The High Latitude Spectroscopic Survey on the Nancy Grace Roman Space Telescope

Yun Wang1, Zhongxu Zhai1, Anahita Alavi1, Elena Massara2,3, Alice Pisani4,5,6, Andrew Benson7, Chris Hirata8,9,10, Lado Samushia11, David H. Weinberg9,10, James Colbert1,12,13, Olivier Doré12,13, Tim Ei14, Chen Heinrich15, Shirley Ho16, Elisabeth Krause14, Nikhil Padmanabhan15, David Spergel5, and Harry I. Teplitz1

1 IPAC, California Institute of Technology, Mail Code 314-6, 1200 E. California Blvd., Pasadena, CA 91125, USA; wang@ipac.caltech.edu
2 Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave. W, Waterloo ON N2L 3G1, Canada
3 Department of Physics and Astronomy, University of Waterloo, ON N2L 3G1, Canada
4 Princeton University, Department of Astrophysical Sciences, 4 Ivy Lane, Princeton, 08540, NJ, USA
5 Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY, 10010, USA
6 The Cooper Union for the Advancement of Science and Art, 41 Cooper Square, New York, NY 10003, USA
7 Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA
8 Department of Physics, The Ohio State University, Columbus, OH 43210, USA
9 Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA
10 Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH 43210, USA
11 Department of Physics, Kansas State University, 116 Cardwell Hall, 1228 N. 17th Street, Manhattan, KS 66506, USA
12 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91101, USA
13 Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA
14 Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
15 Department of Physics, Yale University, New Haven, CT 06520, USA

Received 2021 October 5; revised 2021 December 20; accepted 2022 January 7; published 2022 March 22

Abstract

The Nancy Grace Roman Space Telescope (HLSS) over a large volume at high redshift, using the near-IR grism (1.0–1.93 μm, R = 435–865) and the 0.28 deg2 wide-field camera. We present a reference HLSS that maps 2000 deg2 and achieves an emission-line flux limit of 10−16 erg s−1 cm−2 at 6.5σ, requiring ~0.6 yr of observing time. We summarize the flowdown of the Roman science objectives to the science and technical requirements of the HLSS. We construct a mock redshift survey over the full HLSS volume by applying a semianalytic galaxy formation model to a cosmological N-body simulation and use this mock survey to create pixel-level simulations of 4 deg2 of grism spectroscopy. We find that the reference HLSS would measure ~10 million Hα galaxy redshifts that densely map large-scale structure at z = 1–2 and 2 million [O III] galaxy redshifts that sparsely map structures at z = 2–3. We expect the performance of this survey for measurements of the cosmic expansion history with baryon acoustic oscillations and the growth of large-scale structure with redshift-space distortions. We also study possible deviations from the reference design and find that a deep HLSS at f_{line} > 7 × 10^{-17} erg s^{-1} cm^{-2} over 4000 deg^2 (requiring ~1.5 yr of observing time) provides the most compelling stand-alone constraints on dark energy from Roman alone. This provides a useful reference for future optimizations. The reference survey, simulated data sets, and forecasts presented here will inform community decisions on the final scope and design of the Roman HLSS.

Unified Astronomy Thesaurus concepts: Dark energy (351); Cosmology (343)

1. Introduction

The 2010 decadal survey of astronomy and astrophysics (National Research Council 2010) identified a Wide Field Infrared Survey Telescope (WFIRST) as the highest priority for a large new space initiative, recognizing that advances in large-format near-IR (NIR) detectors enabled a mission that could make major contributions to studies of cosmic acceleration, exoplanet demographics, and a wide range of stellar and galactic astrophysics reaching from the solar neighborhood to redshifts z > 10. A key component of the proposed WFIRST observing program was a large-area slitless spectroscopic survey at high Galactic latitude, designed to map large-scale structure traced by galaxies over an enormous comoving volume at z > 1. This paper describes a “reference” version of this High Latitude Spectroscopic Survey (HLSS) based on the current capabilities of the mission. A companion paper (C. M. Hirata et al., in preparation) describes a reference version of the High Latitude Imaging Survey (HLIS). The HLSS as described here has been defined as a concrete example of the spectroscopic component of the High Latitude Wide Area Survey, in order to formulate mission science requirements and specify needed mission capabilities. The final implementation of the High Latitude Wide Area Spectroscopic Survey will be defined by NASA through a community process prior to launch.

Like other decadal missions before it, WFIRST has undergone substantial changes between the concept proposed to the decadal survey and the design adopted for construction. A crucial early change was the shift from a purpose-built 1.5 m telescope to a preexisting 2.4 m telescope, with correspondingly higher light-gathering power and angular resolution (Dressler et al. 2012). A second crucial change was the adoption of H4RG detectors, with 4x the pixel count of the H2RG detectors assumed in the original design. A third change, prompted by the greater capabilities of the 2.4 m telescope, was the addition of an on-axis coronagraphic instrument to study nearby exoplanetary systems and demonstrate technology for future missions. The
evolution of the WFIRST concept and design can be traced through the series of Science Definition Team reports (Green et al. 2012; Spergel et al. 2013, 2015).

WFIRST is now in its final design and fabrication phase (Phase C in NASA parlance), and it has been renamed the Nancy Grace Roman Space Telescope in honor of one of the pioneers of NASA space astrophysics. We will refer to it hereafter as Roman Space Telescope, or simply as Roman. The mission recently passed a critical milestone of identifying a full complement of flight-qualified H4RG detectors that meet its science requirements. While some detailed design decisions are still to be made, the capabilities of the mission hardware are now well understood. In particular, the Roman grism enables spectroscopy at \( R = 460 \lambda/\mu m \) for \( \lambda = 1–1.93 \mu m \). The Roman capabilities are summarized succinctly in the white paper of Akeson et al. (2019). As emphasized there, for applications in wide-field NIR imaging and spectroscopy Roman is many hundreds of times more powerful than Hubble Space Telescope, enabling ambitious observing programs that were previously inconceivable.

During its 5 yr prime mission, Roman will include a substantial Guest Observer program, but it is expected that the majority of observing time will be devoted to a set of large, coherent surveys as envisioned by Astro2010. The design of these surveys will be guided by quantitative goals in cosmology and exoplanet science, but the powerful public data sets they produce will support an enormous range of science investigations. Allocations of observing time and choices among alternative survey strategies will be made close to mission launch by a community-driven process. Here we present a reference design for the HLSS, one that illustrates the many considerations that go into survey strategy choices and that demonstrates the extraordinary power of an observing program that would (in this reference design) require ∼0.6 yr of observing time (not including calibration and overheads) on Roman, interleaved throughout the prime mission.

In round numbers, this reference survey would map 2000 deg\(^2\) to a typical emission-line flux limit of \( 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2} \) at 6.5σ (8.5 × 10\(^{-17}\) erg s\(^{-1}\) cm\(^{-2}\) at 5σ),\(^{16}\) enabling measurement of ∼10 million galaxy redshifts at \( z = 1–2 \) using H\(_\alpha\) emission lines and ∼2 million galaxy redshifts at \( z = 2–3 \) using [O III] emission lines. We assume that all areas observed by the HLSS will also be observed by the HLIS, providing high-resolution angular images and broadband spectral energy distributions (SEDs) for all detected galaxies. For studies of cosmic acceleration, this galaxy redshift survey would enable precise measurements of the distance–redshift relation \( D_M(z) \) and the expansion rate \( H(z) \) using the standard ruler imprinted by baryon acoustic oscillations (BAOs; Blake & Glazebrook 2003; Seo & Eisenstein 2003) and precise measurements of the growth of dark matter clustering through redshift-space distortions (RSDs; Kaiser 1987; Guzzo et al. 2008; Wang 2008). General discussion of these observational probes in the context of cosmic acceleration can be found in the relevant reviews, e.g., Frieman et al. (2008), Wang (2010), and Weinberg et al. (2013). Our forecasts in this paper draw heavily on the methodology developed by Wang et al. (2010), Wang (2012), Wang et al. (2013), Wang (2017), and Zhai et al. (2019a). As an illustration for broader applications in extragalactic astrophysics, we note that Roman is capable of carrying out a deep galaxy survey over 20 deg\(^2\) in ∼5 days, with the same depth as the influential NIR grism survey of the 3D-HST Treasury Program (5σ line flux limits of \( 5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \) for “typical objects”; Brammer et al. 2012; van Dokkum et al. 2013; Momcheva et al. 2016), but over an area more than 100 times larger!

The HLSS envisioned here has complementary properties to the large galaxy redshift surveys anticipated from the ground-based Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) and the space-based Euclid mission (Laureijs et al. 2011), both of which aim to cover larger areas (14,000 deg\(^2\) and 15,000 deg\(^2\), respectively). Compared to DESI, where the primary DESI target class transitions from luminous red galaxies to emission-line galaxies at \( z > 1 \), the HLSS will achieve a sampling of large-scale structure \( >10 \times \text{ denser at } z = 1.3 \). Compared to Euclid, the HLSS is smaller in volume but much denser in sampling, with an estimated difference of a factor of >5 in galaxy density at \( z = 1.5 \).

Although Roman could execute a shallow and wide-area survey comparable to Euclid’s in approximately 1 yr of observing time, the deeper survey proposed here is a better complement to other surveys and more effectively exploits the capabilities of Roman’s larger aperture. The higher density of galaxies with H\(_\alpha\) redshifts at \( z = 1–2 \) allows much better measurements of higher-order clustering beyond two-point correlations, and the greater depth allows [O III] redshift measurements of \( ∼2 \text{ M galaxys at } z = 2–3 \), a unique probe of structure in this redshift range. Per unit observing time, Roman is an extraordinarily efficient facility for slitless spectroscopic surveys, so it is well positioned to respond to developments in experimental cosmology between now and mission launch in the mid-2020s.

2. Survey Definition
2.1. Science Goal, Objectives, and Requirements
The Science Goal of the Roman High Latitude Survey (HLS) is to answer the two top-level questions on dark energy:

1. Is cosmic acceleration caused by a new energy component or by the breakdown of general relativity (GR) on cosmological scales?
2. If the cause is a new energy component, is its energy density constant in space and time, or has it evolved over the history of the universe?

This Science Goal flows down to the following Roman Science Objectives:

Objective 1: Roman will conduct NIR sky surveys in both imaging and spectroscopic modes, providing an imaging sensitivity for unresolved sources better than 26.5 AB magnitude.

Objective 2: Roman will determine the expansion history of the universe in order to test possible explanations of its apparent accelerating expansion, including dark energy and modification to Einstein’s gravity, using the supernova, weak lensing, and galaxy redshift survey techniques, at redshifts up to \( z = 5 \) with high-precision cross-checks between the techniques.

\[^{16}\] Using the same assumptions, the HLSS 5σ line flux limit is \( 10^{-16} > 5/6.5 \sim 8.0 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \). However, the line flux limit is sensitive to the assumed zodiacal background level, and more artifacts are expected in the grism spectra at S/N = 5 than at S/N = 6.5. Therefore, we have adopted a conservative 5σ line flux limit of \( 8.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \), to allow for a wider range of variation in the zodiacal background and noisier spectra.
Objective 3: Roman will determine the growth history of the largest structures in the universe in order to test the possible explanations of its apparent accelerating expansion, including dark energy and modification to Einstein’s gravity using weak lensing, RSDs, and galaxy cluster techniques, at redshifts up to \( z = 2 \) with high-precision cross-checks between the techniques.

We have flowed down these three Science Objectives into Level 2 Science Requirements for the Roman HLSS, as described below.

HLSS 1.—The area to be surveyed shall be \( \sim 1500 \text{ deg}^2 \) (2000 deg\(^2\) goal) after correcting for edge effects. This area will be contiguous to the extent practical, and at least 90% of the survey area must also be covered by the high-latitude imaging survey. The survey area of \( \sim 1500 \text{ deg}^2 \) is the minimum required to yield robust BAO/RSD measurements. The >90% overlap with the HLIS enables joint analysis of 90% of WL and GRS data, which could maximize the dark energy science from Roman. Imaging also provides undispersed galaxy positions, required for robust redshift determination. The statistical precision of the dark energy constraints is sensitive to the survey area and the survey depth of the HLSS; a trade study of depth versus area will need to be carried out to optimize both, in the context of combination with the weak-lensing data from the Roman HLIS and synergy with Euclid and LSST. Eiffer et al. (2021b) provide an example of such a study. A survey of one or two large, contiguous areas has smaller edge effects and better window functions than a survey composed of many smaller areas. A single contiguous area is preferable for the standard BAO analysis (Doré et al. 2018).

HLSS 2.—The comoving density of galaxies with measured redshifts shall satisfy \( n > 3 \times 10^{-4} / (h/\text{Mpc})^3 \) at \( z = 1.6 \).

This is set by requiring that the condition for shot-noise nondominance, \( nP_{\alpha 2} \sim 1 \) (where \( P_{\alpha 2} \) is the galaxy power spectrum at wavenumber \( k = 0.2 h \ \text{Mpc}^{-1} \)), is met at \( z = 1.6 \), with 20% margin. Requiring \( nP_{\alpha 2} \sim 1 \) implies \( n > 3 \times 10^{-4} / (h/\text{Mpc})^3 \) at \( z = 1.3 \), and \( n > 6.5 \times 10^{-4} / (h/\text{Mpc})^3 \) at \( z = 1.8 \). As before, the predicted \( H_0 \) ELG counts by Zhai et al. (2019b) for an \( H_0 \) line flux limit of \( f_{\text{line}} > 10^{-16} \text{ erg s}^{-1} \text{cm}^{-2} \), \( nP_{\alpha 2} \sim 0.6 \) at \( z = 1.8 \), and \( nP_{\alpha 2} \geq 2 \) at \( z = 1.3 \). We have assumed a bias for \( H_0 \) emission line galaxies (ELGs) of \( b(z) = 1 + 0.5z \). This is more conservative than the predictions from Zhai et al. (2021a) for an \( H_\alpha \) line flux limit of \( f_{\text{line}} > 10^{-16} \text{ erg s}^{-1} \text{cm}^{-2} \). We cannot require a higher galaxy number density than what nature provides, given fixed observing time and area coverage. Here we have chosen a characteristic high redshift, \( z = 1.6 \), the end of the range where DESI can look for ELGs via the [O II] doublet (Levi et al. 2019). We have allowed significant margin to allow for the large uncertainties in the measured \( H_0 \) luminosity function (LF) due to the limited availability of uniform data. The Zhai et al. (2019a) \( H_0 \) LF we have assumed is consistent with Model 3 in Pozzetti et al. (2016). The \( H_\alpha \) line flux limit of \( f_{\text{line}} > 10^{-16} \text{ erg s}^{-1} \text{cm}^{-2} \), at which this science requirement is met, is significantly deeper than the Euclid Galaxy Redshift Survey (GRS), which ensures that the HLSS is deep enough for carrying out robust modeling of systematic effects for BAO/RSD, higher-order statistics, and the combination of weak lensing and RSD as tests of GR.

HLSS 3.—The wavelength range of the HLSS will allow measurement of \( H_\alpha \) emission-line redshifts over the redshift range \( 1.1 < z < 1.9 \).

The redshift range \( 1.1 < z < 1.9 \) corresponding to the wavelength range of 1.38 to 1.9 \( \mu \text{m} \). This wavelength coverage also allows measurements of \([\text{O III}]\) emission-line redshifts over the range \( 1.8 < z < 2.8 \). However, redshift measurement strongly depends on the spectroscopic design. For grism spectroscopy, at least two emission lines are required for a robust redshift determination. Choosing the wavelength range to be \( 1–1.9 \) \( \mu \text{m} \), enables the detection of both \( H_\alpha \) and \([\text{O III}]\) over \( 1.1 < z < 1.9 \), and both \([\text{O III}]\) and \([\text{O II}]\) over \( 1.8 < z < 2.8 \), with sufficient margins. The key consideration is that a space mission should focus on what cannot be accomplished from the ground and be complementary to other space missions in wavelength coverage. Ground-based galaxy redshift surveys can reach \( z \sim 1 \) without great difficulty; thus, we should focus on \( z > 1 \). It is also critical that the Roman HLSS redshift range is complementary to that of Euclid, with its red cutoff at 1.85 \( \mu \text{m} \), or \( z < 1.8 \) for \( H_\alpha \). Euclid GRS can only reach \( z \sim 2 \); its shallow depth does not enable a high enough number density of observed \([\text{O III}]\) ELGs. The HLSS is deep enough to observe both \( H_\alpha \) (656.3nm) and \([\text{O III}]\) (500nm) ELGs, with the number density of the latter sensitive to the survey depth (the deeper the survey, the higher their number density).

HLSS 4.—Redshift measurement errors \( \sigma_z \) shall satisfy \( \sigma_z < 0.001(1 + z) \), excluding galaxies larger than \( 0.5' \) radius and outliers. The fraction of outliers with \( f_{\text{outlier}} = 0.005 \) shall be less than 10% in the sample of galaxies smaller than \( 0.5' \) radius. The incidence of outliers shall be known to a fraction of \( 2 \times 10^{-3} \) of the full sample at each redshift.

The rms redshift error of \( \sigma_z < 0.001(1 + z) \) is sufficient to prevent degradation of dark energy measurements (Wang et al. 2010). Outliers in the redshift measurement are contaminants to both BAO and RSD measurements, as they reduce the signal-to-noise ratio \( (S/N) \) of the BAO measurement and bias the RSD measurement. Adding a contaminant to the galaxy power spectrum with contamination fraction \( \alpha \) corresponds to leaking in an amount of power from the “wrong” line with amplitude \( \alpha^2 \) and diluting the power spectrum of the “real” signal by a factor of \((1 - \alpha^2)\) at the same time.

For BAO the dilution is a minor issue (it reduces \( S/N \)), but for RSD it is a critical problem because it reduces the galaxy bias \( b \) by \((1 - \alpha)\) without changing the linear RSD parameter \( \beta \). So the inferred rate of growth of structure \( b_g = b\beta \) is reduced by a factor of \((1 - \alpha)\). This leads to a stringent requirement on knowledge of \( \alpha \). If the true \( \alpha \) is 9.8% but estimated to be 10%, then the systematic error is 0.2%, which is a reasonable budget for this contribution.

Larger galaxies have larger redshift errors in grism spectroscopy. The current data from the HST grism survey WISP find that 90% of galaxies that would be observed by Roman \( H_\alpha \) line flux \( > 10^{-16} \text{ erg s}^{-1} \text{cm}^{-2} \), \( 1 < z < 1.94 \) have a size less than \( 0.5' \) (semimajor-axis continuum size; Bagley et al. 2020). Since the maximum redshift of the WISP sample is \( z \sim 1.6 \), Roman \( H_0 \) ELGs will likely have smaller sizes on average. The \([\text{O III}]\) ELGs are more compact, so this requirement is also sufficient for the redshift precision for \([\text{O III}]\) ELGs.

HLSS 5.—Relative position measurement uncertainties shall be less than \( 3''/4 \) over the entire survey area.

In order to measure galaxy clustering accurately, galaxy positions are required to be measured to better than \( \sim 0.1 \) Mpc \( h^{-1} \) (which corresponds to \( 6''/9 \), assuming that 105 Mpc \( h^{-1} \) subtends \( \sim 2'' \), e.g., the Planck 2018 cosmology at \( z = 1.5 \)). Given the pixel scale of \( 0''/11 \), this requirement is automatically
met within each field and is tied to the precision of astrometry across different fields.

**HLSS 6.**—The survey completeness shall be 60%, and the redshift purity shall be 90% (i.e., the outlier fraction is less than 10%), excluding galaxies larger than 0.754 in radius. Completeness is defined as the fraction of Hα ELGs with measured redshifts flagged as reliable, and purity is defined as the fraction of measured redshifts flagged as reliable that are actually within 2σ of the true redshifts.

A requirement on completeness and purity is needed to translate the Hα ELG number counts predicted by the Hα LF to the galaxy number density that can be used to measure BAO/RSD by Roman HLSS. The completeness of 60% and redshift purity of 90% are based on extrapolations from Euclid and are being validated using grism simulations. Since Roman has a higher spatial and spectral resolution compared to Euclid and more rolls (four vs. three) per field, we expect a higher completeness and purity for Roman. The requirement on the knowledge on the contamination fraction is set by HLSS 4.

Note that our definitions are the same as specified by the Roman Science Requirements document and are chosen to be simple and convenient: the observing efficiency of the HLSS is simply the product of its completeness and purity. They differ somewhat from the typical definition of completeness as the fraction of the target sample that has measurements classified as reliable and the redshift is actually measured correctly at a given confidence level, and purity as the fraction of the measurements classified as reliable that is actually part of the target sample and correctly measured.

These six science requirements lead to a number of implementation and operations requirements; see Doré et al. (2018). We will omit those here, except for the key requirements that impact the HLSS survey design.

**HLSS Imaging.**—Imaging observations shall be obtained of the fields in the HLSS that reach $J_{AB} = 24.0$, $H_{AB} = 23.5$, and $F184_{AB} = 23.1$ for an $r_{\text{eff}} = 0''3$ source at 10σ to achieve a reference image position, in three filters.

Imaging in at least three filters is required to build a minimal spectral template for grism spectral decontamination. It is not practical to extract spectra from grism images of galaxies that are fainter than $H_{AB} = 25$ in imaging, while the HLIS is required to have a survey depth of $H_{AB} \geq 26.5$ (Objective 1). The HLIS is much deeper than required by the HLSS. Thus, the HLSS imaging requirement is met by overlapping in footprint with the HLIS, so it only needs to be applied to any area not covered by the HLIS, and is consistent with Objective 1.

**HLSS Deep Field.**—There shall be 40 observations of two deep fields, each 11 deg$^2$ in area, sufficient to characterize the completeness and purity of the overall galaxy redshift sample. The 40 observations repeat the HLSS observing sequence of four exposures 10 times, with each deep field observation having the same exposure time as a wide-field observation of the HLSS. The dispersion directions of the 40 observations should be roughly evenly distributed between 0° and 360°.

This requirement is derived from the need to calibrate the HLSS with a spectroscopic subsample with nearly 100% purity. For meaningful comparisons, this subsample should have the same selection criteria as that of the HLSS, but with a redshift purity >99%. To measure the redshift purity to 1% requires 10,000 objects, assuming noise of $1/\sqrt{N}$ from Poisson statistics. This needs to be done in at least four categories (low $z$, high $z$, faint, luminous); hence, we need 160,000 galaxies, to allow for up to 16 subdivisions. Based on the estimated galaxy number density of $\sim 2723$ per deg$^2$ at the flux limit for the HLSS, $10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$, we need a total area for the deep fields of 160,000/2723 = 22 deg$^2$. These can be split into two subfields of 11 deg$^2$ each. Smaller subfields prevent the testing of galaxy clustering statistics in each subfield. Each deep field should be part of the HLSS footprint, so they are representative of the HLSS as a whole.

The visits to the deep field should consist of 10 sets of HLSS-like visits, matching the integration time, dither pattern, and observational time sequence of the HLSS strategy, with each set of HLSS-like visits covering the same areas of 22 deg$^2$. Assuming an efficiency (completeness multiplied by purity) of 50% and uncorrelated sets, the efficiency after 10 sets of visits is $(1-0.5)^{10} = 0.001$, leading to a 99.9% complete sample for calibrating the HLSS. Since each set of observations consists of four roll angles, the total number of deep field observations is 40. The dispersion directions of the 40 visits should be roughly evenly distributed between 0° and 360°, in order to map out possible sources of systematic errors due to inhomogeneity.

**HLSS Relative Flux Calibration.**—The relative spectrophotometric flux calibration shall be known to 2% relative accuracy (with the goal of 1%), in order to understand the effective sensitivity limit for each redshift bin for each area surveyed.

The requirement here is only on the relative spectrophotometry, which impacts the selection function of galaxies. Absolute line flux calibration will only change the overall number of objects and the $dN/dz$ but will not introduce density variations. Large-scale structure measurements require precise knowledge of the selection function of galaxies. Although the overall redshift distribution may be determined by averaging over the entire survey, fluctuations in the selection function can easily contaminate the underlying cosmological density fluctuations.

The HLSS spectroscopic sample is expected to be defined by a line flux limit of $10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$. Spatial errors in the spectrophotometric calibration will introduce artificial spatial fluctuations in the number density of galaxies, which could contaminate the cosmological signal. To control this systematic effect, we require that the noncosmological fluctuations in the mean number density (or the selection function of the survey) be <1% (sqrt variance) when averaged over spatial scales between 10 and 200 Mpc $h^{-1}$. At small scales this is ≈2 orders of magnitude smaller than the cosmological signal, while at the roughly BAO scale of 100 Mpc $h^{-1}$ this is ≈1 order of magnitude smaller than the cosmological signal. These fluctuations equal the cosmological signal at ≈400 Mpc $h^{-1}$. These physical scales correspond to $\sim 0.5$–$6^\circ$ at a redshift of 1.5.

The above requirement can be converted to a requirement on the spectrophotometric calibration accuracy, giving the luminosity function of galaxies (Pozzetti et al. 2016; Zhai et al. 2019a). At the line flux limit of Roman, this yields a requirement of $\sim 1\%$ relative spectrophotometric calibration, averaged over angular scales of $0.5^\circ$–$6^\circ$. This is a very stringent requirement. We have relaxed this requirement from 1% to 2% to add margin for mission success, assuming that we will achieve 1% relative spectrophotometric flux calibration in post-processing by projecting out problematic modes in the analysis, and using Ubercal to minimize zero-point flux fluctuations (Padmanabhan et al. 2008; Marković et al. 2017).
**HLSS Data Processing.**—The data processing system shall provide sufficient knowledge of the 3D selection function so that the artificial correlations due to inaccuracies in the 3D selection function are less than 10% of the statistical error bars on scales smaller than 2° and less than 20% on larger angular scales. This requirement is on the contribution to the total error budget by the uncertainties in the 3D selection function. The BAO scale is less than $\sim 2^\circ$6 (assuming Planck 2018 cosmology) in the redshift range for the HLSS ($1 < z < 3$). To convert the positions of observed galaxies in the large-scale structure into clustering measurements (correlation function, power spectrum, higher-order statistics), we need to know how the “average” number density of objects (in the absence of clustering) changes in the observed volume. The mean number density will vary significantly both in redshift and with angular position owing to effects of target selection, data reduction, and observing conditions. Previous surveys were able to separate the selection function into two independent parts: the radial selection function and the angular selection function. It is likely that the Roman selection function will not be separable in this way, i.e., different parts of the sky will have different radial profiles. Note, however, that most effects are either mostly radial (e.g., target selection, data reduction) or angular (e.g., imaging quality, Galactic extinction).

The knowledge about 3D selection function is usually encoded into sets of random catalogs. When computing clustering statistics, the random catalogs remove the systematic effects of varying mean number density (due to target selection, data reduction, or observing conditions). If the 3D selection function is not correct, the effects will not be completely removed and will generate spurious correlations that can bias the true cosmological signal. The angular mask of the Roman HLSS will vary pixel to pixel on the NIR detector. The full description of the angular mask may turn out to be computationally intractable. For the core science goals we require the description of the mask to be correct with an angular resolution of approximately $3\arcmin$. This corresponds to a spatial resolution of $3 h^{-1}$ Mpc at $z = 1.5$. This is driven by the fact that we need to be able to resolve the BAO peak. In principle, our requirements on the knowledge of the 3D selection function are driven by the main requirement that the spurious correlations should be no more than 10% of statistical errors between the scales of 10 and 150 $h^{-1}$ Mpc in clustering signals (either in correlation function multipoles or in power spectrum).

### 2.2. Survey Design

The main driver for the survey area of the Roman HLSS is to maximize overlap with the Roman HLIS, to use the HLIS for spectroscopic source identification, and to build the template model of contaminating background for each source using multiband photometry. The main driver for the depth of the Roman HLSS is to go significantly deeper than Euclid, for robust modeling of systematic uncertainties, and to increase the power of higher-order statistics as a dark energy probe.

The observing plan for the HLSS has several free parameters: the area versus depth trade, the location of the footprint on the sky, the tiling/dithering pattern, and the timing of the observations. All of these are subject to trade and will be reoptimized based on the science landscape prior to Roman launch, but for the purposes of writing Roman requirements and showing that they can be met in a 5 yr mission, we have designed a Reference Survey.

The tiling pattern for the Reference Survey is built hierarchically, starting from a single 0.28 deg$^2$ pointing and then building a nested series of for loops (i.e., repeating the 0.28 deg$^2$ pointing over and over again, with adjacent pointings properly tiled so that there are no pointing gaps or chip gaps after dithering), until we reach the whole survey with multiple exposures and roll angles, as shown schematically in Figure 1. The pattern itself is constrained by the layout of the Roman focal plane. Roman is a three-mirror anastigmat optical design, which can correct an annular field or portion thereof (e.g., Korsch 1977); the wide-field instrument contains 18 detectors using a portion of this annulus (Pasquale et al. 2018). These are arranged in a $6 \times 3$ pattern, but with an arc to fit all detectors within the annulus (see Figure 1(a)). The detectors each map to a $7/5 \times 7/5$ region on the sky but have gaps of $0/6–1/5$; to mostly cover these gaps, we do a diagonal chip gap dither and take a second exposure (Figure 1(b)).

The next level of the observing pattern hierarchy is to tile the sky, since Roman has only one grism (unlike Euclid) and it takes much longer to roll to a new orientation than to slew to an adjacent field. We tile first along the short axis (Figure 1(c)) to make a strip, and then we lay strips next to each other to make a 2D tiling (Figure 1(d)). The short axis was placed in the inner

![Figure 1](image.png)

**Figure 1.** A cartoon of the HLSS tiling pattern. Panels (a)-(f) show the sequence from a single pointing to successively larger regions.
for loop both for efficiency (the 0.4 short-axis slews are faster than the 0.8 long-axis slews) and because we found in observing simulations that strips with smooth edges can be tiled on a curved sky, whereas the “spaghetti strips” that form if the long axis is in the inner for loop suffer severe inefficiencies.

Finally, we first repeat the pattern at a slightly offset roll angle (Figure 1(e)) and then do two more rolls with the observatory rolled almost 180° (Figure 1(f)). Even more rolls are observed in the Roman Deep Fields (required for calibration). These rolls both reduce spectral confusion and—by providing many repeated observations of the same stars at different points on the focal plane—are expected to improve the uber-calibration solution (Marković et al. 2017). In the absence of chip gaps, this would lead to a total of eight observations (two each at four roll angles), but most points in the survey area fall into a gap at least once. Simulations of the observing plan give a coverage of 487, 1162, 1712, 1961, 2046, 2076, 2102, and 2117 deg² with coverage of ≥8, 7, 6, 5, 4, 3, 2, and 1 grism observation, respectively. The cases with <4 observations are typically due to edge effects where only some roll angles are observed, and they are unlikely to be used for the science analysis. Each observation is ~300 s in duration in the Reference HLSS.

The Reference HLSS requires 214 days (including observing overheads and deep fields for HLSS calibration, but not including additional calibration observations that are budgeted separately). It is not conducted in one single chunk but will be distributed throughout the Roman mission, constrained by (i) the field of regard (Roman observes between 54° and 126° from the Sun), (ii) the roll constraints (Roman can roll ±15° from the optimal angle on the solar panels), and (iii) the other observing programs. In particular, we do not do HLSS observations during the microlensing program (which takes six 72-day “seasons” centered in March and September when the Galactic bulge is available), and the HLIS and HLSS must be interleaved with the supernova program (which requires regular returns to the supernova field over the course of 2 yr, but at only 25% duty cycle). The roll constraint implies that observations where the observatory rolls by ±180° must be carried out with a separation of 6 months (plus an integer number of years), where the Sun is on the opposite side of the observatory. Figure 1 of Eifler et al. (2021a) shows one possible example of the scheduling.

Figure 2 shows the notional footprint of the HLSS in the sky (Galactic and equatorial views). The survey footprint for the HLIS + HLSS is placed at high Galactic latitude (to reduce dust extinction and confusion from stars) and at high ecliptic latitude (to reduce zodiacal light; this also expands the range of roll angles allowed, as we discuss below). Since the Galactic and ecliptic planes are tilted 60° to each other, this leads to two large, contiguous regions on the sky where we could place the Roman footprint: a “Southern” region and a “Northern” region. The Reference HLIS + HLSS was placed in the Southern region because the photometric redshifts for the weak-lensing analysis require both Roman NIR photometry and deep optical photometry, which is likely to come from the Vera Rubin Observatory based in the Southern Hemisphere.17 A portion of this footprint can extend up to ∼8° decl. because the ecliptic is tilted 23° to the celestial equator; we selected this part of the Southern region so that part of the Roman footprint could be observed with Northern Hemisphere telescopes. An alternative to the Reference design would be to move part of (or extend) the Roman footprint to the Northern region and cover it with optical imaging with Hyper Suprime Cam.18

Finally, we come to the choice of roll angles in the survey. While we plan to simulate several choices, the range of possibilities is restricted by observing geometry, especially at low ecliptic latitude. The “easiest” roll angles to schedule (in the sense of passing all the constraints for the greatest number of days during the mission) occur when the Sun is 90° from the target and the solar arrays are oriented directly toward the Sun. Because of the clocking angle of the wide-field instrument relative to the solar array, this corresponds to dispersion direction (vertical direction in Figure 1(a)) at an ecliptic position angle of 30° or 210°. Using spherical trigonometry, one can show that the possible dispersion directions come in two “fans” at 30° ± Δψ and 210° ± Δψ, where

$$\Delta \psi = 15° + \sin^{-1}(\tan 36° \tan |\beta|),$$

(1)

where |β| is the ecliptic latitude of the target, 15° is the maximum roll of the solar array relative to the Sun, and we recall that the allowed range of Sun-to-target angle is 90° ± 36°. Thus, we have Δψ = 15° on the ecliptic, rising to 26° at |β| = 15°, 39° at |β| = 30°, and 62° at |β| = 45°. At ecliptic latitudes |β| > 53° (i.e., Δψ > 90°), the two allowed ranges merge and all roll angles become possible.

### 3. Simulations of Astrophysical Data

To simulate the Roman galaxies, we use the semianalytic model (SAM) of galaxy formation, Galacticus (Benson 2012), to add galaxies to cosmological dark-matter-only N-body simulations. This allows us to forecast the 3D distribution.

---

17 [https://www.lsst.org/](https://www.lsst.org/)
18 [https://www naoj.org/Projects/HSC/](https://www.naoj.org/Projects/HSC/)
and properties of galaxies in the Roman HLSS. Appendix describes the implementation of Galacticus in detail.

3.1. Predicting Emission-line Galaxy Number Densities

Galaxy number density as a function of redshift is the most important ingredient in investigating the constraining power of future galaxy redshift surveys and optimizing their survey designs. With Galacticus, we can perform realistic simulations to produce synthetic galaxy catalogs and predict the characteristics of the galaxy distributions expected from Roman. The Galacticus model can output a set of galaxy properties, such as redshift, stellar mass, star formation rate, metallicity, and photometric luminosities for a set of filter transmission curves. In addition, Galacticus can be coupled with the CLOUDY photoionization code (Ferland et al. 2013) to compute the emission-line luminosity; see Appendix A.2 (see also Merson et al. 2018). The basic idea is to interpolate the tabulated libraries of emission-line luminosities using Galacticus output for each galaxy, including the ionizing photons, metallicity of the interstellar medium (ISM), hydrogen gas density, and so on. Due to the coupled nature of multiple baryonic processes and our poor prior knowledge, the model needs calibration to make reasonable predictions.

Since Roman HLSS will target emission-line galaxies at redshift $z>1$, we calibrate the Galacticus model using observations at higher redshift, instead of focusing on the local universe. In particular, we use the measurements of WISP number counts for Hα emitters at $0.7 < z < 1.5$ (Mehta et al. 2015) and the Hα luminosity function from the ground-based narrowband High-z Emission Line Survey (HiZELS; Geach et al. 2008; Sobral et al. 2009, 2013) at $z=0.4$, 0.84, 1.47, and 2.23. In addition to the parameters that describe the physics governing galaxy formation, we employ the Calzetti et al. (2000) model for dust attenuation and tune the strength of attenuation by varying the parameter $A_v$. The value of this parameter can impact the overall luminosity of the nebula emission of galaxies.

The merger trees of dark matter halos are the input to the Galacticus SAM and are constructed from the UNIT simulation (Chuang et al. 2019), which assumes a spatially flat ΛCDM model with $\Omega_m = 0.3089$, $h = H_0/100 = 0.6774$, $n_s = 0.9667$, and $\sigma_8 = 0.814$ (see Table 4 in Planck Collaboration et al. 2016). The UNIT simulation has a volume of $1\ (h^{-1}\text{Gpc})^3$ with a mass resolution of $\sim 10^9\ h^{-1}\ M_\odot$. This enables the identification of subhalos over a sufficient cosmic volume for statistics of both dark matter halos and galaxies. In order to forecast the number density of ELGs from Roman HLSS, we run the Galacticus model on top of the UNIT simulation to build a light-cone catalog, covering an area of 4 deg$^2$ in the redshift range of $0 < z < 3$. Our Galacticus model constructs the light-cone mock using the methodology presented by Kitzbichler & White (2007), which identifies dark matter halos that intersect the light cone of the observer. We utilize all 129 snapshots from the UNIT simulation back to $z = 3$ in constructing this light cone but do not perform any interpolation of galaxy positions or other properties between snapshots to the precise redshift at which the galaxy intersects the light cone. The snapshots are close enough in separation that interpolation is not required. The model also replicates the simulation box to cover sufficient volume for our mock. The method of Kitzbichler & White (2007) chooses the line of sight through the simulation to minimize any self-intersection of the light cone in the periodic replications of the simulation cube. Our light cone has a cross section of 154 by 154 comoving $h^{-1}\text{Mpc}$ at $z=3$, which is much smaller than the UNIT simulation cube, which is $1\ h^{-1}\text{Gpc}$ (comoving) on a side. As such, self-intersection and any repeated structure in the light cone will be minimal. Although our simulation is carried out for simulating the Roman HLSS, the data products can have much broader use, e.g., application to surveys such as the one planned for Euclid, or the investigation of ELG properties compared with current observations. This simulation also outputs the SED of galaxies within the wavelength range of the Roman grism, as input to the grism simulations for Roman HLSS (see Section 4).

Due to the uncertainties in the dust attenuation model, we present the galaxy mock with two different values of $A_v$. The predicted number density of Roman ELGs as a function of redshift is shown in Figure 3 for both Hα and [O III] emitters, and it is also summarized in Tables 1 and 2. The dust models have $A_v = 1.652$ and $A_v = 1.915$, respectively; they correspond to calibration of Galacticus to match the WISP number counts for Hα emitters (Mehta et al. 2015) at redshift $0.7 < z < 1.5$, and the Hα luminosity function from HiZELS (Sobral et al. 2013). We have investigated the impact of the survey depth on the number density estimates by considering six thresholds in our analysis: $2.0 \times 10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$ (the $3.5\sigma$ depth of a Euclid-like galaxy redshift survey), $1.0 \times 10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$ (the $6.5\sigma$ depth of the Reference Roman HLSS), $8.5 \times 10^{-17}\text{ erg s}^{-1}\text{ cm}^{-2}$ (the $5\sigma$ depth of the Reference Roman HLSS), $1.5 \times 10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$, $7 \times 10^{-17}\text{ erg s}^{-1}\text{ cm}^{-2}$, and $5 \times 10^{-17}\text{ erg s}^{-1}\text{ cm}^{-2}$. The last three are included to facilitate a depth versus area optimization of Roman HLSS. The predicted galaxy number counts are sensitive to the line flux limit, as expected. The Reference Roman HLSS survey provides an improvement by a factor of a few compared with Euclid in survey depth for Hα ELGs and enables observations of a sufficiently large number of [O III] ELGs at higher redshifts for BAO/RSD measurements. For Roman HLSS, we can expect thousands of Hα emitters per square degree in the redshift range $1 < z < 2$. This number density is critical for the clustering analysis to retrieve cosmological information without being dominated by shot noise.

Although we have calibrated the Galacticus SAM using data from Hα observations only, its predictions for [O III] ELGs are consistent with available observational data (see more details in Zhai et al. 2019a). The right panel of Figure 3 shows the predicted number density of [O III] ELGs as a function of redshift for different dust models and flux limits. At redshift $z < 2$, the [O III] emission line can be used along with the Hα line (the primary line) for robust redshift determination for galaxies from the Roman HLSS. At redshift $z > 2$, [O III] becomes the primary line of ELGs, and [O II] becomes the secondary line, in determining redshifts for galaxies from the Roman HLSS. At the nominal depth of $1.0 \times 10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$ ($6.5\sigma$), the Reference Roman HLSS can observe hundreds of [O III] emitters per square degree within $2 < z < 3$. Although this number density is degraded by roughly an order of magnitude compared with the Hα counterparts at lower redshifts, the [O III] ELGs are highly biased tracers of large-scale structure at $2 < z < 3$, which partially offsets the decrease in number density $n(z)$ since the shot noise of the BAO/RSD measurements depends on $n(z)b(z)P_{\text{m}}(k, z)$ (where $b(z)$ and $P_{\text{m}}(k, z)$ are the galaxy bias and matter power spectrum, respectively). Thus, the Roman [O III] ELGs will be a useful cosmological probe out to a redshift of $\sim 3$, enabling the
measurement of cosmic expansion rate and growth history of structure when the universe is only $\sim 2$ Gyr old.

### 3.2. Galaxy Mock Catalog

In order to facilitate the study of BAO/RSD and other cosmological measurements from Roman HLSS, we have produced a mock data set, a light-cone catalog of galaxies with an area of 2000 deg$^2$ using the calibrated parameters for Galacticus, and the dust model, consistent with the current baseline. This catalog is constructed using the same method as described in the previous section but pays particular attention to the emission lines. In addition, it can also be used to study different galaxy statistics, void statistics, and higher-order clustering signals to further explore the cosmological information from Roman HLSS. To validate this mock galaxy catalog for the Roman HLSS, we have performed a clustering analysis of Hz ELGs from it, using techniques that have been used in analyzing actual data. Since this 2000 deg$^2$ mock is much larger than the previous one, the simulation needs multiple replications (three to four replications in each direction transverse to the light cone) of the UNIT box to give continuous distribution over the volume. This will lead to some repetition of structures in the light cone on the largest scales. This can be improved with simulations of larger box size in the future.

For this mock data set, we adopt the dust model based on the measurements from the HiZELS Hz luminosity function and select galaxies with Hz flux higher than $1.0 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$ within $1 < z < 2$. This results in a galaxy sample with $\sim 10$ million Hz-emitting galaxies within a cosmic volume of $7$ (Gpc$/h)^3$. We then perform a likelihood analysis on this galaxy sample, measuring its galaxy clustering signal relative to a random catalog distributed within the same R.A./decl. area, but with a number density 10 times that of galaxies. The likelihood analysis compares the galaxy clustering measured from data to theoretical models in a Markov Chain Monte Carlo (MCMC) process (Lewis & Bridle 2002), with a covariance matrix for the measurements derived from thousands of approximate mocks for galaxy clustering.

We have adopted the EZmock method (Chuang et al. 2015) to make the approximate mocks, and we calibrate the parameters to have consistent clustering prediction with the fiducial model derived from the galaxy mock (the simulated data set). These mocks are constructed with a large enough cosmic volume and later truncated to have the same light-cone specifications as the mock data set. The resultant EZmocks can mimic the fiducial model in terms of spatial distribution and clustering signal. We produced 1000 EZmocks to provide an accurate estimate of the covariance matrix. We analyze the clustering measurement in Fourier space and in the range of $0.02 \ h \ Mpc^{-1} < k < 0.3 \ h \ Mpc^{-1}$, by comparing the power spectrum computed from the simulated galaxy data and theoretical models. We refer the readers to Zhai et al. (2021a) for more details of the model and data analysis.

In the MCMC likelihood analysis of the mock data set, we measure the key dark energy parameters from BAO/RSD: the scaled cosmic expansion rate $H(z)/s$ and the scaled angular diameter distance $D_A(z)/s$ (both from BAO), and the scaled linear growth of large-scale structure, $f_p(z)G(z)$ (from RSD), where $s$ is the BAO scale (the comoving sound horizon at the drag epoch), $f_p(z)$ is the linear growth rate, and $G(z)$ is the linear growth factor. We have marginalized the other parameters as nuisance parameters for this analysis. We find that the true values of the BAO/RSD parameters can be recovered within $\sim 1\sigma$ by our analysis. This verifies that the Roman galaxy mock that we have created contains the correct information on galaxy clustering, and that information can be recovered by using the data analysis method currently used in analyzing actual data from galaxy redshift surveys.

Figure 4 shows the measurement errors from the Roman HLSS mock data (assuming a Planck 2016 cosmology), compared with a compilation from current observations of large-scale structure, including 6dF (Beutler et al. 2012), SDSS MGS (Howlett et al. 2015; Ross et al. 2015), BOSS (Alam et al. 2017), eBOSS (Bautista et al. 2020; de Mattia et al. 2021; Gil-Marín et al. 2020; Hou et al. 2020; Neveux et al. 2020; Tamone et al. 2020), and BOSS-eBOSS Ly$\alpha$ forest (du Mas des Bourboux 2020). Note that the analysis methods that we have used are not yet optimized; thus, the forecast errors for the Roman HLSS in Figure 4 are very conservative as a simple statistical analysis (see Section 5.1). Future analysis should lead to higher precision, e.g., with improved modeling of nonlinear scale information, enhanced BAO signal from...
### Table 1
Number Counts of Hα-emitting Galaxies, \( dN/dz \), per Square Degree as a Function of Redshift for Different Line Flux Limits, Predicted from Galacticus Simulation

| Redshift | Dust Model with \( A_v = 1.65 \) | Dust Model with \( A_v = 1.92 \) |
|----------|----------------------------------|----------------------------------|
| From     | To                               | From                             | To                               |
| 0.5      | 0.6                              | 0.5                              | 0.6                              |
| 34,725 ± 2775 26,812 ± 2109 22,912 ± 1805 19,912 ± 1550 13,662 ± 1070 9820 ± 775 | 29,787 ± 2363 22,795 ± 1800 19,327 ± 1506 16,642 ± 1311 10,935 ± 863 7732 ± 610 |
| 0.6      | 0.7                              | 0.6                              | 0.7                              |
| 42,560 ± 2922 32,052 ± 2197 26,952 ± 1878 23,110 ± 1600 14,995 ± 1057 10,420 ± 770 | 36,100 ± 2481 26,830 ± 1868 22,317 ± 1548 18,820 ± 1308 11,657 ± 839 7820 ± 591 |
| 0.7      | 0.8                              | 0.7                              | 0.8                              |
| 34,850 ± 2025 26,302 ± 1527 22,022 ± 1285 18,567 ± 1088 11,675 ± 686 7802 ± 470 | 29,610 ± 1734 21,915 ± 1280 17,890 ± 1043 14,985 ± 879 8875 ± 541 5815 ± 354 |
| 0.8      | 0.9                              | 0.8                              | 0.9                              |
| 35,750 ± 2205 26,277 ± 1647 21,670 ± 1380 17,942 ± 1132 10,555 ± 670 6897 ± 459 | 29,992 ± 1859 21,562 ± 1374 17,205 ± 1083 14,072 ± 874 7952 ± 511 4920 ± 341 |
| 0.9      | 1.0                              | 0.9                              | 1.0                              |
| 32,662 ± 1836 23,500 ± 1335 18,935 ± 1081 15,570 ± 892 8752 ± 507 5527 ± 340 | 27,007 ± 1531 18,835 ± 1074 14,780 ± 846 11,812 ± 673 6445 ± 401 3960 ± 261 |
| 1.0      | 1.1                              | 1.0                              | 1.1                              |
| 36,517 ± 2225 25,317 ± 1575 20,057 ± 1238 16,235 ± 1013 8780 ± 549 5340 ± 337 | 29,597 ± 1828 19,932 ± 1229 15,470 ± 956 12,120 ± 759 6282 ± 395 3667 ± 242 |
| 1.1      | 1.2                              | 1.1                              | 1.2                              |
| 34,552 ± 1983 23,292 ± 1372 17,970 ± 1075 14,312 ± 847 7422 ± 444 4260 ± 266 | 27,627 ± 1595 17,837 ± 1069 13,595 ± 805 10,620 ± 622 5042 ± 304 2865 ± 182 |
| 1.2      | 1.3                              | 1.2                              | 1.3                              |
| 26,615 ± 2061 17,502 ± 1394 13,287 ± 1060 10,377 ± 840 5280 ± 433 3055 ± 262 | 20,987 ± 1646 13,202 ± 1051 9777 ± 794 7537 ± 607 3625 ± 301 1990 ± 172 |
| 1.3      | 1.4                              | 1.3                              | 1.4                              |
| 31,170 ± 2009 20,030 ± 1288 15,107 ± 949 11,615 ± 740 5552 ± 350 3025 ± 200 | 24,215 ± 1550 14,995 ± 944 10,900 ± 693 8217 ± 521 3727 ± 245 1835 ± 128 |
| 1.4      | 1.5                              | 1.4                              | 1.5                              |
| 26,380 ± 1708 16,642 ± 1078 12,372 ± 831 9270 ± 632 4267 ± 310 2247 ± 172 | 20,185 ± 1309 12,272 ± 823 8692 ± 598 6410 ± 454 2765 ± 204 1327 ± 115 |
| 1.5      | 1.6                              | 1.5                              | 1.6                              |
| 22,125 ± 1442 13,555 ± 934 9770 ± 688 7190 ± 505 2983 ± 219 1355 ± 100 | 16,742 ± 1116 9675 ± 679 6625 ± 460 4790 ± 345 1785 ± 127 800 ± 59 |
| 1.6      | 1.7                              | 1.6                              | 1.7                              |
| 19,727 ± 1144 11,752 ± 689 8432 ± 503 6075 ± 373 2342 ± 159 1050 ± 78 | 14,670 ± 869 8350 ± 497 5582 ± 346 3887 ± 254 1352 ± 94 667 ± 56 |
| 1.7      | 1.8                              | 1.7                              | 1.8                              |
| 17,817 ± 1174 9895 ± 680 6795 ± 477 4705 ± 322 1740 ± 135 832 ± 70 | 12,645 ± 848 6732 ± 472 4345 ± 298 2997 ± 212 1072 ± 82 542 ± 54 |
| 1.8      | 1.9                              | 1.8                              | 1.9                              |
| 11,732 ± 957 6157 ± 505 4092 ± 343 2792 ± 233 955 ± 83 477 ± 51 | 8095 ± 661 4047 ± 339 2562 ± 220 1640 ± 140 575 ± 56 310 ± 34 |
| 1.9      | 2.0                              | 1.9                              | 2.0                              |
| 9955 ± 547 5070 ± 285 3340 ± 199 2232 ± 139 820 ± 58 390 ± 32 | 6792 ± 388 3305 ± 196 2027 ± 127 1367 ± 96 502 ± 38 255 ± 22 |

Note. The uncertainties are estimated from the jackknife resampling method and cosmic variance. The results for both of the dust models are shown. Note that the observational efficiency is not included here.
Table 2

Same as Table 1, but for [O III]-emitting Galaxies

| Redshift (z) | From (x10^{-1}) | To (x10^{-1}) | Dust Model with \( A_V = 1.65 \) | Dust Model with \( A_V = 1.92 \) |
|-------------|-----------------|----------------|---------------------------------|---------------------------------|
|              | (5 \times 10^{-12}) | (7 \times 10^{-12}) | (8.5 \times 10^{-12}) | (1 \times 10^{-12}) | (1.5 \times 10^{-12}) | (2 \times 10^{-12}) | (5 \times 10^{-12}) | (7 \times 10^{-12}) | (8.5 \times 10^{-12}) | (1 \times 10^{-12}) | (1.5 \times 10^{-12}) | (2 \times 10^{-12}) |
| 1.0         | 1.1             | 10,437 ± 617   | 6160 ± 381          | 4447 ± 287  | 3440 ± 230  | 1610 ± 117  | 977 ± 74   | 6880 ± 426   | 3920 ± 259       | 2892 ± 197       | 2117 ± 140       | 1032 ± 76       |
| 1.1         | 1.2             | 10,212 ± 589   | 5907 ± 364          | 4225 ± 257  | 3167 ± 204  | 1550 ± 114  | 922 ± 72   | 6590 ± 400   | 3707 ± 228       | 2657 ± 170       | 1942 ± 131       | 957 ± 76        |
| 1.2         | 1.3             | 8360 ± 609     | 4797 ± 350          | 3380 ± 254  | 2560 ± 191  | 1155 ± 101  | 697 ± 67   | 5362 ± 387   | 2992 ± 222       | 2145 ± 164       | 1562 ± 132       | 727 ± 69        |
| 1.3         | 1.4             | 10,027 ± 616   | 5727 ± 344          | 4125 ± 246  | 3095 ± 189  | 1540 ± 105  | 875 ± 64   | 6442 ± 388   | 3627 ± 218       | 2610 ± 160       | 1967 ± 128       | 935 ± 67        |
| 1.4         | 1.5             | 8987 ± 578     | 5147 ± 361          | 3690 ± 275  | 2732 ± 196  | 1292 ± 109  | 752 ± 67   | 5752 ± 394   | 3215 ± 244       | 2267 ± 166       | 1677 ± 139       | 790 ± 71        |
| 1.5         | 1.6             | 7900 ± 484     | 4530 ± 288          | 3217 ± 209  | 2355 ± 163  | 1015 ± 77   | 562 ± 50   | 5155 ± 328   | 2780 ± 191       | 1857 ± 137       | 1350 ± 99        | 585 ± 52        |
| 1.6         | 1.7             | 8365 ± 495     | 4792 ± 295          | 3267 ± 209  | 2355 ± 150  | 962 ± 75    | 490 ± 45   | 5377 ± 335   | 2840 ± 177       | 1772 ± 126       | 1315 ± 93        | 505 ± 46        |
| 1.7         | 1.8             | 6957 ± 460     | 3870 ± 254          | 2690 ± 182  | 1930 ± 134  | 812 ± 64    | 455 ± 44   | 4362 ± 286   | 2340 ± 153       | 1555 ± 107       | 1110 ± 77        | 477 ± 44        |
| 1.8         | 1.9             | 5412 ± 415     | 2857 ± 236          | 1937 ± 169  | 1305 ± 122  | 480 ± 48    | 262 ± 31   | 3252 ± 266   | 1640 ± 147       | 1050 ± 97        | 682 ± 66         | 277 ± 32        |
| 1.9         | 2.0             | 4442 ± 251     | 2277 ± 141          | 1532 ± 103  | 1052 ± 68   | 445 ± 33    | 275 ± 28   | 2680 ± 161   | 1305 ± 89        | 845 ± 53         | 605 ± 40         | 282 ± 29        |
| 2.0         | 2.1             | 3055 ± 237     | 1725 ± 135          | 1185 ± 104  | 875 ± 85    | 430 ± 51    | 302 ± 37   | 1962 ± 152   | 1047 ± 99        | 715 ± 71         | 555 ± 62         | 317 ± 39        |
| 2.1         | 2.2             | 2265 ± 243     | 1215 ± 138          | 817 ± 100   | 597 ± 69    | 282 ± 39    | 190 ± 30   | 1422 ± 160   | 710 ± 84         | 490 ± 58         | 345 ± 45         | 200 ± 31        |
| 2.2         | 2.3             | 3210 ± 287     | 1642 ± 151          | 1102 ± 116  | 715 ± 79    | 307 ± 38    | 227 ± 31   | 1827 ± 175   | 937 ± 99         | 567 ± 63         | 395 ± 46         | 230 ± 31        |
| 2.3         | 2.4             | 3170 ± 230     | 1615 ± 129          | 1097 ± 93   | 820 ± 69    | 440 ± 37    | 302 ± 32   | 1845 ± 137   | 980 ± 81         | 687 ± 56         | 522 ± 41         | 310 ± 34        |
| 2.4         | 2.5             | 2722 ± 219     | 1305 ± 100          | 817 ± 71    | 520 ± 54    | 212 ± 30    | 122 ± 21   | 1547 ± 114   | 667 ± 61         | 375 ± 44         | 267 ± 33         | 122 ± 21        |
| 2.5         | 2.6             | 2360 ± 169     | 1087 ± 88           | 677 ± 65    | 477 ± 50    | 237 ± 27    | 162 ± 21   | 1297 ± 103   | 587 ± 59         | 377 ± 42         | 275 ± 32         | 175 ± 23        |
| 2.6         | 2.7             | 2090 ± 169     | 850 ± 76            | 480 ± 44    | 312 ± 33    | 117 ± 22    | 67 ± 12    | 1067 ± 93    | 395 ± 43         | 245 ± 31         | 160 ± 23         | 72 ± 12         |
| 2.7         | 2.8             | 1932 ± 149     | 795 ± 67            | 435 ± 36    | 295 ± 33    | 105 ± 14    | 60 ± 11    | 960 ± 78     | 372 ± 34         | 225 ± 28         | 140 ± 19         | 60 ± 11         |
| 2.8         | 2.9             | 1942 ± 187     | 847 ± 95            | 477 ± 60    | 337 ± 46    | 155 ± 23    | 105 ± 20   | 1045 ± 106   | 405 ± 54         | 265 ± 39         | 195 ± 29         | 110 ± 20        |
| 2.9         | 3.0             | 1337 ± 117     | 522 ± 61            | 345 ± 40    | 222 ± 29    | 100 ± 19    | 67 ± 15    | 625 ± 66     | 290 ± 33         | 167 ± 22         | 122 ± 18         | 70 ± 14         |

Note: All values are in erg s^{-1} cm^{-2}.
cosmological measurements from the analysis of large-scale structure in the universe; the cosmic distance scales from the BAO measurements (top panel), and the growth rate measurements (bottom panel). The current measurements compiled from eBOSS Collaboration et al. (2021) include observations from 6dFGS (Beutler et al. 2012), SDSS MGS (Howlett et al. 2015; Ross et al. 2015; BOSS (Alam et al. 2017), eBOSS (Bautista et al. 2020; de Mattia et al. 2021; Gil-Marín et al. 2020; Hou et al. 2020; Neveux et al. 2020; Tamone et al. 2020), and BOSS-eBOSS Lyα forest (du Mas des Bourboux 2020). The forecast for Roman Reference HLSS is shown as open symbols, based on the clustering analysis of Hα ELGs at $1 < z < 2$ and [O III] ELGs at $2 < z < 3$, respectively, assuming a Planck 2016 cosmology; these are very conservative by assuming S/N of 6.5 for the slitless spectra (see Section 5.1). Note that the prediction of constraints for the Reference Roman HLSS is from “Model A” in Zhai et al. (2021a) at $1 < z < 2$ and from the Fisher matrix results presented in the left panel of Table 3 at $2 < z < 5$.

reconstruction, and more robust estimate of the covariance matrix.

Our Roman HLSS galaxy mock catalog provides the first realistic simulation of the galaxy distribution in the Roman HLSS. This data set can be used to study the optimization of constraints on the properties of dark energy and cosmological parameters using galaxy clustering statistics and to explore the systematic effects on the cosmological measurements.

We have also created a 4 deg$^2$ galaxy mock with galaxy SEDs as input for the Roman grism simulations (see Section 4). We split the construction of SEDs into two components: continuum and emission lines. For the former, Galacticus tracks the star formation history of each galaxy on a grid of time and metallicity (optimized to provide high time resolution close to the time at which the galaxy is observed, and lower resolution farther in the past). This star formation history is convolved with simple stellar population spectra from a stellar population library. This convolution process is carried out in the final production step such that all the millions of the galaxies in the catalog have a SED generated in a timely manner. For emission lines, we assume a Gaussian line profile, with an FWHM determined by the rotation curve at the half-mass radius of the galaxy, normalized to the total line luminosity (see Section 3.1 and Zhai et al. 2019a). The final SED is the sum of the continuum and emission lines, coupled with the calibrated Calzetti et al. (2000) dust model over the entire wavelength range. Our final SED data have a wavelength range of 0.2–4 $\mu$m and redshift range of $0 < z < 10$. This fully covers the Roman HLSS requirement of 1.0–1.93 $\mu$m as the wavelength range within $0 < z < 3$. Depending on redshift, the resolution of the SED is $R \sim$ a few hundreds to 1000. In addition, the combination of this SED and any filter transmission curve can be used to compute broadband photometry. These characteristics make the catalog not only useful for Roman HLSS but also applicable to other surveys, such as the Euclid surveys (Laureijs et al. 2011) and the surveys proposed for ATLAS Probe (Wang et al. 2019).

We note that measurements from our realistic Roman galaxy mock indicate that the Roman science requirements derived in Section 2.1 are fully met by the Reference HLSS.

3.3. Linear Bias of Emission-line Galaxies

With the large-scale 2000 deg$^2$ galaxy mock for Roman HLSS, we can perform a detailed clustering analysis under different selection conditions to forecast the linear bias of both Hα and [O III] ELGs. This is the key element for a follow-up Fisher matrix analysis to predict the constraining power of Roman HLSS. The different emission lines have different wavelengths and thus correspond to different redshift ranges within the Roman grism coverage of 1–1.93 $\mu$m. We focus on Hα and [O III] lines in the current analysis. Additional emission lines are present: [O II] as the second line for [O III] ELGs, [N II] and Hβ as contaminants. These will be investigated in future work. Since the Hα emission line is detectable at $z < 2.0$, we split the measurement of linear bias into two subsamples with $z < 2.0$ and $z > 2.0$.

For galaxies with $z < 2.0$, both Hα and [O III] are observed and can be used in the redshift determination. We apply flux limits to both lines. The first limit $f_{\text{lim},1}$ is the lower limit of the stronger line (either Hα or [O III]), chosen to be [0.5, 1.0, 2.0] $\times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$. The second parameter $f_{\text{lim},2R}$ sets the lower limit of the weaker line in units of $f_{\text{lim},1}$, and we apply three values $f_{\text{lim},1} = 0.25, 0.5, 1.0$ to investigate its impact on the measurement of linear bias. For galaxies with $z > 2.0$, we only impose the flux limit $f_{\text{lim},1}$ on [O III] emission, since Galacticus does not yet give reliable predictions for [O II] emission. In Figure 5, we present the angular distributions of a subsample of our mock galaxies for three different line flux limits (for both Hα and [O III] lines): $5 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ (3D-HST 5σ depth), $10^{-16}$ erg s$^{-1}$ cm$^{-2}$ (Reference Roman HLSS 6.5σ depth), and $2 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$ (Euclid 3.5σ depth). It clearly shows the impact of the survey depth on the measured galaxy distribution that traces the underlying matter distribution. The galaxy number density is very sensitive to the line flux limit. The Reference Roman HLSS (middle panel) is deep enough to trace the underlying matter distribution in detail.

We measure the galaxy bias by comparing the galaxy correlation function with the matter correlation function using $\xi(g,m) = b^2(r)c_{\text{mean}}$. The galaxy bias estimated from this method is a function of scale. However, it is close to a constant at scales from 10 to 50 h$^{-1}$ Mpc. At smaller scales, the nonlinearity of dark matter dynamics becomes significant, while at large scales the bias deviates from a constant owing to a combination of factors, including RSD effect, sample variance, and mode coupling. Therefore, the measurement of correlation function at 10–50 h$^{-1}$ Mpc is best for providing reasonable estimates of the galaxy linear bias. We measure the galaxy bias using different random catalogs and use the mean as the results. In Figure 6, we present the measurement of bias as a function of redshift, for galaxy samples selected with different flux limits. We find that the flux limits and dust models only have mild effects on the linear bias, and the resultant measurement is
close to a linear function of redshift, \( b_{\text{lin}} = az + b \), with parameters \( a \) and \( b \) determined through a likelihood analysis. For H\( \alpha \) galaxies, the bias measurement can be approximated by \( b_{\text{lin}} = 0.9z + 0.5 \), while for [O III] galaxies the bias is close to \( b_{\text{lin}} = z + 0.5 \). These linear fits are roughly within 1\( \sigma \) of measurements using different sample selections; see Figure 6. More details of the measurement are described in Zhai et al. (2021b).

### 4. Simulation of Grism Data

We have developed a pipeline to simulate grism spectroscopic observations for the Roman HLSS. Roman grism enables spectroscopy at \( R = 460\lambda/\mu\text{m} \) for \( \lambda = 1–1.93 \, \mu\text{m} \). Using the pipeline, we have produced simulated data products that are representative of the planned HLSS data, for use by the Science Investigation Team (SIT) and the community. Figure 7 illustrates an outline of the pipeline and products.

The simulations cover an area of 4 deg\(^2\) over the redshift range of \( 0 < z < 3 \). By incorporating the survey parameters of the planned HLSS, such as detection limits, exposure times, roll angles, and dithering, the simulation products have been created to resemble as closely as possible the future observations. In addition to the grism slitless spectroscopy, we also simulate the direct imaging that is used to locate the spectra.

#### 4.1. Motivation behind the Grism Simulation

Slitless grism spectroscopic surveys, including HLSS, can suffer from incompleteness and low purity for a variety of reasons. Sources that we fail to detect because their fluxes are close to the detection limit of the survey will result in incompleteness. Furthermore, some sources may be missed owing to their flux and/or emission lines blending with those of nearby objects. In addition to completeness, the fraction of sources with correctly identified emission lines, known as purity, is also a crucial parameter in galaxy redshift surveys. We have designed the mock slitless spectroscopy observations so that they can be used to measure the completeness and purity for the Roman HLSS (see HLSS 6 science requirements in Section 2.1).
4.2. Simulation Inputs

To create a realistic simulation of the sky, the pipeline uses input galaxies and stars with a wide variety of physical properties. We use the aXeSIM software (Kümmel et al. 2009) to generate a synthetic grism spectrum for each source. aXeSIM is a simulation software package developed in support of Hubble Space Telescope slitless spectroscopy modes. It can, however, also simulate spectral images for other slitless spectroscopic instruments, including the Roman Wide Field Instrument (WFI). aXeSIM requires tables of sources with information about their coordinates, brightness, shapes, redshifts, and spectra. In addition, aXeSIM requires configuration files describing the trace and dispersion solutions for the instrument.

Here we describe the assumptions we made for the creation of the input tables of sources and more details on how we simulate the grism spectra.

1. Galaxies: We use the Roman HLSS 4 deg² galaxy mock catalog with SEDs (see Section 3.2) as the input catalog, from which we select galaxies spanning the redshift range of 0–3. The input galaxy catalog includes galaxy coordinates, sizes, brightness, and redshifts—we use these to create the galaxy object tables that later become the inputs to the grism simulation machinery.
   (a) Coordinates: The sky coordinates (R.A., decl.) in the input catalog are converted to the detector coordinates (X, Y) on each Sensor Chip Assembly (SCA).
   (b) Inclination angle: We assume random inclination angle for each galaxy.
   (c) Size: The input catalog provides galaxy sizes in terms of disk and spheroidal radius. These sizes, along with the random inclination angle, are used to calculate the galaxy morphology, in terms of semimajor (a) and semiminor (b) axes. To avoid including very thin galaxies in our sample, we limited our sources to having a major-to-minor axis ratio $a/b = 1$–4 (e.g., see Figure 11 in Odewahn et al. 1997). It is noteworthy that we had to rescale the sizes to improve agreement with observations. Using data from the WFC3 Infrared Spectroscopic Parallel survey (WISP; Atek et al. 2010; Colbert et al. 2013), we examined the size distribution of galaxies in the galaxy mock catalog. We compared the normalized distributions of the semimajor sizes of mock galaxy catalog and WISP galaxies. To make the comparison, we only used mock galaxies down to the same magnitude limit as the WISP subsample (i.e., $H$ band = 25 AB magnitude). Although the normalized size distribution of the mock catalog has the same peak as the WISP galaxies, the mock catalog contains more large galaxies. In order to make the distributions match, we need to rescale the size of large mock galaxies, semimajor axis $>2.5$ pixels (i.e., $0''275$), by a factor of 1/2. We apply the same rescaling factor to the semiminer axis sizes in the mock galaxy catalog. Incorporating these rescaled sizes, we then use two-dimensional Gaussian profiles as image templates to model the light profiles of galaxies in our final images.
   (d) Brightness: We apply a magnitude cut of $m = 28$ AB magnitude in the Roman $H$-band filter (F158). This value corresponds to the $5\sigma$ point-source detection limit for a 1 hr exposure time in the $H$-band direct imaging.
   (e) SED: In addition to the galaxy properties, we also use the SEDs (see Section 3.2 for more details) that accompany the input catalog, as the spectral template.

Figure 7. Illustration of the steps in our Roman grism observation simulations. In this flow chart, we display the input catalogs and output products with orange and pink parallelograms, respectively. We also represent various calculation processes in blue rectangles. As described in Section 4, in addition to the input galaxy mock catalog produced using Galacticus, the pipeline allows the addition of stars to the simulated images.
for each object. We provide a detailed description of how these SEDs and their various components, continuum and emission lines, are constructed in Section 3.2. We truncate the SEDs to span an observed wavelength range of \( \lambda = 10000-20000 \) Å to match the wavelength range of Roman WFI grism observations.

1. Stars: For a realistic sky scene, we also incorporate stars in our simulation. We use the model presented in Chang et al. (2010) for the number of stars per unit of apparent magnitude, \( m \), within a certain solid angle (\( \Delta \Omega \)) as follows:

\[
\phi(m) = \Delta \Omega \int \psi(M, r) n(r) r^2 dr,
\]

where \( \psi(M, r) \) is the local stellar luminosity function taken from Just et al. (2015). The stellar density profile, \( n(r) \), is based on a model presented in Juric et al. (2008) as the sum over disk and spheroidal halos. It is a good approximation to neglect the angular variation in stellar density over 4 deg\(^2\). Note that we only use a single stellar type of G0V for the stellar spectral templates for simplicity. In order to model starlight profiles, we use point-spread function (PSF) images as the image templates. Models of the PSFs are constructed using the WebbPSF Python package from STScI (Perrin et al. 2014), which transforms models of the telescope and instrument optical state into PSFs taking into account detector pixel scales, rotations, filter profiles, and point-source spectra. In order to optimize both the code runtime and the final spatial resolution simultaneously, we made the PSFs with an oversample factor of 2 inside a 364 × 364 pixel\(^2\) box.

2. Roll angle and dither: In HLSS, spectra will be obtained at different roll angles. This will help resolve possible ambiguity in the correct source to associate with pixels illuminated by overlapping spectra, and thus mitigate the incompleteness caused by spectral confusion and emission lines obscured by nearby neighbors. The simulated products assume four different roll angles: 0°, 5°, 170°, and 175°. In addition, to cover the gap between the detectors, we include a two-point dither pattern, \( \Delta X (\Delta Y) = 0 \) and \( \Delta X(\Delta Y) = 1/4 \) SCA size (where “SCA” denotes the area covered by one H4RG detector), at each roll angle.

3. Background and exposure time: For the grism observation background we use a single value of 0.57 e s\(^{-1}\) pixel\(^{-1}\), based on a zodiacal background calculation representative of the HLSS sky coverage. The direct imaging filter has a narrower bandpass, so for the direct image simulations we use a lower value of 0.25 e s\(^{-1}\) pixel\(^{-1}\) for the background. Note that these are typical values and that the sky may vary within the HLSS area. For the simulated grism and direct imaging exposures, we use 301 s and 141 s exposure times, respectively, which are the current survey parameters of the Reference HLSS.

In our current simulations, we have not reproduced the precise pattern of observations represented by the illustrative survey strategy in Figure 1. However, our simulations of four roll angles and a 1/4 SCA dither can be used to approximate the outcome of this strategy and to examine some alternatives. Full optimization that uses the grism simulations to test the impact of different strategies on survey completeness, contamination, and uniformity is a goal for future work.

### 4.3. Simulation Overview

To run the full simulation, we first divide the 4 deg\(^2\) area into 15 regions each covering, at most, one Roman field of view. The pipeline simulates one of these regions at a time, with all WFI 18 detectors in their proper configurations. The pipeline first builds a large input object table for each region that includes all the sources contained within that region. Source coordinates in this input table are in the flat plane local coordinate system of the WFI. Next, the pipeline creates 18 smaller lists associated with 18 detectors and converts the source coordinates to the local detector coordinate system (\( X, Y \)) in SCA coordinates; see above. The input list for each individual detector is big enough that it will not miss any sources even when the detector field of view is rotated.

The pipeline then runs aXeSIM to simulate the images and spectra. We note that aXeSIM has some limitations that we had to solve for these simulations. First, aXeSIM runs one detector at a time. Therefore, we parallelize the pipeline so that all 18 detectors run simultaneously. Second, aXeSIM disperses light only along the x-axis. However, Roman WFI images will be dispersed along the y-axis. To overcome this limitation, we combine a set of 90° rotations in both the input files and final products. Third, aXeSIM does not account for instrument distortion from sky coordinates to detector coordinates. In our input object tables, we add distortion to the coordinates. We then manually add distortion coefficients to the header file of the aXeSIM products.

The final simulation data products include essentially all of the effects that need to be considered in evaluating plans to extract spectra. However, a few limitations were necessary. In the final simulation products, all detector-level artifacts have already been removed. Therefore, the images do not contain cosmic rays, bad pixels, saturation, or dark currents. Additionally, the products are already flat-fielded. We note that these simulations do not include the zeroth- and second-order spectra, the wavelength-dependent PSF, or changes to sky background within the HLSS coverage. Additionally, in these simulations we had to exclude some very narrow ranges of redshifts from our final products. The WFI sensitivity curve that we used for the simulation goes slightly bluer (9000 Å) than the blue wavelength cut of our spectral templates at 10000 Å (the final SEDs from Section 3.2 covers 0.2-4 μm, but were not used in our grism simulations). As aXeSIM extrapolates the SEDs to the blue edge of the sensitivity curve, we would see unrealistic broad spectral features for galaxies with strong emission lines that are redshifted to 10000 Å. Thus, we had to exclude such redshifts from our samples. These narrow redshift ranges are 0.517 < \( z < 0.531 \), 0.993 < \( z < 0.997 \), 1.012 < \( z < 1.016 \), 1.053 < \( z < 1.057 \), and 1.676 < \( z < 1.687 \).

Our final products will be publicly released on the IPAC website.\textsuperscript{19} The full 4 deg\(^2\) simulation products are provided in four different roll angles and two different dithering patterns. Each pair of roll angle and dithering includes 18 simulated direct images and 18 simulated 2D slitless dispersed images (i.e., grism spectra). Figure 8 displays an example of our simulated direct

\textsuperscript{19}https://roman.ipac.caltech.edu/
5. Expected Science Results

5.1. Constraints on Cosmic Acceleration

In illuminating the unknown nature of cosmic acceleration, we need to measure two free functions of time: the cosmic expansion history $H(z)$, and the growth rate of large-scale structure $f_g(z)$. These can tell us whether dark energy varies with time, and whether it is an unknown energy component (e.g., a cosmological constant), or the consequence of the modification of general relativity as the theory of gravity (see, e.g., Wang 2008).

Roman will have extraordinary capabilities in probing dark energy and constraining gravity. Here we present the expected fractional errors on \{ $H(z)s$, $D_A(z)/s$, $\beta$, $f_g(z)G(z)$ \}, where $s$ denotes the BAO scale (the comoving sound horizon at the drag epoch), $D_A(z)$ is the angular diameter distance, $\beta$ is the linear RSD parameter, and $G(z)$ is the growth factor. Note that $f_g(z)G(z) \equiv f_g(z)G(z)P_0/s^5$ is a derived observable independent of galaxy bias measurements (see Wang et al. 2013), where $P_0$ is the normalization of the matter power spectrum.

We have assumed an observing efficiency of 60% for all of our forecasts.

We have used the Fisher matrix code calibrated to match analysis results from actual data (see Wang et al. 2013), and adopted the UNIT input cosmology model as the fiducial model. Our forecast assumptions are conservative: $k_{\text{min}} = 0.02 \text{ h Mpc}^{-1}$, $k_{\text{max}} = 0.25 \text{ h Mpc}^{-1}$, and 100% nonlinearity (parameterized as $k_* = 0.12 \text{ h Mpc}^{-1}$). We find that our Fisher matrix code predicts errors that are approximately a factor of 2 smaller than those found by Zhai et al. (2021a) under the same assumptions; this is not surprising since the latter analysis was illustrative using the simplest models for RSD and not yet optimized via the use of RSD models complex enough to provide high precision.

Zhai et al. (2021a) carried out the MCMC likelihood analysis for the $k$ range of 0.02–0.3 h Mpc$^{-1}$ and computed the covariance matrix using 1000 EZmocks (Chuang et al. 2019). This is a preliminary study aimed at validating the HLSS mock, and it shows that the input cosmological model can be recovered at $\sim 1\sigma$ using an analysis method that has been applied to real data (Wang 2017). The model used is not matched to the statistical precision of the HLSS mock, as evidenced by the fact that the constraints on cosmological parameters are not significantly tightened when we increase $k_{\text{max}}$ from 0.2 to 0.3 h Mpc$^{-1}$.

Figure 8. An example of simulation final products. This example displays a direct image on the left and a 2D slitless dispersed image on the right for detector number 1. The images are shown for two roll angles of 0° and 5°.

and dispersed images for detector number 1 (i.e., SCA1) at two different roll angles of 0° and 5°.
Table 3
The Fractional Errors on $H(z)/s$, $D_A(z)/s$, $f_{\text{nu}}(z)G(z)/s$, $\beta$, for the Reference Roman HLSS over 6 Months (Not Including Overheads and Calibration), at the 6.5$\sigma$ and 5$\sigma$ Line Flux Limits, Respectively

| $\alpha$ (deg$^2$) | $10^{-16}$ erg s$^{-1}$ cm$^{-2}$ (6.5$\sigma$) | 6 Months | $8.5 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ (5$\sigma$) | 6 Months |
|---------------------|---------------------------------------------|----------|---------------------------------------------|----------|
| $H(z)/s$            | $D_A(z)/s$                                  | $f_{\text{nu}}(z)G(z)/s$ | $\beta$ | $H(z)/s$ | $D_A(z)/s$ | $f_{\text{nu}}(z)G(z)/s$ | $\beta$ |
| 2000                | 0.0177                                      | 0.0148   | 0.0599                                      | 0.0329   | 0.0173 | 0.0144 | 0.0552 | 0.0314 |
| 1.0–1.2             |                                             |          |                                             |          |       |       |       |       |
| 1.2–1.4             |                                             |          |                                             |          |       |       |       |       |
| 1.4–1.6             |                                             |          |                                             |          |       |       |       |       |
| 1.6–2.0             |                                             |          |                                             |          |       |       |       |       |
| 2.0–2.4             |                                             |          |                                             |          |       |       |       |       |
| 2.4–3.0             |                                             |          |                                             |          |       |       |       |       |

Note. We have assumed an observing efficiency of 60%.

Therefore, we consider our Fisher matrix forecasts with $k_{\text{max}} = 0.25$ h Mpc$^{-1}$ and 100% nonlinearity (Wang et al. 2013) to be reasonable given the anticipated progress in modeling galaxy clustering data before Roman is launched in 2027.

Zhai et al. (2021a) assumed a line flux limit of $10^{-16}$ erg s$^{-1}$ cm$^{-2}$ at 6.5$\sigma$, which is overly conservative since observers routinely extract and utilize slitless spectra at 5$\sigma$. Note that the emission-line galaxies have a median radius of $\sim 0.025$ at $z \sim 1.5$, based on WISP data (see Figure 17 of Doré et al. 2018). In this paper, we adopt line flux limits derived for galaxies with radius $0.025$ at 1.5 $\mu$m, using the as-built throughput and wave-front errors of the Roman telescope and instrument.

The limiting sensitivity to a line emitter depends on reaching a particular S/N. The expected S/N, a property of the galaxy and the instrument, is obtained as the ratio of the signal (calculated using the true line flux) to the noise, with the “signal” defined as the number of photons received from the galaxy. At a given S/N cut, the point-source sensitivity is significantly better than for galaxies with a finite size, and there is a continuous degradation in sensitivity as the source gets larger. A detailed discussion of the S/N for Roman can be found in Hirata et al. (2012). The Reference HLSS (see Section 2.2) line flux limit at 1.5 $\mu$m, for galaxies with radius $0.025$, is $\sim 8.5 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ at 5$\sigma$ (see footnote 1 in Section 1) and $\sim 1.0 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$ at 6.5$\sigma$.

The Reference HLSS is rather minimal; it covers 2000 deg$^2$ to the H$\alpha$ line flux limit of $10^{-16}$ erg s$^{-1}$ cm$^{-2}$ at 6.5$\sigma$ ($\sim 8.5 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ at 5$\sigma$), over approximately 6 months of observing time (not including overheads and calibration). Table 3 shows the fractional errors on $\{H(z)/s$, $D_A(z)/s$, $f_{\text{nu}}(z)G(z)/s$, $\beta\}$, for the Reference HLSS over 6 months, at the 6.5$\sigma$ and 5$\sigma$ line flux limits, respectively. Note that the constraints at $1 < z < 2$ come from galaxies with H$\alpha$ lines detected at S/N $\geq 6.5$ (or 5). Requiring both H$\alpha$ and [O III] to be detected at S/N $\geq 6.5$ (or 5) is overly stringent and significantly reduces the galaxy number density in the sample (Zhai et al. 2021b).

Given the same amount of observing time, the Roman HLSS can cover a wider area at shallower depth, or a smaller area at increased depth. Table 4 shows the fractional errors on $\{H(z)/s$, $D_A(z)/s$, $f_{\text{nu}}(z)G(z)/s$, $\beta\}$, for an example 18-month deep Roman HLSS at $f_{\text{line}} > 7 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ over 4000 deg$^2$, an example 10-month medium-wide Roman HLSS with $f_{\text{line}} > 1.5 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$ over 10,000 deg$^2$, and an example 10-month shallow and very wide Roman HLSS with $f_{\text{line}} > 2 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$ over 15,000 deg$^2$ (the same survey area and line flux limit as Euclid but at 5$\sigma$ instead of 3.5$\sigma$). We have assumed an observing efficiency of 60% and 5$\sigma$ line flux limits. The line flux limits are calculated for a galaxy with half-light radius $0.025$ at 1.5 $\mu$m ($H_\alpha$ at $z = 1.3$).

Based on our study, the 18-month deep Roman HLSS at $f_{\text{line}} > 7 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ over 4000 deg$^2$ provides the most compelling stand-alone constraints on dark energy from Roman alone. This provides a useful reference for future optimization of the Roman HLSS given the landscape of results from other projects. This survey gives the tightest constraints on cosmic acceleration at $z \gtrsim 1.6$, beyond the reach of ground-based projects using galaxy tracers, e.g., DESI, with its ELG redshift reach of $z \sim 1.5$. Note that given the very high number density of this example survey, higher-order statistics will add powerful constraints on cosmic acceleration, which can be combined with the two-point statistics to make up the difference and provide competitive constraints at $z \lesssim 1.6$ as well.

Note that the estimated observing times for the example surveys in Tables 3 and 4 do not include overheads and calibration. These are meant as illustrative examples of what the Roman HLSS could look like.

5.2. Systematic Uncertainties and Mitigation Strategies

The systematic uncertainties of the Roman HLSS can be discussed in broad categories: (1) astrophysical and theoretical uncertainties—uncertainties in the modeling of astrophysical and physical processes relevant to BAO/RSD measurements; (2) observational uncertainties—systematic effects due to the slitless spectroscopy.

5.2.1. Astrophysical and Theoretical Uncertainties

The modeling and mitigation of astrophysical and theoretical uncertainties is an extremely active field of study. The main astrophysical uncertainties arise primarily from nonlinear clustering of matter on intermediate and small scales and the RSDs.

The inclusion of intermediate and small scales greatly enhances cosmological constraining power of BAO/RSD; thus, the modeling of nonlinear effects is of critical importance. A lot of progress has been made, with the reconstruction of the linear matter density field on intermediate scales being the most effective approach.

RSD is a source of uncertainties on small scales (the random peculiar velocities of galaxies), especially on scales smaller than $5$ Mpc h$^{-1}$, and the source of signal from cosmic structure growth on large scales (the flattening of 3D galaxy distribution due to the coherent bulk flows). Thus, the correct modeling of RSD is essential to meeting the Roman BAO/
Table 4

The Fractional Errors on $H(z)s, D_A(z)/s, \frac{\bar{f}_s(z) G(z)}{f_s}$, and $\beta$, for an Example Deep Roman HLSS at $f_{\text{line}} > 7 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$, an Example Medium Roman HLSS with $f_{\text{line}} > 1.5 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$, and an Example Shallow Roman HLSS with $f_{\text{line}} > 2 \times 10^{-16}$ erg s$^{-1}$ cm$^{-2}$.

| $z$      | $H(z)s$  | $D_A(z)/s$ | $\bar{f}_s(z) G(z)$ | $\beta$ | $H(z)s$  | $D_A(z)/s$ | $\bar{f}_s(z) G(z)$ | $\beta$ | $H(z)s$  | $D_A(z)/s$ | $\bar{f}_s(z) G(z)$ | $\beta$ |
|---------|----------|------------|----------------------|---------|----------|------------|----------------------|---------|----------|------------|----------------------|---------|
| 1.0–1.2 | 0.0116   | 0.0099     | 0.0309               | 0.0200  | 0.0084   | 0.0071     | 0.0271               | 0.0158  | 0.0077   | 0.0064     | 0.0259               | 0.0148  |
| 1.2–1.4 | 0.0111   | 0.0091     | 0.0326               | 0.0219  | 0.0086   | 0.0070     | 0.0290               | 0.0182  | 0.0083   | 0.0067     | 0.0279               | 0.0177  |
| 1.4–1.6 | 0.0110   | 0.0085     | 0.0346               | 0.0242  | 0.0093   | 0.0072     | 0.0312               | 0.0215  | 0.0097   | 0.0076     | 0.0308               | 0.0226  |
| 1.6–2.0 | 0.0086   | 0.0062     | 0.0351               | 0.0233  | 0.0096   | 0.0073     | 0.0338               | 0.0254  | 0.0113   | 0.0089     | 0.0355               | 0.0299  |
| 2.0–2.4 | 0.0144   | 0.0100     | 0.0496               | 0.0456  | 0.0219   | 0.0165     | 0.0673               | 0.0701  | 0.0229   | 0.0176     | 0.0697               | 0.0733  |
| 2.4–3.0 | 0.0156   | 0.0106     | 0.0592               | 0.0572  | 0.0383   | 0.0287     | 0.1289               | 0.1393  | 0.0504   | 0.0385     | 0.1682               | 0.1830  |

Note. We have assumed an observing efficiency of 60% and 5σ line flux limits. The estimated observing times do not include overheads and calibration.
RSD science goals. This is a research area less advanced than the modeling of nonlinear effects, but progress is being made continuously.

5.2.2. Observational Uncertainties

Roman spectroscopy suffers from the observational uncertainties due to its slitless spectroscopy and limited spectral coverage. The slitless nature of the spectroscopy leads to the overlap/blending of spectra from different galaxies. The limited spectral coverage of 1–1.93 μm leads to the misidentification of emission lines in a given spectrum. The spectral overlap can be mitigated via the requirement of having at least three roll angles per field. The mitigation of line misidentification is more complex and challenging to address.

As a worked example, Massara et al. (2021) studied the line confusion in spectroscopic surveys and its possible effect on the BAO analysis. [O III] emitters are the main target of Roman at z > 1.94. The [O III] feature on galaxy spectra is a doublet that could be detected by Roman as two separate lines. However, low-S/N spectra might make the detection of the second line very difficult and [O III] more prone to line confusion. Of particular interest is the possibility of misidentifying Hβ lines as [O III] features. Indeed, the Hβ line is very close in wavelength to [O III], so that the two lines cannot be distinguished by photometric measurements. Misidentifying Hβ as [O III] leads to a wrong inference of the galaxy distance, which will appear 90 Mpc h⁻¹ closer than its true position. Since the BAO peak present in the galaxy correlation function is at about 100 Mpc h⁻¹, shifting some galaxies by 90 Mpc h⁻¹ will lead to an enhancement of the correlation function around the same scale. The interloper-induced bump will convolve with the BAO peak, broadening and shifting it. Misidentifying Hβ as [O III] introduces a systematic error in the BAO analysis, which might be worrisome but seems to be detectable. Since the shift of galaxies happens only along the line of sight, it creates an anisotropic pattern in the galaxy correlation, which should be observed in the quadrupole of the correlation function. We are developing a theoretical model to capture this effect, and enable its correction.

5.3. Additional Science

In addition to the standard BAO and RSD measurements from the two-point statistics of the observed galaxy field, the Roman HLSS galaxy samples will enable additional cosmological probes of high-redshift cosmology utilizing higher-order statistics and small-scale clustering. Most of the additional methods either partially or fully rely on the simulation-based modeling. The fact that Roman samples are above z = 1, where the nonlinearities in the gravitational evolution are still mild, makes the production of the simulations necessary for their science analysis significantly cheaper. The methodology for performing these types of analysis has evolved significantly in recent decades, and we expect it to mature even more by the time Roman starts collecting data. Below we will briefly describe the current status quo for a number of key additional probes, along with rough projections on how we expect them to perform on Roman.

5.3.1. Higher-order Clustering

An obvious way of going beyond standard BAO/RSD analysis is to analyze higher-order correlation functions (Fry 1984). The three-point correlation functions have been measured and analyzed since the 1970s (Groth & Peebles 1977). More recently, the three-point functions have been used to measure the BAO signal (Slepian et al. 2017; Pearson & Samushia 2018), supplement the two-point function to tighten RSD constraints (Gil-Marin et al. 2017), constrain neutrino mass (Hahn et al. 2020), detect relativistic effects (Di Dio et al. 2019), and measure primordial non-Gaussianity (Jeong & Komatsu 2009; Karagiannis et al. 2018).

The most pressing issues for the three-point function analysis are the reduction of the data vector (Gualdi et al. 2018, 2019), accurate modeling of covariance matrices (Sugiyama et al. 2020), and the modeling of the three-point function as a function of cosmological parameters (Lazanu & Liguori 2018). On all these issues significant progress has been made in recent years, and we expect the methodology to be settled by the time Roman data are delivered for analysis. We expect higher-order clustering to significantly enhance the dark energy constraints from the Roman HLSS.

5.3.2. Void Cosmology

In the past decade cosmic voids, the underdense regions in the galaxy distribution, have emerged as a robust probe of cosmology—complementary to traditional analysis. Given its high galaxy number density coupled with a large volume, Roman holds the power to provide strong constraints on cosmology from voids (Pisani et al. 2015, 2019).

The Roman high galaxy number density provides access to the hierarchy of voids and subvoids, allowing measurements of the void—galaxy cross-correlation function (i.e., the void density profile) down to the smallest scales. The study of the void—galaxy cross-correlation function is a rich field of research and has already provided stringent constraints from current data (e.g., Hamaus et al. 2016; Hawken et al. 2017; Hamaus et al. 2020; Aubert et al. 2020; Nadathur et al. 2020), through its use of voids as standard spheres for the Alcock-Paczynski test (Alcock & Paczynski 1979; Lavaux & Wandelt 2012) and the RSD analysis around void centers. Unlocking small scales with Roman will strongly contribute to tightening such constraints. Analogously, accessing small scales permits us to constrain the void size function (giving void numbers as a function of size) in the new regime of small void radii, so far unexplored by other surveys.

voids are particularly sensitive to dark energy (e.g., Lee & Park 2009; Lavaux & Wandelt 2012; Bos et al. 2012; Pisani et al. 2015; Verza et al. 2019), modified gravity (e.g., Cai et al. 2015; Clampitt & Jain 2015; Sahlen 2019; Contarini et al. 2021), and neutrinos (e.g., Massara et al. 2015; Banerjee & Dalal 2016; Schuster et al. 2019; Kreisch et al. 2019; Sahlein 2019; Bayer et al. 2021; Kreisch et al. 2021). We have measured ~90,000 voids in the Roman HLSS galaxy mock (see Section 3.2), using the publicly available Voronoi-watershed-based void finder VIDE, and will be studying these to forecast cosmological constraints. In the redshift range z = 1–2, Roman will be a unique survey for void science to investigate cosmology and physics beyond ΛCDM.

5.3.3. Small-scale Clustering

Perturbation theory cannot be used to predict clustering patterns of galaxies on small scales. Uncertainties in how the galaxies populate their host dark matter halos
(Desjacques et al. 2018) complicate theoretical modeling in addition to nonlinear gravitational evolution. A number of recent works have tried to extract cosmological constraints from small-scale galaxy clustering despite these difficulties. Reid et al. (2014) were able to constrain the growth rate of structure with a 2.5% precision using the CMASS galaxy sample, better than the conventional large-scale RSD analysis of the same sample by a factor of two. Subsequent similar works have improved the modeling (Kwan et al. 2015; Wang 2017; Zhai et al. 2019b; Nishimichi et al. 2019).

Cosmological analysis of the Roman small-scale clustering patterns is a very promising avenue, since the high density of its tracers will allow for high-precision measurements. The modeling of Roman measurements will likely require a suite of high-resolution simulations and well-tested simulation-based modeling techniques (similar to, e.g., Heitmann et al. 2010; Lange et al. 2022).

6. Conclusion

6.1. Summary

We have presented a comprehensive summary of the extensive work on the Roman HLSS by the Roman HLS Cosmology Science Investigation Team. Our work has included the following:

1. Derivations of science requirements from the Roman dark energy science goal and science objectives.
2. Definition of a Reference HLSS meeting all of the Roman dark energy science requirements.
3. Simulation of realistic galaxy mock catalogs and SEDs for galaxies in the Reference Roman HLSS based on Galacticus, a physical model of galaxy formation.
4. Forecast of galaxy number densities and linear biases using the realistic Reference Roman HLSS galaxy mock, verifying that all Roman dark energy science requirements are met.
5. Forecast of dark energy constraints from the Reference Roman HLSS, as well several possible alternative surveys with different characteristics.
6. Pixel-level grism simulations for the Reference Roman HLSS using our realistic Roman galaxy mock and SEDs as input.
7. Study on the observational systematics for the Roman HLSS.
8. Study on the additional science from the Roman HLSS, utilizing higher-order statistics, void statistics, and small-scale clustering, to significantly enhance dark energy constraints from Roman.

In order to make this paper a useful reference to others, we have included summaries of previous work, in addition to presenting new results obtained as part of this paper. Note that the Reference HLSS is a worked example of the High Latitude Wide Area Spectroscopic Survey on Roman. The actual survey that Roman will execute will be defined in an open community process prior to launch, taking into consideration the landscape of dark energy projects and their synergies.

6.2. Public Resources for Further Work

This paper provides a comprehensive description of work on the Roman HLSS, with useful results that can help facilitate further work. A companion paper on the Roman HLIS is C. M. Hirata et al. (2021, in preparation). Eifler et al. (2021a) present multiprobe strategies of cosmology with Roman. Eifler et al. (2021b) present a study on the synergies in cosmological studies of Roman with the Rubin Observatory Legacy Survey of Space and Time. A paper considering synergies of Roman with cosmic microwave background lensing from Simons Observatory is L. Wenzl et al. (2021, in preparation).

We plan to make all of the simulated Roman data available to the public. The galaxy catalogs and SEDs will be hosted by the NASA IRSA archive at IPAC, jointly presented by the Roman Science Support Center (SSC) at IPAC and IRSA. The grism simulations will be hosted by the Roman SSC website.

The Roman Exposure Time Calculator (Hirata et al. 2012) is available on the Roman SSC website, at https://roman.ipac.caltech.edu/sims/tools/wfDepc/wfDepc.html.

This work is supported in part by NASA grant 15-WFIRST15-0008, Cosmology with the High Latitude Survey Roman Science Investigation Team. This work was done in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank Jeff Kruk for cross-checking the Roman ETC and input on the Roman grism simulations, James Rhoads for communications on Reference Roman HLIS+HLSS sky background assumptions, and Lukas Wenzl for helpful comments on a draft of the paper.

Appendix

Modeling Galaxies with Galacticus

To construct simulated galaxy surveys built on top of cosmological, dark-matter-only N-body simulations, we make use of the Galacticus SAM of galaxy formation (Benson 2012) to predict galaxy properties. We do not attempt to give a complete description of the physical motivations for the ingredients of the model, which mostly follow standards that have been established in other SAMs, but rather focus on the specifics of the implementation in Galacticus. The reader is referred to Baugh (2006) and Benson (2010) for discussions of the physical motivations for SAMs. Galacticus is designed to be highly modular and flexible. As such, the description given here refers to the specific configuration of Galacticus used to generate the simulated surveys described in this paper.

Galacticus operates on merger trees of dark matter halos (in this case extracted from an N-body simulation as discussed in Section 3.1). A galaxy is potentially formed within each branch of each merger tree and is defined by a set of properties that we describe below. Most of these properties are evolved according to a set of differential equations that are detailed in Appendix A.2. This differential evolution is sometimes interrupted by impulsive events (such as galaxy mergers) as described in Appendix A.3. Finally, some properties (such as galaxy sizes) are determined under assumptions of equilibrium as described in Appendix A.4.

The Galacticus model contains several free parameters. The values of these parameters have been chosen by manually searching the parameter space and seeking models that provide a reasonable match to a variety of observational data (e.g., the $z = 0$ stellar mass function of galaxies; Li & White 2009) as described in Benson (2012), for example.
A.1. Properties of Halos and Galaxies

In Galacticus each node of the merger tree is represented by a set of physical components, each of which has one or more properties associated with it. Together, these components, which we will detail below, describe the dark matter halo and the galaxy associated with that node of the tree.

The dark matter halo of each node is modeled with a spherical, NFW profile (Navarro et al. 1997), with virial density contrast given by the spherical collapse model (e.g., Percival 2005). The hot atmosphere of gas that fills each dark matter halo is modeled as a spherical, isothermal β-profile, with core radius equal to 0.3 times the virial radius, and with temperature equal to the virial temperature of the halo. The disk of each galaxy is modeled as a razor-thin, exponential disk in which gas and stars share the same radial scale length. A galaxy may also have a spheroid component, modeled as a Hernquist sphere (Hernquist 1990) in which gas and stars share the same radial scale length. Finally, each galaxy may contain a single supermassive Kerr black hole (Bardeen 1970) at its center.

Merger trees are initialized by populating each progenitorless node with a mass of primordial gas equal to the universal baryon fraction multiplied by the dark-matter-only simulation halo mass for halos existing prior to reionization at \( z = 10.5 \), or whose virial velocities exceed \( 35 \text{ km s}^{-1} \). For lower-velocity halos post-reionization (\( z < 10.5 \)) we assume that no gas is accreted. Masses of all other components are set to zero. Scale lengths of the NFW profiles are set using the concentration—mass—redshift relation of Gao et al. (2008), while spin parameters (Peebles 1969) are assigned using the approach of Cole et al. (2000). Specifically, a spin is selected at random from the cosmological distribution of B Cv et al. (2007). This spin is assumed to remain unchanged until a halo has doubled in mass, at which time a new spin is selected at random from the same distribution. Halo masses and scale radii are interpolated linearly in time along each branch of the tree, while halo spins are held constant between snapshots. \(^{21}\) When a halo merges into a large halo, becoming a subhalo, we assign orbital parameters by drawing at random from the cosmological distribution measured by Benson (2005), and we assign a time until merging for the subhalo using the model of Jiang et al. (2008) (boosted by a factor 2.25). After this time until merging has passed, any galaxy in the subhalo is merged with the central galaxy of the host halo as described in Appendix A.3 below.

A.2. Differential Evolution

The properties of all components in a node are evolved forward in time according to a set of differential equations. Primordial gas from the intergalactic medium is accreted into each halo at a rate equal to the mass accretion rate of the associated dark-matter-only halo scaled by the universal baryon fraction if the halo meets the reionization conditions described above. In this way, any mass growth of a halo that is not accounted for through merging of smaller halos is included as accretion. This accreted mass has a specific angular momentum equal to that of the accreted dark matter (itself determined by the halo spin parameter).

20 A “node” here corresponds to a dark matter halo identified in a dark-matter-only cosmological N-body simulation.

21 Note that this means that the halo angular momentum will evolve as the mass and energy of the halo change in time.

The rate at which gas from the diffuse halo cools and flows into the galaxy is computed using a standard cooling radius model (e.g., White & Frenk 1991) from the evolution of a cooling radius, \( r_{\text{cool}} \). The cooling radius is defined as the radius at which \( t_{\text{cool}}(r) = \tau_e^{-1} - t_0 \), where \( \tau_e \) is the dynamical time of the halo, \( t_0 \) is the age of the universe, and \( \gamma = 0.84 \) is a tunable parameter of the model. The cooling time is defined as \( t_{\text{cool}} = (d^2/2) k_B T n_i / \Lambda (T, Z) \), where \( d = 3 \) is the number of degrees of freedom for the cooling gas, \( n_i \) is the total number density of particles assuming that the gas is fully ionized, and \( Z_0 \) is the metallicity of the hot gas. The cooling function, \( \Lambda (T, Z) \), is computed under the assumption of collisional ionization equilibrium and solar abundances using the CLOUDY code. \(^{22}\) (v13.03; Ferland et al. 2013). The angular momentum of this inflowing gas is estimated using the approach of Cole et al. (2000), and the gas is assumed to settle into a rotationally supported disk.

We adopt the instantaneous recycling approximation in which we assume that stars with main-sequence lifetimes less than around 10 Gyr evolve instantaneously and return some fraction of their mass to the ISM. We adopt a recycled fraction of \( R = 0.46 \) and a yield of \( p = 0.035 \) for a Chabrier (2001) initial mass function (Benson 2010), and we further assume that metals are well mixed into their associated gas at all times. For spheroids we assume that the star formation rate, \( \phi_s \), exhibits a simple scaling with available gas mass and dynamical time, \( \phi_s = \epsilon_{\text{sf}, s} (M_g / \tau_s) (200 v_0^3 / V_s)^{1/2} \), where \( \epsilon_{\text{sf}, s} = 0.0822 \), \( \alpha_{\text{sf}, s} = 1.964 \), and \( \tau_s \) is the dynamical time of the spheroid. For disks, \( \phi_d = \int_0^\infty dR 2 \pi R^2 \Sigma_{\text{sf}, d}(R) \), where \( \Sigma_{\text{sf}, d}(R) \) is the star formation rate surface density and is computed using the expression given by Krumholz et al. (2009). We assume that stellar winds and supernova explosions drive outflows from galaxies at a rate proportional to the rate of star formation with mass-loading factor \( \beta = (V_o / V)^{\alpha_o} \), where \( V_o = 400 \) and 85 km s\(^{-1}\) for disks and spheroids, respectively, and \( \alpha_o = 2.1 \) and 1.2 for disks and spheroids, respectively. Gas outflowing from the galaxy is assumed to be gradually reincorporated into the hot halo (see Henriques et al. 2013, for example) at a rate \( M_{\text{inc}} = \alpha_{\text{inc}} M_d / \tau_e \), where \( \alpha_{\text{inc}} = 5.0 \) controls the rate of return of outflowed gas and \( \tau_e \) is the dynamical time of the halo.

As a consequence of star formation, stellar luminosity is produced at a rate \( L_{\text{sf}, i} = \phi_i \tau_i^{-1} (Z_{\text{sf}, i} / t_{\text{obs}} - t) \), where \( T(Z, t) \) is the mass-to-light ratio of a simple stellar population of metallicity \( Z \) and age \( t \), and \( t_{\text{obs}} \) is the age of the universe at which the luminosity will be observed. Note that many different luminosities—corresponding to different \( t_{\text{obs}} \) and different filters (and so to different functions \( T(Z, t) \))—are computed simultaneously. The mass-to-light ratio is computed by integrating the SEDs of simple stellar populations under the relevant filter transmission curve. These SEDs are computed using the FSPS code (v2.5; Conroy et al. 2009; Conroy & Gunn 2010) to tabulate for a grid of times and metallicities assuming a Chabrier (2001) initial mass function. When computing \( T(Z, t) \), we interpolate in this table as necessary.

We use the stability criterion originally proposed by Efstathiou et al. (1982) to judge whether galactic disks are

---

\(^{22}\) Specifically, we tabulate cooling rates on a grid of temperatures and metallicities. We generally use CLOUDY’s solar abundance ratios, with the exception that we also include a zero-metallicity case using CLOUDY’s primordial abundances). The helium mass fraction is interpolated linearly with metallicity between CLOUDY’s primordial and solar values.
unstable to the formation of a spheroidal component (Ostriker & Peebles 1973). If the disk of a galaxy is deemed to be gravitationally unstable, we allow mass, metals, and angular momentum to be transferred from the disk component to the spheroid component on a timescale comparable to the dynamical timescale of the disk.

For satellite halos, the hot halo can be stripped owing to ram pressure from the hot halo of the host. The ram pressure radius is found following Font et al. (2008), with mass outside this radius being removed on a timescale based on the ram pressure acceleration (Roediger & Brüggen 2007).

The supermassive black hole is allowed to accrete from both the hot gas halo and the ISM of the spheroid component at rates proportional to the Bondi rate, but limited to the Eddington rate. Bondi accretion rates are boosted by factors of 6.3 and 4.7 for accretion from the hot halo and the ISM, respectively. For accretion from the hot halo we assume that accretion can only occur from the fraction of the hot halo mass actually in the hot mode. We do not explicitly model whether halos are undergoing hot- or cold-mode accretion, and so instead we impose a simple transition from cold-mode to hot-mode behavior at the point where a halo (were it in the hot mode) is able to cool out to the virial radius (Benson & Bower 2010).

As discussed in detail by Begelman (2014), accretion flows with accretion rates close to the Eddington limit will be radiatively inefficient as they struggle to radiate the energy they release, while flows with accretion rates that are much smaller than Eddington \( M_{\text{acc}} < \alpha^2 M_{\text{Edd}} \), where \( \alpha \approx 0.1 \) is the usual parameter controlling the rate of angular momentum transport in a Shakura & Sunyaev (1973) accretion disk) are also radiatively inefficient, as radiative processes are too inefficient at the associated low densities to radiate energy at the rate it is being liberated. Therefore, accretion disk structure is assumed to be a radiatively efficient, geometrically thin, Shakura & Sunyaev (1973) accretion disk if the accretion rate is between \( 0.01 M_{\text{Edd}} \) and \( 0.3 M_{\text{Edd}} \), and an advection-dominated accretion flow (ADAF) otherwise (Begelman 2014). For thin disks and high accretion rate ADAFs the radiative efficiency is given by \( \epsilon_{\text{rad}} = 1 - E_{\text{ISCO}} \), where \( E_{\text{ISCO}} \) is the specific energy of the innermost stable circular orbit (in dimensionless units) for the given black hole spin. For low accretion rate ADAFs the radiative efficiency is matched to that of the thin-disk solution at the transition point \( (0.01 M_{\text{Edd}}) \) and is decreased linearly with accretion rate below that. For the jet efficiency in thin accretion disks we use the results of Meier (2001), interpolating between their solutions for Schwarzschild black holes and rapidly rotating Kerr black holes. For the case of ADAF accretion flows we use the jet efficiency computed by Benson & Babul (2009).

We model quasar-mode feedback by allowing the black hole to drive a wind from the spheroid component following the model of Ostriker et al. (2010) and the assumption that energy deposition into the spheroid results in a wind carrying a mass flux of \( \dot{M}_{\text{wind}} = c_s L_{\text{wind}} / V_z^2 \), where \( c_s = 0.01 \), when the density of the ISM in the spheroid exceeds the critical value defined by Ostriker et al. (2010). We model radio-mode feedback as jet power deposition into the hot halo. We assume that energy is deposited by the jet into the hot halo at a rate \( P_{\text{jet}} = \epsilon_{\text{jet}} \dot{M}_{\text{acc}} \), and that this causes a reduction in the mass inflow rate of \( \dot{M}_{\text{in}} / V_z^2 \). Inflow rates of metals and angular momentum are reduced proportionally. If the inflow rate is reduced to zero, any remaining jet power is used to push material out of the hot halo at a rate \( \dot{M}_{\text{rdo}} = P_{\text{rdo}} / V_z^2 - \dot{M}_{\text{in}}, \) where \( \dot{M}_{\text{in}} \) is the rate of inflow in the absence of any radio-mode feedback.

The resulting network of differential equations described above is solved numerically using a backward differentiation formula integrator, with adaptive time stepping chosen to enforce a fractional precision of 1% in all evolved quantities.\(^{23}\)

We evolve all properties of all components in a node simultaneously, but we evolve only one node at a time. The order in which nodes are evolved is determined by the causal connections between them (i.e., progenitor halos must be evolved before their descendants, while satellites must be evolved in lock step with their centrals since their properties are mutually interdependent).

The differential equation networks include some cross-terms between nodes (e.g., due to ram pressure stripping). As we evolve only one node at a time, these cases must be treated separately. Our approach is to accumulate any mass, metals, or angular momentum transferred from a halo being evolved to another halo over some time step, \( \Delta t \). After that time step, differential evolution is paused, the accumulated quantities are transferred to the receiving node, and differential evolution resumes (possibly in another node, possibly the node which just received the accumulated properties, according to the rules described above). The size of the time step, \( \Delta t \), is chosen to be sufficiently small to give converged results.

### A.3. Impulsive Evolution

In addition to the differential evolution described above, we interrupt this evolution when two galaxies merge. Galaxy properties are changed instantaneously during this event, after which differential evolution resumes. The effects of galaxy mergers are described below. Mergers between galaxies are classified as major if \( M_{\text{sat}} > f_{\text{major}} M_{\text{cen}} \), where \( M_{\text{sat}} \) and \( M_{\text{cen}} \) refer to the total galactic baryonic mass of the satellite and central galaxies, respectively, and \( f_{\text{major}} = 0.57 \). In a major merger, all mass in the central galaxy galactic disk is transferred to the spheroid, while in a minor merger the central galaxy galactic disk remains in place. In either case, the mass of the satellite galaxy is transferred to the central galaxy spheroid, after which the satellite galaxy no longer exists as a distinct entity.

Any remaining hot halo and outflowed gas associated with the satellite galaxy halo is transferred to the central galaxy halo, with the satellite gas assumed to have a specific angular momentum relative to that of the central galaxy halo (which is proportional to the virial specific angular momentum of that halo). The supermassive black holes in the two merging galaxies are assumed to also merge instantaneously. The mass of the final black hole is taken to be the sum of the masses of the two initial black holes (i.e., we ignore any loss of mass into gravitational radiation, which is typically a few percent; Rezzolla et al. 2008). The spin of the final black hole is computed using the fitting function given by Rezzolla et al. (2008) assuming that the initial spins of the two black holes and their orbital angular momentum are all parallel (i.e., \( \cos \alpha = \cos \beta = \cos \gamma = 1 \) in the notation of Rezzolla et al. 2008).

\(^{23}\) We set small, absolute precisions for all evolved quantities to prevent the differential evolution requiring arbitrarily small time steps in regimes where some parameter values become extremely small.
A.4. Equilibrium-determined Properties

Galaxy sizes are determined on the assumption that both disk and spheroid are in equilibrium in the gravitational potential given their angular momenta, following the general approach of Cole et al. (2000). The radial sizes of the disk and spheroid components are determined from their angular momenta (actually the pseudo-angular momentum in the case of the spheroid, defined below) and the assumption of rotational support in the gravitational potential created by the combined baryonic and dark matter mass distributions in the halo, with the dark matter profile being allowed to respond to the (adiabatic) growth of the galaxy. Sizes are assumed to adjust instantaneously to any changes in angular momentum or gravitational potential and so are always in equilibrium. While model spheroids are nominally pressure support, we solve for their sizes by defining a pseudo-angular momentum, which is the angular momentum that the spheroid would have if it were rotation supported at the same size. The model of Cole et al. (2000) accounts for adiabatic contraction of the dark matter halo. We also include this effect but adopt the updated model of Gnedin et al. (2004). During galaxy mergers, the pseudo-angular momentum of any newly formed spheroid is computed using the method proposed by Cole et al. (2000).

ORCID iDs

Yun Wang @ https://orcid.org/0000-0002-4749-2984
Zhongxu Zhai @ https://orcid.org/0000-0001-7984-5476
Anahita Alavi @ https://orcid.org/0000-0002-8630-6435
Alice Pisani @ https://orcid.org/0000-0002-6146-4437
Andrew Benson @ https://orcid.org/0000-0001-5501-6008
Lado Samushia @ https://orcid.org/0000-0002-1609-5687
David H. Weinberg @ https://orcid.org/0000-0001-7775-7261
James Colbert @ https://orcid.org/0000-0001-6482-3020
Olivier Doré @ https://orcid.org/0000-0001-7432-2932
David Spergel @ https://orcid.org/0000-0002-5151-0006
Harry I. Teplitz @ https://orcid.org/0000-0002-7064-5424

References

Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:1902.05569
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Atek, H., Malkan, M., McCarthy, P., et al. 2010, ApJ, 723, 104
Andrew Benson
Olivier Doré
David Spergel
Zhongxu Zhai
Harry I. Teplitz

Chuang, C.-H., Kitaura, F.-S., Prada, F., Zhao, C., & Yepes, G. 2015, MNRAS, 446, 2621
Chuang, C.-H., Yepes, G.; Kitaura, F.-S., et al. 2019, MNRAS, 487, 48
Clampitt, J., & Jain, B. 2015, MNRAS, 454, 3357
Colbert, J., Teplitz, H., Atek, H., et al. 2013, ApJ, 779, 34
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Conroy, C., & Gunn, J. E. 2010, Astrophysics Source Code Library, ascl:1010.043
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Contini, S., Marulli, F., Moscardini, L., et al. 2021, MNRAS, 504, 5021
Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2021, MNRAS, 501, 5616
Desjacques, V., Jeong, D., & Schmidt, F. 2018, PhR, 733, 1
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Di Dio, E., Durrer, R., Maartens, R., Montanari, F., & Umeh, O. 2019, JCAP, 2019, 053
Doré, O., Hirata, C. M., Wang, Y., et al. 2018, arXiv:1804.03628
Dressler, A., Spergel, D., Mountain, M., et al. 2012, arXiv:1210.7809
Yun Wang
Zhongxu Zhai
Harry I. Teplitz

Wang et al.
