSH). I. Program strategy and clustering properties of ~ 2000 Ly α emitters at $z = 6-7$ over the $0.3-0.5$ Gpc 2 survey area*, *Publ. Astron. Soc. Japan* **70** (Jan., 2018) S13, [[arXiv:1704.07455](#)].
- [23] K.-S. Lee, E. Gawiser, C. Park, et al., *The one-hundred-deg 2 decam imaging in narrowbands (odin): Survey design and science goals*, 2023.
- [24] O. Le Fevre, P. Cassata, O. Cucciati, et al., *The VIMOS VLT Deep Survey final data release: a spectroscopic sample of 35016 galaxies and AGN out to $z \sim 6.7$ selected with $17.5_j=i_{-}\{AB\}_j=24.7$* , *arXiv e-prints* (July, 2013) arXiv:1307.0545, [[arXiv:1307.0545](#)].
- [25] E. Mentuch Cooper, K. Gebhardt, D. Davis, et al., *HETDEX Public Source Catalog 1: 220 K Sources Including Over 50 K Ly α Emitters from an Untargeted Wide-area Spectroscopic Survey*, *Astrophys. J.* **943** (Feb., 2023) 177, [[arXiv:2301.01826](#)].
- [26] M. White, A. Raichoor, A. Dey, et al., *The clustering of Lyman Alpha Emitting galaxies at $z=2-3$* , *Journal of Cosmology and Astro-Particle Physics* **2024** (Aug., 2024) 020, [[arXiv:2406.01803](#)].
- [27] H. Hildebrandt, J. Pielorz, T. Erben, et al., *CARS: the CFHTLS-Archive-Research Survey. II. Weighing dark matter halos of Lyman-break galaxies at $z = 3-5$* , *Astron. Astrophys.* **498** (May, 2009) 725–736, [[arXiv:0903.3951](#)].
- [28] O. Le Fèvre, L. A. M. Tasca, P. Cassata, et al., *The VIMOS Ultra-Deep Survey: 10 000 galaxies with spectroscopic redshifts to study galaxy assembly at early epochs $2 < z \leq 6$* , *Astron. Astrophys.* **576** (Apr., 2015) A79, [[arXiv:1403.3938](#)].
- [29] V. Ruhlmann-Kleider, C. Yèche, C. Magneville, et al., *High redshift LBGs from deep broadband imaging for future spectroscopic surveys*, *Journal of Cosmology and Astro-Particle Physics* **2024** (Aug., 2024) 059, [[arXiv:2404.03569](#)].
- [30] C. Payerne, W. d’Assignies Doumerg, C. Yèche, et al., *Selection of high-redshift Lyman-Break Galaxies from broadband and wide photometric surveys*, *arXiv e-prints* (Oct., 2024) arXiv:2410.08062, [[arXiv:2410.08062](#)].
- [31] A. Raichoor, *Lae lbg selection with supprime medium-bands photometry*, *In prep* (2025).
- [32] Planck Collaboration, N. Aghanim, Y. Akrami, et al., *Planck 2018 results. VI. Cosmological parameters*, *arXiv e-prints* (July, 2018) arXiv:1807.06209, [[arXiv:1807.06209](#)].
- [33] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, *Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds*, *Astrophys. J.* **500** (June, 1998) 525–553, [[astro-ph/9710327](#)].
- [34] R. B. Partridge and P. J. E. Peebles, *Are Young Galaxies Visible?*, *Astrophys. J.* **147** (Mar., 1967) 868.
- [35] P. Guhathakurta, J. A. Tyson, and S. R. Majewski, *A Redshift Limit for the Faint Blue Galaxy Population from Deep U Band Imaging*, *Astrophys. J. Lett.* **357** (July, 1990) L9.
- [36] K. Nagamine, M. Ouchi, V. Springel, and L. Hernquist, *Lyman- α Emitters and Lyman-Break Galaxies at $z = 3-6$ in Cosmological SPH Simulations*, *Publ. Astron. Soc. Japan* **62** (Dec., 2010) 1455, [[arXiv:0802.0228](#)].
- [37] M. Ouchi, K. Shimasaku, H. Furusawa, et al., *Statistics of 207 ly α emitters at a redshift near 7: Constraints on reionization and galaxy formation models**, *The Astrophysical Journal* **723** (oct, 2010) 869.

- [38] T. Garel, J. Blaizot, B. Guiderdoni, et al., *The UV, Lyman α , and dark matter halo properties of high-redshift galaxies*, *Mon. Not. R. Astron. Soc.* **450** (June, 2015) 1279–1294, [[arXiv:1503.06635](#)].
- [39] S. Gurung-López, Á. A. Orsi, S. Bonoli, et al., *Ly α emitters in a cosmological volume - I. The impact of radiative transfer*, *Mon. Not. R. Astron. Soc.* **486** (June, 2019) 1882–1906, [[arXiv:1807.00006](#)].
- [40] H. Miyatake, Y. Harikane, M. Ouchi, et al., *First Identification of a CMB Lensing Signal Produced by 1.5 Million Galaxies at $z \sim 4$: Constraints on Matter Density Fluctuations at High Redshift*, *Phys. Rev. Lett.* **129** (Aug., 2022) 061301, [[arXiv:2103.15862](#)].
- [41] H. Umeda, M. Ouchi, S. Kikuta, et al., *SILVERRUSH. XIV. Ly α Luminosity Functions and Angular Correlation Functions from 20,000 Ly α Emitters at $z \sim 2.2$ –7.3 from up to 24 deg² HSC-SSP and CHORUS Surveys: Linking the Postreionization Epoch to the Heart of Reionization*, *Astrophys. J. Suppl.* **277** (Apr., 2025) 37, [[arXiv:2411.15495](#)].
- [42] H. Ebina and M. White, *Cosmology before noon with multiple galaxy populations*, *Journal of Cosmology and Astro-Particle Physics* **2024** (June, 2024) 052, [[arXiv:2401.13166](#)].
- [43] D. Herrera, E. Gawiser, B. Benda, et al., *ODIN: Clustering Analysis of 14,000 Lyman Alpha Emitting Galaxies at $z=2.4$, 3.1, and 4.5*, *arXiv e-prints* (Mar., 2025) arXiv:2503.17824, [[arXiv:2503.17824](#)].
- [44] Z. Zheng, R. Cen, H. Trac, and J. Miralda-Escudé, *Radiative Transfer Modeling of Ly α Emitters. II. New Effects on Galaxy Clustering*, *Astrophys. J.* **726** (Jan., 2011) 38, [[arXiv:1003.4990](#)].
- [45] C. Behrens, C. Byrohl, S. Saito, and J. C. Niemeyer, *The impact of Lyman- α radiative transfer on large-scale clustering in the Illustris simulation*, *Astron. Astrophys.* **614** (June, 2018) A31, [[arXiv:1710.06171](#)].
- [46] H. Khoraminezhad, S. Saito, M. Gronke, and C. Byrohl, *Simulating realistic Lyman- α emitting galaxies including the effect of radiative transfer*, *arXiv e-prints* (July, 2025) arXiv:2507.16707, [[arXiv:2507.16707](#)].
- [47] R. Momose, K. Shimasaku, K. Nagamine, et al., *Catch Me if You Can: Biased Distribution of Ly α -emitting Galaxies according to the Viewing Direction*, *Astrophys. J. Lett.* **912** (May, 2021) L24, [[arXiv:2104.10580](#)].
- [48] R. J. Bouwens, G. Illingworth, R. S. Ellis, et al., *$z \sim 2$ –9 galaxies magnified by the hubble frontier field clusters. ii. luminosity functions and constraints on a faint-end turnover*, *The Astrophysical Journal* **940** (Nov., 2022) 55.
- [49] J. M. Sullivan, T. Prijon, and U. Seljak, *Learning to concentrate: multi-tracer forecasts on local primordial non-Gaussianity with machine-learned bias*, *Journal of Cosmology and Astro-Particle Physics* **2023** (Aug., 2023) 004, [[arXiv:2303.08901](#)].
- [50] J. F. Crenshaw, B. Leistedt, M. L. Graham, et al., *Quantifying the Impact of LSST u -band Survey Strategy on Photometric Redshift Estimation and the Detection of Lyman-break Galaxies*, *arXiv e-prints* (Mar., 2025) arXiv:2503.06016, [[arXiv:2503.06016](#)].
- [51] S. Miyazaki, Y. Komiyama, M. Sekiguchi, et al., *Subaru Prime Focus Camera – Suprime-Cam*, *Publ. Astron. Soc. Japan* **54** (Dec., 2002) 833–853, [[astro-ph/0211006](#)].
- [52] N. Kaifu, T. Usuda, S. S. Hayashi, et al., *The First Light of the Subaru Telescope: A New Infrared Image of the Orion Nebula*, *Publ. Astron. Soc. Japan* **52** (Feb., 2000) 1.
- [53] M. Iye, H. Karoji, H. Ando, et al., *Current Performance and On-Going Improvements of the 8.2 m Subaru Telescope*, *Publ. Astron. Soc. Japan* **56** (Apr., 2004) 381–397, [[astro-ph/0405012](#)].
- [54] A. Dey and D. Schlegel, *The intermediate band imaging survey*, [[arXiv:25MM.NNNNN](#)].

- [55] H. Aihara, N. Arimoto, R. Armstrong, et al., *The Hyper Suprime-Cam SSP Survey: Overview and survey design*, *Publ. Astron. Soc. Japan* **70** (Jan., 2018) S4, [[arXiv:1704.05858](#)].
- [56] B. Flaugher, H. T. Diehl, K. Honscheid, et al., *The Dark Energy Camera*, *AJ* **150** (Nov., 2015) 150, [[arXiv:1504.02900](#)].
- [57] D. A. Lang, J. Moustakas, E. F. Schlafly, et al., “legacypipe: Image reduction pipeline for DESI Legacy Imaging Surveys.” Astrophysics Source Code Library, record ascl:2502.024, Feb., 2025.
- [58] D. Lang, D. W. Hogg, and D. J. Schlegel, *WISE Photometry for 400 Million SDSS Sources*, *AJ* **151** (Feb., 2016) 36, [[arXiv:1410.7397](#)].
- [59] A. Dey, S. Saito, M. Gronke, and C. Byrohl, *High-redshift target selection using medium band photometry*, *In prep.* (2025).
- [60] Bailey et al. in preparation.
- [61] M. Levi, C. Bebek, T. Beers, et al., *The DESI Experiment, a whitepaper for Snowmass 2013*, *arXiv e-prints* (Aug., 2013) arXiv:1308.0847, [[arXiv:1308.0847](#)].
- [62] DESI Collaboration, A. Aghamousa, J. Aguilar, et al., *The DESI Experiment Part II: Instrument Design*, *arXiv e-prints* (Oct., 2016) arXiv:1611.00037, [[arXiv:1611.00037](#)].
- [63] DESI Collaboration, B. Abareshi, J. Aguilar, et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *AJ* **164** (Nov., 2022) 207, [[arXiv:2205.10939](#)].
- [64] J. H. Silber, P. Fagrelus, K. Fanning, et al., *The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)*, *AJ* **165** (Jan., 2023) 9, [[arXiv:2205.09014](#)].
- [65] T. N. Miller, P. Doel, G. Gutierrez, et al., *The Optical Corrector for the Dark Energy Spectroscopic Instrument*, *AJ* **168** (Aug., 2024) 95, [[arXiv:2306.06310](#)].
- [66] E. F. Schlafly, D. Kirkby, D. J. Schlegel, et al., *Survey Operations for the Dark Energy Spectroscopic Instrument*, *AJ* **166** (Dec., 2023) 259, [[arXiv:2306.06309](#)].
- [67] C. Poppett, L. Tyas, J. Aguilar, et al., *Overview of the Fiber System for the Dark Energy Spectroscopic Instrument*, *AJ* **168** (Dec., 2024) 245.
- [68] Raichoor et al. in preparation.
- [69] J. Guy, S. Bailey, A. Kremin, et al., *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *AJ* **165** (Apr., 2023) 144, [[arXiv:2209.14482](#)].
- [70] G. Desprez, V. Picouet, T. Moutard, et al., *Combining the CLAUDS and HSC-SSP surveys. $U + grizy(+YJHK_s)$ photometry and photometric redshifts for 18M galaxies in the 20 deg² of the HSC-SSP Deep and ultraDeep fields*, *Astron. Astrophys.* **670** (Feb., 2023) A82, [[arXiv:2301.13750](#)].
- [71] C. Gronwall, R. Ciardullo, T. Hickey, et al., *Ly α Emission-Line Galaxies at $z = 3.1$ in the Extended Chandra Deep Field-South*, *Astrophys. J.* **667** (Sept., 2007) 79–91, [[arXiv:0705.3917](#)].
- [72] A. E. Shapley, C. C. Steidel, M. Pettini, and K. L. Adelberger, *Rest-Frame Ultraviolet Spectra of $z \sim 3$ Lyman Break Galaxies*, *Astrophys. J.* **588** (May, 2003) 65–89, [[astro-ph/0301230](#)].
- [73] N. Padmanabhan, M. White, P. Norberg, and C. Porciani, *The real-space clustering of luminous red galaxies around $z \sim 0.6$ quasars in the Sloan Digital Sky Survey*, *Mon. Not. R. Astron. Soc.* **397** (Aug., 2009) 1862–1875, [[arXiv:0802.2105](#)].
- [74] A. D. Myers, M. White, and N. M. Ball, *Incorporating photometric redshift probability density information into real-space clustering measurements*, *Mon. Not. R. Astron. Soc.* **399** (Nov., 2009) 2279–2287, [[arXiv:0903.3121](#)].

- [75] S. D. Landy and A. S. Szalay, *Bias and Variance of Angular Correlation Functions*, *Astrophys. J.* **412** (July, 1993) 64.
- [76] A. D. Myers, J. Moustakas, S. Bailey, et al., *The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument*, *AJ* **165** (Feb., 2023) 50, [[arXiv:2208.08518](#)].
- [77] A. A. Khostovan, D. Sobral, B. Mobasher, et al., *The clustering of typical Ly α emitters from $z \sim 2.5$ –6: host halo masses depend on Ly α and UV luminosities*, *Monthly Notices of the Royal Astronomical Society* **489** (aug, 2019) 555–573.
- [78] R. Bielby, P. Tummuangpak, T. Shanks, et al., *The vlt lbg redshift survey - v. characterising the $z = 3.1$ lyman alpha emitter population*, *Monthly Notices of the Royal Astronomical Society* **456** (01, 2015).
- [79] E. Gawiser, H. Francke, K. Lai, et al., *Ly α -emitting galaxies at $z = 3.1$: L^* progenitors experiencing rapid star formation**, *The Astrophysical Journal* **671** (dec, 2007) 278.
- [80] R. M. Bielby, T. Shanks, P. M. Weilbacher, et al., *The VLT LBG Redshift Survey - I. Clustering and dynamics of ≈ 1000 galaxies at $z \approx 3$* , *Mon. Not. R. Astron. Soc.* **414** (June, 2011) 2–27, [[arXiv:1005.3028](#)].
- [81] A. Durkalec, O. Le Fèvre, A. Pollo, et al., *Evolution of clustering length, large-scale bias, and host halo mass at $2 < z < 5$ in the VIMOS Ultra Deep Survey (VUDS)**, *Astron. Astrophys.* **583** (Nov., 2015) A128, [[arXiv:1411.5688](#)].
- [82] Z. Zheng, A. L. Coil, and I. Zehavi, *Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering*, *Astrophys. J.* **667** (Oct., 2007) 760–779, [[astro-ph/0703457](#)].
- [83] N. A. Maksimova, L. H. Garrison, D. J. Eisenstein, et al., *AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations*, *Monthly Notices of the Royal Astronomical Society* **508** (09, 2021) 4017–4037, [<https://academic.oup.com/mnras/article-pdf/508/3/4017/40811763/stab2484.pdf>].
- [84] L. H. Garrison, D. J. Eisenstein, D. Ferrer, et al., *The ABACUS cosmological N-body code*, *Mon. Not. R. Astron. Soc.* **508** (Nov., 2021) 575–596, [[arXiv:2110.11392](#)].
- [85] R. H. Wechsler and J. L. Tinker, *The Connection Between Galaxies and Their Dark Matter Halos*, *Ann. Rev. Astron. & Astrophys.* **56** (Sept., 2018) 435–487, [[arXiv:1804.03097](#)].
- [86] J. Ravi, B. Hadzhiyska, M. J. White, et al., *Examining Lyman-alpha emitters through simulations in anticipation of the DESI-II survey*, *Phys. Rev. D* **110** (Nov., 2024) 103509, [[arXiv:2403.02414](#)].
- [87] J. M. Sullivan, C. Cuesta-Lazaro, M. M. Ivanov, et al., *High-redshift Millennium and Astrid galaxies in effective field theory at the field level*, *arXiv e-prints* (May, 2025) arXiv:2505.03626, [[arXiv:2505.03626](#)].
- [88] Planck Collaboration, Y. Akrami, F. Arroja, et al., *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *arXiv e-prints* (July, 2018) arXiv:1807.06205, [[arXiv:1807.06205](#)].
- [89] K. M. Górski, E. Hivon, A. J. Banday, et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *Astrophys. J.* **622** (Apr., 2005) 759–771, [[astro-ph/0409513](#)].
- [90] K. Kovač, R. S. Somerville, J. E. Rhoads, et al., *Clustering of Ly α Emitters at $z \sim 4.5$* , *Astrophys. J.* **668** (Oct., 2007) 15–22, [[arXiv:0706.0893](#)].
- [91] H. Kusakabe, K. Shimasaku, M. Ouchi, et al., *The stellar mass, star formation rate and dark matter halo properties of LAEs at $z \sim 2$* , *Publ. Astron. Soc. Japan* **70** (Jan., 2018) 4, [[arXiv:1707.09373](#)].

- [92] C.-N. Hao, J.-S. Huang, X. Xia, et al., *A Deep Ly α Survey in ECDF-S and COSMOS. I. General Properties of Ly α Emitters at $z \sim 2$* , *Astrophys. J.* **864** (Sept., 2018) 145, [[arXiv:1808.02704](#)].
- [93] S. Hong, A. Dey, K.-S. Lee, et al., *Statistics of two-point correlation and network topology for Ly α emitters at $z \approx 2.67$* , *Mon. Not. R. Astron. Soc.* **483** (Mar., 2019) 3950–3970, [[arXiv:1811.10631](#)].
- [94] F. Bian, X. Fan, L. Jiang, et al., *The LBT Boötes Field Survey. I. The Rest-frame Ultraviolet and Near-infrared Luminosity Functions and Clustering of Bright Lyman Break Galaxies at $Z \sim 3$* , *Astrophys. J.* **774** (Sept., 2013) 28, [[arXiv:1307.4835](#)].
- [95] N. Kokron, S.-F. Chen, M. White, et al., *Accurate predictions from small boxes: variance suppression via the Zel’dovich approximation*, *Journal of Cosmology and Astro-Particle Physics* **2022** (Sept., 2022) 059, [[arXiv:2205.15327](#)].
- [96] J. DeRose, S.-F. Chen, N. Kokron, and M. White, *Precision redshift-space galaxy power spectra using Zel’dovich control variates*, *Journal of Cosmology and Astro-Particle Physics* **2023** (Feb., 2023) 008, [[arXiv:2210.14239](#)].
- [97] B. Hadzhiyska, M. J. White, X. Chen, et al., *Mitigating the noise of DESI mocks using analytic control variates*, *The Open Journal of Astrophysics* **6** (Oct., 2023) 38, [[arXiv:2308.12343](#)].
- [98] S.-F. Chen, Z. Vlah, and M. White, *Consistent Modeling of Velocity Statistics and Redshift-Space Distortions in One-Loop Perturbation Theory*, *JCAP* **07** (2020) 062, [[arXiv:2005.00523](#)].
- [99] DESI Collaboration, A. G. Adame, J. Aguilar, et al., *DESI 2024 V: Full-Shape Galaxy Clustering from Galaxies and Quasars*, *arXiv e-prints* (Nov., 2024) [[arXiv:2411.12021](#)].
- [100] M. Maus, Y. Lai, H. E. Noriega, et al., *A comparison of effective field theory models of redshift space galaxy power spectra for DESI 2024 and future surveys*, *arXiv e-prints* (Apr., 2024) arXiv:2404.07272, [[arXiv:2404.07272](#)].
- [101] M. Maus, S. Chen, M. White, et al., *An analysis of parameter compression and full-modeling techniques with Velocileptors for DESI 2024 and beyond*, *arXiv e-prints* (Apr., 2024) arXiv:2404.07312, [[arXiv:2404.07312](#)].
- [102] Z. Vlah and M. White, *Exploring redshift-space distortions in large-scale structure*, *Journal of Cosmology and Astro-Particle Physics* **2019** (Mar, 2019) 007, [[arXiv:1812.02775](#)].
- [103] J. Torrado and A. Lewis, *Cobaya: code for Bayesian analysis of hierarchical physical models*, *Journal of Cosmology and Astro-Particle Physics* **2021** (May, 2021) 057, [[arXiv:2005.05290](#)].
- [104] J. Torrado and A. Lewis, “Cobaya: Bayesian analysis in cosmology.” *Astrophysics Source Code Library*, record ascl:1910.019, Oct., 2019.
- [105] V. Desjacques, D. Jeong, and F. Schmidt, *Large-scale galaxy bias*, *Phys. Rep.* **733** (Feb., 2018) 1–193, [[arXiv:1611.09787](#)].
- [106] E. Mons, V. P. Maranchery, M. S. S. Sivadas, and C. Jose, *Non-separable halo bias from high-redshift galaxy clustering*, [[arXiv:2506.07662](#)].
- [107] A. Baleato Lizancos, U. Seljak, M. Karamanis, et al., *Selecting samples of galaxies with fewer Fingers-of-God*, *arXiv e-prints* (Jan., 2025) arXiv:2501.10587, [[arXiv:2501.10587](#)].
- [108] H. K. Herrera-Alcantar, *The lyman- α forest 3d correlation function from lbg spectra: First measurement with desi and bao prospects.*, *In prep* (2025).
- [109] C. Modi, M. White, and Z. Vlah, *Modeling CMB lensing cross correlations with CLEFT*, *Journal of Cosmology and Astro-Particle Physics* **2017** (Aug., 2017) 009, [[arXiv:1706.03173](#)].

- [110] N. Galitzki, A. Ali, K. S. Arnold, et al., *The Simons Observatory: instrument overview*, in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX*, vol. 10708 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 1070804, July, 2018. [arXiv:1808.04493](#).
- [111] N. Sailer, J. DeRose, S. Ferraro, et al., *Evolution of structure growth during dark energy domination: Insights from the cross-correlation of DESI galaxies with CMB lensing and galaxy magnification*, *Phys. Rev. D* **111** (May, 2025) 103540, [[arXiv:2503.24385](#)].
- [112] R. de Belsunce, A. Krolewski, S. Chiarenza, et al., *Cosmology from Planck CMB Lensing and DESI DR1 Quasar Tomography*, *arXiv e-prints* (June, 2025) arXiv:2506.22416, [[arXiv:2506.22416](#)].
- [113] N. Sailer, J. Kim, S. Ferraro, et al., *Cosmological constraints from the cross-correlation of DESI Luminous Red Galaxies with CMB lensing from Planck PR4 and ACT DR6*, *arXiv e-prints* (July, 2024) arXiv:2407.04607, [[arXiv:2407.04607](#)].
- [114] G. S. Farren, A. Krolewski, N. MacCrann, et al., *The Atacama Cosmology Telescope: Cosmology from Cross-correlations of unWISE Galaxies and ACT DR6 CMB Lensing*, *Astrophys. J.* **966** (May, 2024) 157, [[arXiv:2309.05659](#)].
- [115] G. Piccirilli, G. Fabbian, D. Alonso, et al., *Growth history and quasar bias evolution at $z \lesssim 3$ from Quiaa*, *Journal of Cosmology and Astro-Particle Physics* **2024** (June, 2024) 012, [[arXiv:2402.05761](#)].
- [116] M. Pinon, A. de Mattia, P. McDonald, et al., *Mitigation of DESI fiber assignment incompleteness effect on two-point clustering with small angular scale truncated estimators*, *Journal of Cosmology and Astro-Particle Physics* **2025** (Jan., 2025) 131, [[arXiv:2406.04804](#)].
- [117] S. Cole and N. Kaiser, *Biased clustering in the cold dark matter cosmogony.*, *Mon. Not. R. Astron. Soc.* **237** (Apr., 1989) 1127–1146.
- [118] C. Modi, M. White, A. Slosar, and E. Castorina, *Reconstructing large-scale structure with neutral hydrogen surveys*, *JCAP* **11** (2019) 023, [[arXiv:1907.02330](#)].
- [119] B. A. Reid, H.-J. Seo, A. Leauthaud, et al., *A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies*, *Mon. Not. R. Astron. Soc.* **444** (Oct., 2014) 476–502, [[arXiv:1404.3742](#)].
- [120] D. Bianchi and W. J. Percival, *Unbiased clustering estimation in the presence of missing observations*, *Mon. Not. R. Astron. Soc.* **472** (Nov., 2017) 1106–1118, [[arXiv:1703.02070](#)].
- [121] L. Anderson, E. Aubourg, S. Bailey, et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample*, *Mon. Not. R. Astron. Soc.* **427** (Dec., 2012) 3435–3467, [[arXiv:1203.6594](#)].
- [122] E. Hawkins, S. Maddox, S. Cole, et al., *The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe*, *Mon. Not. R. Astron. Soc.* **346** (Nov., 2003) 78–96, [[astro-ph/0212375](#)].
- [123] S. Alam, J. A. Peacock, K. Kraljic, et al., *Multitracer extension of the halo model: probing quenching and conformity in eBOSS*, *Mon. Not. R. Astron. Soc.* **497** (Sept., 2020) 581–595, [[arXiv:1910.05095](#)].