Cosmology
with the WFIRST
High Latitude Survey

JPL Caltech ipac OHIO STATE

CORNELL UNIVERSITY · Carnegie Mellon University · Penn UNIVERSITY of PENNSYLVANIA

PRINCETON UNIVERSITY · Yale · KANSAS STATE UNIVERSITY

A Proposal for a WFIRST Science Investigation Team:
Galaxy Redshift Survey (Topic A), and
Weak Lensing and Cluster Growth (Topic C)
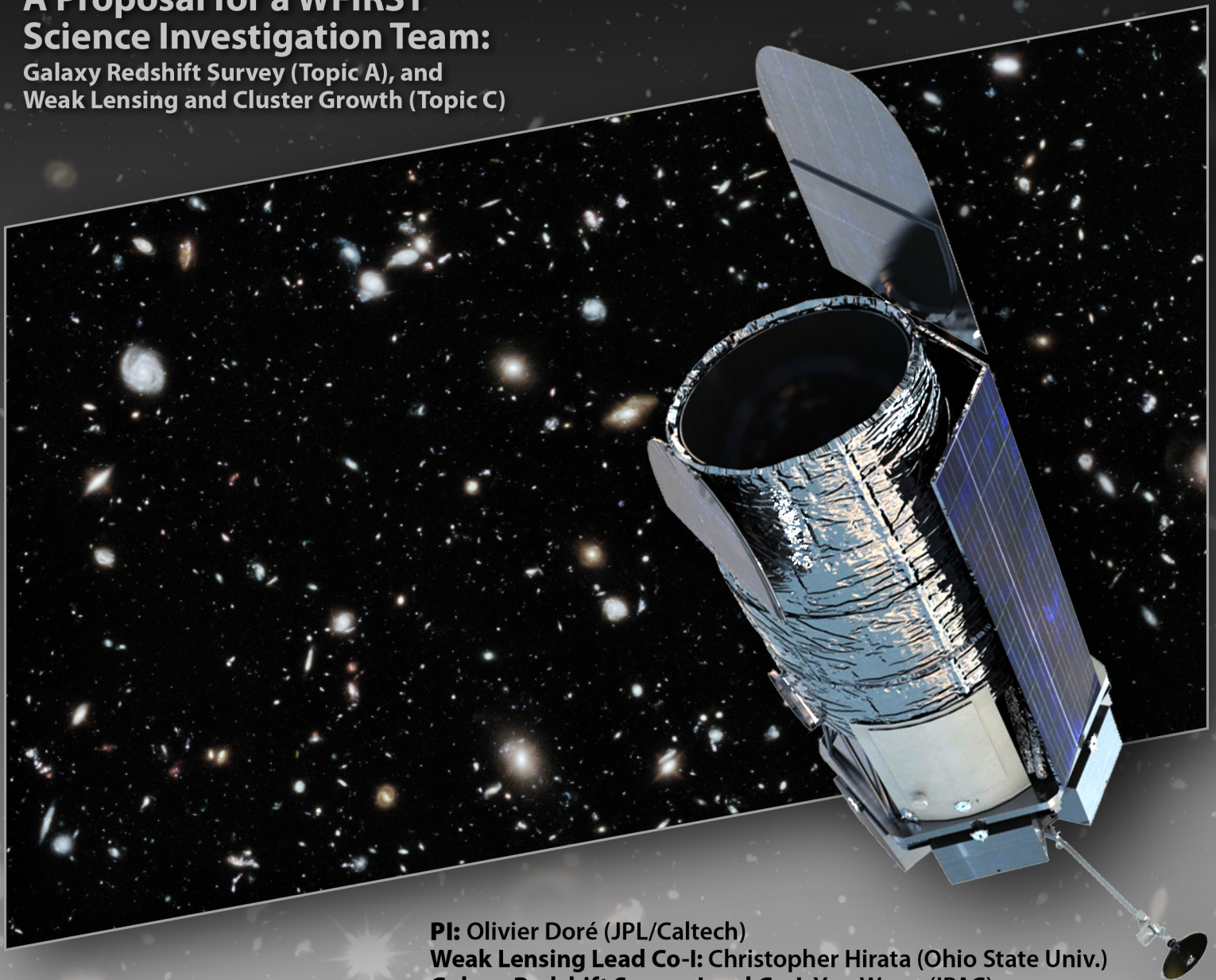
PI: Olivier Doré (JPL/Caltech)
Weak Lensing Lead Co-I: Christopher Hirata (Ohio State Univ.)
Galaxy Redshift Survey Lead Co-I: Yun Wang (IPAC)
Cluster Growth Sub-lead Co-I: David Weinberg (Ohio State Univ.)

Co-Is: Rachel Bean (Cornell), Peter Capak (IPAC), Tim Eifler (JPL), Shirley Ho (Carnegie-Mellon Univ.), Bhuvnesh Jain (Univ. of Penn.), Mike Jarvis (Univ. of Penn.), Alina Kiessling (JPL), Robert Lupton (Princeton), Rachel Mandelbaum (Carnegie-Mellon Univ.), Nikhil Padmanabhan (Yale), Lado Samushia (Kansas State Univ.), David Spergel (Princeton), Harry Teplitz (IPAC)

Collaborators: Andrew Benson (Carnegie Obs.), Katrin Heitmann (ANL), George Helou (IPAC), Michael Hudson (Univ. of Waterloo), Elisabeth Krause (Stanford), Hironao Miyatake (JPL), Eduardo Rozo (U. of Arizona), Mike Seiffert (JPL), Chaz Shapiro (JPL), Kendrick Smith (Perimeter Insti.), Masahiro Takada (Univ. of Tokyo), Anja von der Linden (Stony Brook Univ.), Naoki Yoshida (Univ. of Tokyo)

Table of Contents

Cosmology with the WFIRST High Latitude Survey

1 Executive Summary [Oli: Olivier, 2 pages]

[**Oli**: To be updated]

Cosmic acceleration is the most surprising cosmological discovery in many decades. Even the least exotic explanation of this phenomenon requires an energetically dominant component of the universe with properties never previously seen in nature, pervading otherwise empty space, with an energy density that is many orders of magnitude higher than naive expectations. Testing and distinguishing among possible explanations requires cosmological measurements of extremely high precision that probe the full history of cosmic expansion and structure growth. This program is one of the defining objectives of the Wide-Field Infrared Survey Telescope (WFIRST), as set forth in the *New Worlds, New Horizons* report (NWNH) [**?**]. The WFIRST-AFTA mission, as described in the Science Definition Team (SDT) reports [**? ?** , hereafter SDT13 and SDT15 respectively], has the ability to improve these measurements by $1-2$ orders of magnitude compared to the current state of the art, while simultaneously extending their redshift grasp, greatly improving control of systematic effects, and taking a unified approach to multiple probes that provide complementary physical information and cross-checks of cosmological results.

We have assembled a team with the expertise and commitment needed to address the stringent challenges of the WFIRST dark energy (DE) program through the Project's formulation phase. After careful consideration, we have elected to address investigations A (Galaxy Redshift Survey, GRS) and C (Weak Lensing (WL) and Cluster Growth (CL)) of the WFIRST Science Investigation Team (SIT) NASA Research Announcement (NRA) with a unified team, because the two investigations are tightly linked at both the technical level and the theoretical modeling level. The imaging and spectroscopic elements of the High Latitude Survey (HLS) will be realized as an integrated observing program, and they jointly impose requirements on instrument and telescope performance, operations, and data transfer. The methods for simulating and interpreting weak lensing and galaxy clustering observations largely overlap, and many members of our team have expertise in both areas. The WFIRST supernova cosmology program (investigation B) is more distinct in its methods and requirements, so it is feasible to integrate the supernova and HLS investigations at the level of the Formulation Science Working Group (FSWG).

The team PI, Olivier Doré, is a cosmologist with broad expertise in cosmic microwave background and large scale structure (LSS) studies. He brings extensive experience with complex data analysis (e.g., the Wilkinson Microwave Anisotropy Probe (WMAP), Planck) and mission design (e.g., Joint Dark Energy Mission (JDEM) Destiny and the SMall EXplorer (SMEX) concept SPHEREx currently under Phase A study, for which he is the Project Scientist). Yun Wang and Chris Hirata will serve as Lead Co-Investigators for topics A and C, respectively and David Weinberg will serve as Lead for sub-topic "Cluster growth" within topic C. Many members of our team have been involved with the design and requirements of a dark energy space mission for a decade or more, including the Co-Chair (Spergel) and four additional members (Hirata, Hudson, Wang, Weinberg) of the 2013-2015 WFIRST-AFTA SDTs. Our team includes authors of the two most comprehensive reviews of observational methods for probing dark energy [**? ?**] and the Chair and Vice-Chair (Spergel, Weinberg) of the Astro2010 Science Frontier Panel on Cosmology and Fundamental Physics, whose report played a central role in the NWNH recommendation of WFIRST as the highest priority large space-based program. Our team of Co-Is includes world leading experts on image processing and weak lensing (Eifler, Jain, Jarvis, Kiessling, Lupton, Hirata, Mandelbaum), on design and analysis of galaxy redshift surveys (Ho, Padmanabhan, Samushia, Wang,

Weinberg), on space-based slitless spectroscopy analogous to that planned for WFIRST (Teplitz), on photometric calibration (Padmanabhan), on photometric redshifts (Capak) from large imaging surveys, and on cosmological forecasting and parameter estimation from combinations of cosmic microwave background (CMB), WL, and LSS data (Bean, Doré, Eifler, Hirata, Ho, Jain, Mandelbaum, Samushia, Spergel, Wang, Weinberg).

This team of Co-Is brings close connections to most of the major current or planned cosmological experiments that will provide the context for the WFIRST dark energy program. This includes the WMAP and Planck CMB missions, the Sloan Digital Sky Survey (SDSS), the Baryon Oscillation Spectroscopic Survey (BOSS), the Dark Energy Survey (DES), the Subaru Hyper Suprime-Cam (HSC) and Prime Focus Spectrograph (PFS) projects, the Dark Energy Spectroscopic Instrument (DESI), the Euclid mission and the Large Synoptic Survey Telescope (LSST) Dark Energy Science Collaboration (DESC). Our team of U.S. and international collaborators brings extensive expertise in detector characterization, cosmological simulations, detailed simulations of observational data sets, and the theoretical modeling and cosmological interpretation of weak lensing and galaxy clustering data. Notably, members of our team are responsible for nearly all of the tables and figures in §§ 2.2.3-2.2.5 of the SDT15 report, describing the HLS dark energy program. We therefore have an unparalleled understanding of the current design of WFIRST-AFTA and of the challenges ahead in achieving its science goals.

We have structured our planning and our proposal around the series of deliverables described in §2. Because development of requirements is at the core of our proposed investigation, we present some broad aspects of our strategy in §?? before turning to a more detailed discussion of the WL and GRS program elements in §§3 and 4. We address questions of survey operations and optimization in §6 and our plans for broad community engagement in §7. We conclude with our management plan in §??, re-emphasizing the value of a unified approach to the HLS dark energy science program.

2   Proposed Deliverables [Oli: Olivier, 4 pages]

[**Oli**: In this section, we summarize the deliverables we promised and what we actually accomplished]

In §§??,3,4, we describe our work plan and explain how we will iteratively flow down science objectives to the measurements to be conducted, develop observational strategies, simulate synthetic astronomical 'truth' data and the observational data output (including calibration), develop a methodology for validating dark energy constraints, and define scientific performance requirements and a complete plan for the science investigation. Our work plan maps the six SIT tasks, identified in parentheses (numbered *T 1-6* as in §3.1 of the WFIRST SIT call), to the deliverables below. We also explicitly reference the deliverables (D1-12) in the section titles of our proposal.

- *(D1) Full requirements flow-down* from the high-level science goals of the HLS galaxy clustering and weak lensing survey to detailed performance of the telescope, wide field instrument, software, operations, and data transfer. We will evaluate the mission design, as prepared for the Critical Design Review (CDR), against these requirements (*T1, T3, T4*).

- *(D2) Forecasts of the cosmological performance of the HLS Imaging and Spectroscopy data sets*, including expected constraints on dark energy, modified gravity, neutrino masses, and inflation, from analyses that include the measurement of the location of the Baryon Acoustic Oscillations (BAO), Redshift-Space Distortions (RSD), galaxy power spectrum and higher order statistics, cosmic shear, galaxy-galaxy lensing, and cluster demographics. These forecasts will incorporate realistic assessments of observational systematics and theoretical modeling systematics, and they will examine the expected constraints from different probes individually, in concert with each other, and in concert with expected constraints from the WFIRST supernova program,

CMB experiments, and other cosmological surveys such as DESI, LSST, and Euclid. We will use our forecasting tools to investigate trades, e.g., the impact of survey or instrument design choices (area, depth, pixel size, spectral resolution, etc.) on cosmological performance. (*T1, T2*)

- **(D3) Simulated imaging and spectroscopic data sets** for testing pipeline performance and evaluating systematic biases — e.g., from confusion, noise, and incompleteness in images and spectra, or errors in Point Spread Function (PSF) determination or shape measurement. These data sets will be created with varying levels of complexity in the source catalogs and instrumental effects, to allow isolation of individual contributions to statistical and systematic uncertainties. Some of these artificial data sets will be made publicly available, and some will take the form of data challenges, where the underlying parameters are initially known only to the creators of the data set, in the spirit of the Shear Testing Program (STEP) and Gravitational Lensing Accuracy Test (GREAT) weak lensing data challenges [**? ? ? ? ?**]. (*T3, T5*)

- **(D4) Proto-type imaging and spectroscopic pipelines**, including weak lensing shape measurement and galaxy redshift measurement, tested against the above artificial data sets. These proto-type pipelines will provide building blocks for development of full pipelines during the implementation phase, and they will allow us to sharpen definitions of software requirements and to identify challenges to and strategies for meeting these requirements. (*T3, T5*)

- **(D5) Calibration strategies** for photometry, shape measurement, spectroscopy, and redshift completeness. Evaluation of the expected performance of these strategies against the science requirements. (*T4*)

- **(D6) A strategy for the determination and calibration of photometric redshifts** using WFIRST data and anticipated external data (e.g., LSST optical photometry), and defining ground-based data that are needed to implement this strategy (e.g., spectroscopic training sets, large redshift surveys for calibration via cross-correlation). Evaluation of the impact of remaining photometric redshift uncertainties on statistical and systematic errors in weak lensing and clustering analyses. Definition of requirements for WFIRST photometric redshifts informed by this strategy and evaluation. (*T4*)

- **(D7) A detailed operations concept for the HLS Imaging and Spectroscopy program,** extending the work presented in SDT13 and SDT15. (*T2, T6*)

- **(D8) Development of methods for modeling and interpreting the cosmological measurements anticipated from WFIRST**. Determination of the effects of non-linear gravitational clustering, realistically complex relations between the galaxy and dark matter distributions, and the influence of the baryon component on matter clustering. The study of techniques to remove systematic biases, e.g., by marginalization over nuisance parameters. Utilization of cosmic shear, galaxy-galaxy lensing, cluster mass functions and cluster weak lensing, BAO, RSD, the galaxy power spectrum, and higher order statistics for galaxy clustering, weak lensing, and various combinations. Identification of areas where further improvements of theoretical modeling would significantly enhance the cosmological return from WFIRST. (*T3*)

- **(D9) Simulated light-cone observations** based on cosmological simulations for guiding this methodology development and testing its performance. Most of these data sets will be at the level of galaxy redshift and shape catalogs rather than the pixel-level imaging and spectroscopy simulations described above. They will incorporate varying degrees of complexity regarding galaxy bias, redshift evolution, survey geometry, and observational systematics such as incompleteness, shape measurement errors, and photometric redshift biases. Many of these artificial data sets will be made publicly available, and some will take the form of data challenges, where the underlying parameters are initially known only to the creators of the data set. (*T3, T5*)

- **(D10) Pilot survey proposals with associated figures of merits,** to be executed during the first months of WFIRST operations. These would become part of the final dark energy data set but also pin down remaining astrophysical or instrument performance uncertainties at the level needed to optimize the HLS. We will develop the figures of merit required to quickly assess the

data-quality and make operational decisions regarding the cosmological surveys. *(T2, T6)*

- **(D11) A prioritized program of observations from other facilities,** ground and space-based, needed to calibrate or finalize strategy decisions on the WFIRST dark energy program. *(T6)*
- **(D12) Broad engagement with the cosmological community**, through workshops, talks, publications, and public release of codes and artificial data sets, with the goals of (a) building awareness of and broad support for the WFIRST dark energy program and (b) inspiring the community to develop methods and carry out investigations that will maximize the cosmological return from WFIRST. *(T6)*

3 Weak Lensing and Cluster Growth Investigation [Oli: Chris, Rachel M, Mike, David W., 15 pages]

As discussed in §2.2.3 of SDT15 (written by members of our team), the HLS Imaging survey will (in its current design) measure the shapes of nearly 400 million galaxies in 3 near-infrared (NIR) bands, plus fluxes in a 4th band to improve photometric redshifts (photo-z). With a data set two orders-of-magnitude larger than the current state of the art [? ?], the WFIRST weak lensing program will measure the cosmic expansion history and the growth of structure with exquisite statistical precision, demanding corresponding advances in the control of WL systematics. The cosmic shear power spectrum, which is the basic WL observable, depends on both the distance-redshift relation $D(z)$ and the power spectrum of matter clustering $(\Omega_m h^2)^2 P_m(k, z)$. The WL survey will also enable high-precision cosmological constraints from galaxy-galaxy lensing (GGL) and from galaxy clusters, which can be identified in either the HLS or external data sets and characterized with the help of WFIRST WL. The `CosmoLike` forecasting tool can predict the constraints from these methods individually and in combination with complementary probes such as BAO, RSD, supernovae, and the CMB. While all aspects of our investigation are interconnected, we separate the discussion into more tightly connected loops: requirements, image simulations, and algorithm development at level of pixels and shape measurement; issues related to photometric calibration and photometric redshifts; development of methods for cosmological analysis of WFIRST data and testing them on simulated data; and finally methods for systematic error testing and mitigation.

3.1 Requirements (D1, D3, D4, D5)

[**Authors**: Chris, Rachel M, Mike]

The definition of WFIRST requirements will start in pre-Phase A and continue through CDR. In the early stages, we will focus on identifying the driving requirements for the mission (e.g., those related to total throughput, wavefront error and wavefront stability, and any calibration requirements that would require dedicated hardware or special operational modes). Simple simulation tools will be used for this stage, where fast turnaround times and conservative assumptions will be prioritized over the true end-to-end simulations that will be required for the analysis. As the project matures, we will consider the more complete list of requirements (e.g., specifications on data products) and work with the WFIRST Science Centers (WSC) to build the fully realistic simulations needed to test reduction and analysis pipelines.

Our existing tools (the ETC, operations simulations, and `CosmoLike`) allow us to forecast the statistical power of the WL survey for a specified observing strategy and time allocation and to evaluate the impact of hardware or strategy changes on cosmological constraints (see §6.2). The challenge for the SIT is to define requirements that ensure control of systematic uncertainties at the level of these statistical errors. WL systematics fall into two broad categories, instrumental/algorithmic effects tied directly to the measurements and astrophysical effects tied to the cosmological interpretation of the measurements. The former affect engineering requirements, while

the latter must be predicted theoretically or constrained through observations.

Our basic strategy for defining instrumental/algorithmic requirements is to create simulated HLS images with varying levels of realism, analyze them with proto-type data reduction and WL pipelines, and compare the resulting measurements to the simulation inputs. We will determine the sensitivity of WL shear measurements to each effect, and then flow down requirements from nuisance parameters in `CosmoLike` (e.g., spurious shear) to hardware requirements that can be compared to integrated modeling results (e.g., wavefront stability). Our simulation framework will build on the public code `GalSim` [?], which already includes a WFIRST-AFTA module and to which Co-Is Jarvis and Mandelbaum are lead contributors. Our proto-type pipelines will build on codes for image analysis and shear measurements developed over many years by Co-Is Hirata, Jarvis, Lupton, and Mandelbaum and their collaborators for analysis of SDSS, DES, HSC, and LSST imaging [? ? ? ? ? ?]. WL algorithms are an active area of research, and we will integrate promising new approaches as our investigation progresses. Our overall framework is analogous to, but more focused than, the GREAT3 challenge led by Mandelbaum [?], including the use of different simulation branches where effects are turned on and off (separately or together) to probe the biases resulting from individual or combined effects.

In the domain of instrument- and pipeline-related systematic errors, each error must be evaluated according to the following criteria: (i) What is the raw magnitude of the systematic error compared to statistical errors? (ii) What is the approach to modeling and removing that systematic error? What cross-checks will be necessary to validate that this has been done correctly? (iii) Are there implications to the optimal observing strategy (e.g., dithering, repeat observations, dedicated calibration observations)? (iv) Are there implications for the hardware requirements (e.g., stability, pre-flight characterization, or dedicated flight calibration hardware)?

As an example: optical aberrations induce PSF ellipticity that is 2 orders of magnitude greater than WFIRST weak lensing requirements if uncorrected. It is not practical to eliminate the effect in hardware by requiring the wavefront to be perfect at the few nm level in a wide-field instrument, so a model of the PSF will have to be built. Some contributions to the PSF (e.g., jitter and guiding errors) will need to be measured separately for each exposure. Others vary on longer time scales (e.g., due to thermal changes in the optics) and imply a trade-off between stability requirements and the noise on a measurement of the relevant parameter from a set of N exposures. We will use these considerations to turn high-level requirements on knowledge of the PSF into lower-level requirements on thermal stability and on the operations plan.

Detector systematics are special because their absolute amplitudes may be difficult for the vendor to control or even test. We thus anticipate absolute requirements only on the largest systematic effects such as inter-pixel capacitance (IPC) and persistence, which are being addressed as part of the technology development program. There will be many subtler effects arising in the detectors (examples might include NIR-detector analogues of the brighter-fatter effect, color-dependent charge diffusion, etc.); we will set requirements on the knowledge of these. We will pay special attention to methods of measuring these effects on-orbit using a combination of ground characterization, dedicated calibration modes, and the survey data itself, and the interaction with survey operations and stability requirements (on e.g., focal plane temperature). We will coordinate the best approach to these effects with the Project and with collaborators Seiffert and Shapiro. Shapiro leads the Precision Projector Laboratory (PPL), a JPL facility designed to emulate WFIRST weak lensing data using scenes focused onto WFIRST NIR-detectors. Systematics identified thusly will be studied and mitigated in collaboration with the SIT. Co-I Hirata advises the PPL on image analysis software and interpretation of results [? ?].

Cluster cosmology is generally considered to be less demanding in terms of hardware requirements than cosmic shear, since the large galaxy over-densities and shear signals are not as easily masked by subtle optical aberrations or detector behaviors. Nevertheless, it may place new requirements on survey footprint/operations (to ensure overlap with other data sets); pipeline behavior

in crowded fields (e.g., [**?**]); and ancillary data products and simulations to describe, e.g., changes in selection effects and source redshift distributions in the presence of blending and magnification. WL requirements lead Hirata will work closely with cluster lead Weinberg to ensure that these requirements are captured in the flow-down.

3.2   Requirements (D1, D3, D4, D5)

[**Authors**: Chris, Rachel M, Mike]

3.3   Photometric Calibration

[**Authors**: Nikhil, Chris]

Achieving a uniform photometric calibration of the survey is an essential requirement for achieving the scientific goals of WFIRST, since calibration variations on large scales can swamp the cosmological signals and make dedicated photo-z calibration fields unrepresentative of the whole survey. Our calibration requirements and strategy development will be led by Co-I Padmanabhan, the chief architect of the "uber-calibration" method used for the SDSS [**?**]; members of our team also have experience designing calibration strategies for Euclid and LSST. Co-I Capak is an active member of the Euclid calibration team. Co-I Lupton plays a key role in designing LSST and HSC calibration strategies. Using our simulations and forecasting tools, we will determine both absolute and relative calibration requirements for the survey as a function of angular scale on the sky. Based on these requirements, we will define a calibration plan for WFIRST, including overlapping regions for self-calibration over the entire survey and repeated fields or use of the microlensing survey data for small scale calibration (e.g., flat fields and intra-pixel response). We will also define methods and metrics for tracking the photometric stability and calibration quality across the survey. We will determine formats for making the calibrations and calibration error maps publicly available to enable related analyses. We will pay special attention to the relation of the photometric calibration strategy to instrument design, including the impact of field-dependent bandpass variations and NIR detector effects such as persistence and rate-dependent nonlinearity. We will track which such effects can be measured internally to the WFIRST data, and when dedicated observing modes might be needed (such as observations of the sky with an internal illumination source also on). The latter considerations will feed into the observing strategy and hardware requirements.

3.4   Photometric Redshifts

[**Authors**: Peter, Shooby, Dan]

Accurate photo-zs are crucial to all WFIRST probes of dark energy. We plan to combine calibrations of the color-redshift relation [**?**] and cross-correlation techniques to achieve the required photo-z performance. These will be supplemented by consistency checks such as the expected level of WL cross-correlations for sources in different photo-z bins. Our team is actively developing the color-redshift mapping for WFIRST (supported by WFIRST Preparatory Support (WPS); PI Capak), with initial tests demonstrating a path to meeting the WFIRST requirements on photo-z bias. Our team has worked with NASA HQ to allow the NASA Strategic Keck Proposals to obtain the required spectra. A key piece of the proposed work will be carrying out these spectroscopic surveys. Cross-correlation methods [**? ? ? ?**] require spectroscopic overlap and rely on assumptions about the bias of galaxies. We will determine the applicability of these methods and their implications for survey design. We also plan to investigate promising new methods of photo-z inference [**? ? ?**] as additional consistency checks.

Our team will refine the performance estimates and calibration data requirements for each of these photo-z methods. This will include studying the impact on WFIRST operations (e.g., incorporating observations of spectroscopic calibration fields in the observing plan). Co-I Capak

will lead the photo-z effort, with support from other team members (Mandelbaum, Hirata, Hudson, Jain, Eifler, Krause, Miyatake) who have extensive experience using photo-z calibration techniques for WL applications with SDSS (e.g., [? ?]), Canada-France-Hawaii Telescope (CFHT), DES and HSC survey data.

3.5 Cosmological Forecasting, Simulations, and Methodology Development (D2, D8, D9)

Forecasting. Cosmological forecasting plays an essential role in connecting strategies and requirements defined at the instrument and observation level to WFIRST's top-level science goals. Co-Is Bean, Hirata, Padmanabhan, Wang and Weinberg have been involved in the forecasting for the WFIRST reports and for cosmological surveys and space missions in general. One of our early activities will be to develop a forecasting framework for WFIRST building on Eifler & Krause's `CosmoLike`, incorporating all of the dark energy probes from the HLS Imaging, Spectroscopy, and Supernova surveys. We will incorporate the impact of the leading observational and theoretical systematics through "nuisance parameters" that can be marginalized over in cosmological parameter estimates. This framework will be much more thorough than the one used for the SDT reports, and it will enable complete flow-down and flow-up analyses between the instrument, strategy, and software requirements and the expected cosmological performance of WFIRST. We will initially use analytic approximations for error covariance matrices and the dependence of observables on model parameters. We will steadily update these approximations using the simulations described below. For cosmic shear, we will extend existing work on the WL bispectrum (e.g., [? ?]), which should produce complementary constraints to the power spectrum and may have similar statistical power given the high source density expected in HLS Imaging. Figure 1 illustrates the application of `CosmoLike` to WFIRST multi-probe forecasting.

Cosmological simulations. As emphasized in SDT15, WFIRST will improve the statistical precision of cosmic expansion and structure growth measurements by factors of $5-50$ compared to current data, and will require commensurate improvements in modeling the lensing and clustering signals and the astrophysical contaminants thereof. We will devote significant effort to modeling the nonlinear regime via a combination of numerical simulations, perturbation theory, and empirical measurements, to address issues such as baryonic effects on matter clustering [? ?]; the validity of the Born approximation [?]; and galaxy intrinsic alignments [? ? ?]. We will develop analysis strategies that take advantage of our methodological improvements and marginalize over remaining uncertainties. The computational requirements for cosmological simulations will increase significantly in the years just prior to launch. Where these resources will be sourced from is currently an unsolved issue that is being investigated at the project level. Co-I Kiessling has been part of this effort at the request of the project and she will be responsible for leading the team efforts to quantify computing requirements beyond FY20.

Most of our work will use large-volume N-body simulations (gravity only) in concert with statistical recipes or semi-analytic models for assigning galaxies to dark matter (sub-)halos. A large set of such simulations ranging in size from tens of billions to a trillion particles already exists. These will be made available to the collaboration by collaborator Heitmann and will be augmented over time with more simulations as needed. We will use hydrodynamic simulations of smaller volumes, provided by collaborator Yoshida, for targeted investigations such as the impact of baryonic physics on the matter power spectrum or as a realistically complicated test of non-linear galaxy bias models. From the N-body simulations we will produce mock galaxy catalogs both in fixed redshift cubes and in light cones with realistic survey geometry and redshift evolution. We will use these mocks to understand the effect of observational complications (survey geometry, variable depth and completeness, etc.) on WL and clustering measurements and to provide realistically clustered inputs for the pixel-level simulations described earlier.
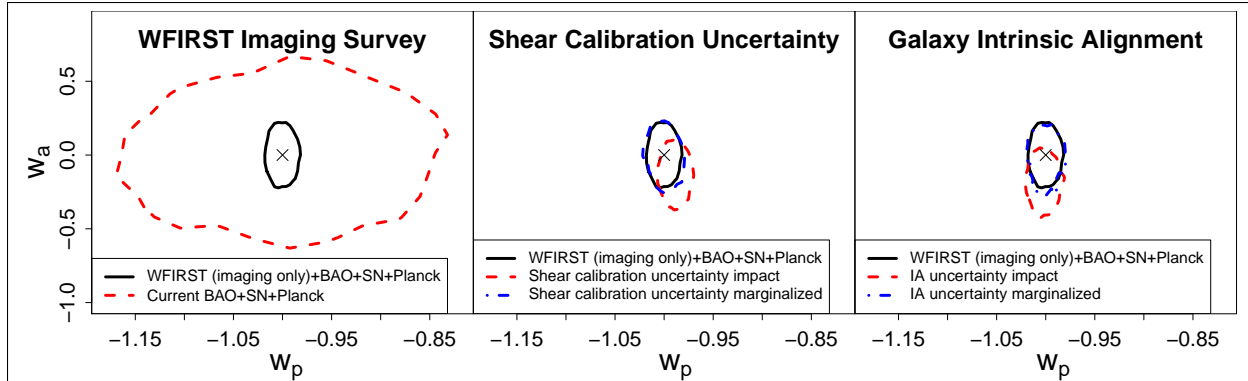
Figure 1: `CosmoLike` forecasts of constraints on the dark energy equation-of-state from combining current CMB+BAO+Supernovae (SN) data with anticipated WFIRST WL, GGL, projected clustering, cluster number counts, and cluster weak lensing measurements. We adopt a flat universe with a DE equation of state $w(z) = w_0 + w_a[1 - (1 + z)^{-1}]$, with the $x-$axis showing the value of w at the pivot redshift (z_p) where it is best constrained by this data combination. The left panel shows the improvement from WFIRST imaging data with statistical errors only. In the middle panel, red contours show the effect of an uncorrected shear calibration error of magnitude 0.004, while blue contours show the result with the same calibration error after marginalizing over five shear calibration nuisance parameters with Gaussian priors of width 0.002. The right panel shows a similar example for an astrophysical systematic, the presence of galaxy intrinsic alignments (see [**?**] for details). In both cases, marginalization removes the systematic error without significantly increasing the statistical uncertainty.

Our team includes Co-I Kiessling and collaborators (Heitmann, Takada, Yoshida) who are leading similar efforts for Euclid, Subaru HSC and PFS, DES, DESI and LSST. In §**??** we describe the computing facilities that will enable these computations. We will publish papers documenting our methodology development and will simultaneously release the associated simulations. This growing library of publicly available simulations will encourage the broader community to develop analysis strategies that will ultimately be applied to WFIRST (and other surveys).

Galaxy-galaxy lensing. The combination of GGL with galaxy clustering is an alternative route to extracting cosmological constraints from an imaging survey. Systematics are significantly different from those affecting cosmic shear analysis, and theoretical studies suggest that the statistical power is comparable [**?**]. The need to mitigate systematics favors a joint modeling approach to cosmic shear, GGL, and galaxy clustering, which requires devising and testing models of non-linear galaxy clustering and its dependence on redshift, building on studies such as [**? ? ? ? ? ? ?**] and including the possible impact of "assembly bias" connected to halo formation histories. We will include GGL in our cosmological forecasting framework and identify any GGL-specific requirements distinct from those tied to cosmic shear.

Clusters. Our efforts in cluster analysis methodology will parallel those for cosmic shear and GGL, facing many of the same issues but in the (somewhat simpler) high mass halo regime. Co-I Weinberg will lead the cluster effort with support from collaborators Rozo and von der Linden. While the WFIRST data set presents some unique issues, we will draw extensively on the machinery being developed for DES by Rozo, which follows the broad strategy laid out in chapter 6 of [**?**], and for LSST, described in [**?**].

The first branch of the cluster effort will focus on the identification and characterization of clusters in WFIRST+LSST (or WFIRST+HSC) imaging. This optical+NIR combination will lead to the best statistical precision, since optical/IR selection can select clusters (at typical redshifts $z \sim 0.5 - 1.5$) down to mass thresholds significantly lower than X-ray or Sunyaev-Zeldovich detection; we expect $\sim 40,000$ clusters in the HLS with masses $> 10^{14} M_\odot$. Building on DES methodology [**?**], we will design and test (in simulations) prototype cluster finders for WFIRST+LSST data. Among the issues we will address are: completeness and contamination of the cluster catalogs as a

function of redshift and richness; biases in estimated richness or cluster redshift; expected scatter between richness and halo mass; accuracy of galaxy photometry in cluster regions; accuracy of shear measurement in cluster regions; reliable separation of foreground, member, and background galaxy populations for weak lensing analysis; and ellipticity- or orientation-dependent cluster selection.

A second branch recognizes WFIRST's unique potential for calibrating cluster mass-observable relations at $z \gtrsim 1$ via weak lensing. Cluster-galaxy lensing (CGL, e.g., [? ?]) is the method of choice for all cluster surveys overlapping weak-lensing surveys, but requires sufficient number densities of galaxies behind the clusters with accurate photo-zs. For cluster mass calibration at $z \gtrsim 1$, space-based shear measurements and photo-zs from the combination of LSST and deep NIR photometry, such as delivered by WFIRST, are therefore necessary. Calibrating cluster mass-observable relations furthermore benefits tremendously from the availability of multi-wavelength mass proxies (e.g. [?]), hence we will consider the synergies with X-ray (eROSITA) and SZ (Planck, AdvACT, SPT-3G, CMB-S4) measurements, with special attention to the impact on survey footprint placement. We will also investigate the potential for HLS-CMB cross-correlations to extract the kinetic Sunyaev-Zeldovich signature to constrain dark energy, gravity and and neutrino mass sum [? ?]. Magnification of special galaxy subsamples can provide a cross-check on the shear-based lensing effort (e.g., [?]).

For modeling methods, we will develop a comprehensive approach that combines CGL and cluster-galaxy cross-correlations to extract cosmological information from fully non-linear, trans-linear, and linear scales, extending and unifying methods based on the cluster mass function (e.g., [?], [?]), cluster mass-to-light or mass-to-number ratios [? ?], large scale cluster-mass correlations [?]. We will test the robustness of these methods to all of the observational effects listed above. For our cosmological forecasting, we will integrate clusters with our `CosmoLike` treatment of WL and the GRS to account for correlated statistical errors (from large scale structure in the HLS survey volume) and common systematics (shear calibration, photo-z errors).

3.6   Systematics Testing and Mitigation (D8)

Achieving WFIRST's precision cosmology goals requires eliminating systematic biases while minimizing statistical losses, and *demonstrating* that biases have been removed. We will develop a 3-pronged approach to this challenge.

Marginalization. As illustrated in Figure 1, a general strategy for both instrumental and astrophysical systematics is to describe their possible impact by nuisance parameters and marginalize over these in cosmological analysis. Important elements in making this strategy effective are: (i) devising concise templates that describe the systematics with minimal numbers of parameters; (ii) setting realistic priors on nuisance parameters; (iii) combining multiple observables that can break degeneracies. We will develop this approach for treating observational effects, particularly shear measurement systematics and photo-z biases, and astrophysical effects, particularly the impact of baryons on the matter power spectrum [? ? ? ?] and galaxy intrinsic alignments (IA). For the observational effects, we will use our simulations (§4.1) and photo-z investigations (§4.2) to design templates and determine appropriate priors. For the astrophysical effects, we will draw on our team's extensive experience with analytic and numerical modeling of IAs [? ? ? ? ? ? ? ? ?] and on the work of Eifler and Krause in incorporating baryonic and IA effects into cosmological WL analysis [? ? ? ?]. This will include the principal components approach, which allows one to marginalize out the directions in observable space that are most sensitive to the choice of prescriptions for baryonic physics in simulations [?]. Cosmic shear, GGL, CGL, and galaxy clustering each respond differently to these systematics, so we anticipate that joint analyses will allow much tighter constraints on both systematics and cosmology than any probe in isolation.

Systematics maps. A second approach to identifying and removing observational systematics is to cross-correlate the signal being measured with maps of possible systematic effects, such as

stellar density, PSF size, or Galactic extinction (e.g., [?]). These methods both measure the impact of systematics and provide a template for removing them. We will devise a system of such methods for WFIRST WL and galaxy clustering measurements and test their efficacy on our simulated data sets. This effort will be led by Co-Is Ho and Padmanabhan, who developed such approaches for their analyses of large scale galaxy and quasar clustering in the SDSS [? ?].

Null tests, internal consistency, and external data sets. WL analyses must be validated using internal tests that the measurements should pass for any set of cosmological parameters but may fail in the presence of systematics (e.g., [?]). For example, the cross-correlation between PSF-corrected galaxy shapes and star shapes should be consistent with zero, many statistics associated with B-mode shear (e.g., [?]) should vanish, and consistent cosmological results should be obtained when using the largest or smallest 50% of the source galaxies. Drawing on our team's experience with other surveys, we will ensure that there is a coherent pipeline for carrying out these standard tests on WFIRST data, including the many consistency tests enabled by having shape measurements in multiple bands. We will pay close attention to tests that make use of unique properties of WFIRST data; for example, comparison of shapes measured on subsets of an exposure with multiple non-destructive reads would test for the impact of detector non-linearities.

Cross-correlations with external imaging and spectroscopic surveys (Kilo Degree Survey (KiDS), HSC, DES, PFS, DESI, LSST, Euclid) offer multiple opportunities for improving the HLS analysis, including photo-z calibration and tests for shear systematics. For example, the cross-correlation of WFIRST and LSST shapes will evade additive systematics that impact only one or the other survey, e.g., those coming from the atmosphere for LSST or from detector effects specific to WFIRST. Other validations can be performed using surveys that measure similar quantities as the HLS but using different techniques, such as cluster masses estimated via CMB lensing. We will investigate a variety of possible tests, evaluate the hardware and operations implications (e.g., footprint overlap, joint data management), and ensure that they are represented in the FSWG process.

4　Galaxy Redshift Survey Investigation [Oli: Yun, Lado, Shirley et al., 15 pages]

As discussed extensively in §2.2.4 of SDT15 (written by members of our team), the defining goal of HLS spectroscopy is to derive constraints on dark energy from a slitless spectroscopic (grism) redshift survey of approximately 20 million emission line galaxies (ELG) in the redshift range $z = 1-3$. The galaxy redshift survey will enable high-precision measurements of the cosmic expansion history via BAO and structure growth via RSD. Acoustic oscillations in the pre-recombination universe imprint a characteristic scale on matter clustering, which can be measured in the transverse and line-of-sight directions to determine the angular-diameter distance $D_A(z)$ and Hubble parameter $H(z)$, respectively [? ? ?]. Anisotropy of clustering caused by galaxy peculiar velocities constrains (in linear perturbation theory) the combination $\sigma_m(z)f_g(z)$, where σ_m describes the rms amplitude of matter fluctuations and $f_g(z) \equiv d\ln\sigma_m(z)/d\ln a$ is the fluctuation growth rate. Thus the GRS on its own can address the key questions identified by NWNH: whether cosmic acceleration is caused by modified gravity or by dark energy, and whether (in the latter case) the dark energy density evolves in time [? ?]. These tests become more powerful in combination with weak lensing and cluster measurements from HLS Imaging and high-precision relative distance measurements from the Supernova Survey [? ?]. The broadband shape of the galaxy power spectrum and higher order measures of galaxy clustering provide additional diagnostics of dark energy, neutrino masses, and inflation, and insights on the physics of galaxy formation. While all aspects of our GRS investigation are interconnected, we organize it in a structure similar to that of §3 for clarity: requirements, simulations, and prototype pipelines in §4.1, cosmological forecasting, modeling, and cosmological simulations in §4.3, and systematics testing and mitigation in §4.4.

4.1   Requirements

[**Authors**: Yun, Lado]

The most important task of the SIT in guiding development of the WFIRST HLS spectroscopy is to set and validate the requirements of the instrument, the data reduction software, and the survey. The GRS Lead Co-I Wang will work closely with PI Doré and Co-Is Hirata and Teplitz on setting requirements for the GRS. To make fully informed decisions, the team requires high fidelity simulations of both instrument performance and the observable sky that the instrument will measure. The team must also ensure that the analysis of these data by the reduction pipeline will be of sufficient quality to enable measurement with the high precision needed for cosmology. These simulation and pipeline activities will require the team to coordinate with the WSC. Several members of our team (Wang, Teplitz, Capak, Helou) are located at the Infrared Processing and Analysis Center (IPAC), and work closely with the WSC. To the extent practical, we will draw on tools created by the WSC and design our own software and simulations to be useful to them. We note that our work primarily demands the ability to quickly and flexibly simulate different configurations and analyze the results with different algorithms, while the WSC has the task of developing tools for the community and production ready pipelines that integrate with the full WFIRST data system.

Deriving requirements. As with WL (§3.2), we will focus first on GRS requirements that may drive hardware choices, i.e., those that may be demanding in terms of grism design, detector properties, stability and repeatability of pointing, or dedicated calibration hardware. We will include a prioritized list of effects to incorporate in grism simulations. Over time, we will use our increasingly realistic network of simulations to evaluate the impact of requirements and possible trades from the pixel level through to cosmological inferences. The starting points for this process are the WFIRST ETC and survey planning software and the `CosmoLike` forecasting tool.

The maximum achievable statistical power of the GRS is determined mainly by the telescope aperture, throughput, detector area and pixel size, and allotted observing time. However, the statistical power and uniformity of the GRS are further affected by numerous aspects of instrument performance and survey design, e.g., spectral resolution, detector read noise and persistence, dither and roll angle pattern, image quality, complexity of non-1st order features, spatially varying thermal background due to the warm telescope, scattered light from bright stars, repeatability of the grism positioning, and accuracy of calibration of the wavelength-dependent PSF and distortion map.

Non-uniformity of the survey, which is inevitable to some degree, can be corrected in clustering measurements by weighting galaxies to account for incompleteness. However, large corrections typically come at a cost in statistical power, and imperfect knowledge of the non-uniformity leads to systematic errors in the inferred clustering. The other important source of observational systematics is contamination of the redshift catalog by artifacts or objects with incorrectly determined redshifts, and loss of objects from the catalog because of catastrophic redshift errors or uncertainties in the flux calibration. We will define requirements such that (a) the statistical power of the GRS is close to the maximum allowed by the telescope aperture and detector area and (b) the impact of uncorrected observational systematics is small compared to the statistical errors. The expected precision of the galaxy power spectrum provides a useful guide to the statistical power of the GRS, but the full question of cosmological constraining power depends on the astrophysical modeling techniques used to interpret the measured clustering, as discussed in §4.3 below. By the end of our investigation we will have a complete set of tools to evaluate the impact of hardware or strategy trades, changes in requirements, or changes in astrophysical inputs on the expected cosmological return from the GRS.

Simulations. Our development of requirements and a prototype spectroscopic pipeline will rely critically on realistic simulations of the pixel-level grism images. We will work closely with the WSC on developing these simulations for a variety of cases, ranging from simple widely separated

sources to realistically clustered galaxy populations drawn from the cosmological simulations described in §4.3. Our team has extensive experience in producing such simulations for the Hubble Space Telescope (HST) and Euclid. Co-I Tepliz is one of the leaders of the HST Wide-Field Camera 3 (WFC3) IR Spectroscopic Parallel survey (WISPs), for which pixel simulations are vital in assessing completeness and other parameters [**?**]; Co-I Wang (with Teplitz and Capak) is developing simulation techniques for the Euclid grism survey. Co-I Teplitz will lead our grism simulations for WFIRST. A critical astrophysical input for these simulations is the redshift-dependent luminosity function of Hα and [OIII] emitters, which is currently uncertain at levels that have an important impact on WFIRST strategy and performance forecasts. Co-Is Teplitz and Wang are part of an HST archival study to reprocess existing data from multiple HST projects to mitigate systematic uncertainties of the Hα luminosity function (LF) measurement. Through WISPs, Teplitz is also working to obtain significantly more HST data to improve the Hα LF measurement. In addition, realistic galaxy templates are vital to the forecasting of grism measurements, and we are working to improve both line diagnostics and prediction of line vs. continuum properties.

Prototype pipeline. We will build a prototype spectroscopic pipeline for the analysis of slitless spectroscopic data to produce a redshift catalog. This is a complex, multi-step process. Co-I Teplitz has extensive experience with HST slitless spectroscopy using the WFC3, HST Near Infrared Camera and Multi-Object Spectrometer (NICMOS), and HST Space Telescope Imaging Spectrograph (STIS) instruments [**?   ?   ?**]; he will lead our work on prototype pipelines for the GRS. We will take the basic steps implemented for WFC3 processing as the starting point for a prototype WFIRST pipeline. This pipeline must clean the grism images of contamination from detector artifacts and cosmic rays, register and combine images from separate dithers and roll angles, match objects in the dispersed and direct imaging exposures, extract wavelength- and flux-calibrated 2D spectra, infer redshifts based on detected emission lines, and measure emission-line fluxes and other spectroscopic characteristics. The resulting catalogs are the input for the clustering analyses discussed further in §4.3.

Analysis challenges. Slitless spectroscopic analysis presents several important challenges. First, the pipeline must mitigate the confusion caused by overlapping spectra. The standard solution (used by the HST data pipeline) is to subtract a model of neighboring objects from each source as it is extracted. We will investigate the use of HST-like algorithms for WFIRST, as well as more sophisticated solutions that could produce better results, such as fitting the full pixel set for regions of the frame. Model dependent solutions, with iterative fitting, are potentially promising, but biases would have to be carefully understood. While the HLS obtains exposures at multiple roll angles, these may be greatly separated in time. This could introduce new problems for variable sources, or in fields with foreground moving objects. We will also develop methods to automate quality assessment and flagging of extracted spectra, as the sheer volume of WFIRST GRS will make human review of spectra (standard practice in current grism surveys) impossible.

A second major challenge is the need to mitigate catastrophic redshift errors. Such failures arise from the misidentification of redshifts (e.g., confusing [OIII] for Hα in low signal to noise (S/N) spectra) or false-positive line detections caused by noise peaks or unflagged cosmic rays. Redshift fidelity can be greatly improved by using the photometric redshift estimates derived from the multi-band photometry as a prior in the redshift determination. Co-I Capak is spearheading multidimensional analysis of galaxy color information for WFIRST photometric redshifts, and that work will be folded into the spectroscopic pipeline.

Completeness maps. Nearly as important as the redshift catalogs themselves is the production of completeness maps that characterize the spatially varying depth of the survey, the level of contamination, and regions that should be masked because the catalog is unreliable. These completeness maps are used to weight galaxies in clustering analysis and/or to create random catalogs such that the local number density of points is proportional to the likelihood of successfully measuring a redshift of a galaxy if it were at that point. Co-I Samushia is leading a similar effort

in DESI and has previously worked on quantifying and removing systematic effects associated with inaccuracies in random catalogues [**?**]. Co-I Ho has also led the effort in creating the SDSS-BOSS LSS catalog and randoms [**?**] and led the effort in removing observational effects in BOSS LSS catalog [**?**]. Co-I Samushia will lead our work to develop tools for creating these completeness maps by a full forward-modeling method, where artificial sources are assigned random angular positions and redshifts, added to grism images, and pushed through the data pipeline. Compared to existing large redshift surveys (from ground-based fiber spectroscopy), the WFIRST completeness map will have much more complex small scale structure because of the varying numbers of exposures at individual points on the sky and sensitivity variations across the focal plane. Because of source confusion and sky background effects, the completeness and contamination will be a function of the local galaxy surface density. We will develop strategies and tools for recording these large and complex completeness maps in formats that can be efficiently used to create random catalogs and weight galaxies for clustering analyses. By the end of the investigation period, we will be able to create full pixel-level simulations from an input cosmological simulation (see §4.3), run them through our proto-type pipeline to create a redshift catalog and completeness map, and analyze the resulting artificial data set with our clustering analysis tools to compare to the idealized case that has the complete galaxy catalog of the cosmological simulation.

Calibration strategies. We will define the absolute and relative calibration requirements for the GRS, such as the angular scale and temporal stability. We will develop methods for calibrating the relative and absolute flux measurements along with wavelength calibration and redshift accuracy and completeness. For flux and wavelength calibration we will set requirements on the ground testing, in flight calibration sources, and calibration observations based on experience with other missions including Spitzer, HST, and Euclid. Furthermore, we will investigate self calibration strategies based on optimizing dither patterns and exposure times for the science observations and the use of touch-stone fields that both calibrate and provide long-term trending of the data. Both the primary and self calibration procedures will be tested with simulations specified by this SIT and conducted by the WSC. Finally, we will use the large spectroscopic surveys necessary for the weak lensing photo-z calibration to verify the calibration by directly testing the redshift accuracy and completeness estimates from the simulations. Co-Is Capak and Padmanabhan, both with extensive experience from similar work for Euclid and BOSS, will lead our calibration work.

4.2   Simulations

[**Authors**: Shirley, Elena, Andrew, Alina, Alex, Yun]

4.3   Cosmological Forecasting, Modeling, and Simulations (D2, D8, D9)

Forecasting. Our initial forecasts for the cosmological constraints from the WFIRST GRS will adopt the model-independent approach (incorporating both BAO and RSD) that Co-I Wang has developed [**?**] and used for the WFIRST SDT reports and similar forecasts for Euclid [**?**]. This approach offers a fast way to forecast how uncertainties in $H(z)$, $D_A(z)$, growth rates, and other cosmological parameters change in response to changes of the survey strategy, instrument performance, or astrophysical inputs such as the Hα luminosity function. Early in this investigation, we will incorporate a full description of redshift-space galaxy clustering into `CosmoLike`, using a halo occupation density (HOD) framework similar to that already implemented for angular galaxy clustering [**?**]. In the medium term, we will also incorporate effects of clustering measurement and theoretical modeling systematics via nuisance parameters, analogous to our existing treatments of observational and theoretical systematics in weak lensing analysis. The expected level of these systematics will be informed by the studies described in §4.1 and below. As with the imaging survey, this comprehensive forecasting framework will enable us to connect low-level technical requirements to our top-level science goals.

GRS modeling. The development of the methodology for the interpretation of GRS data is centered on the mitigation of the astrophysical systematic effects for galaxy clustering measurements: nonlinear effects, RSD (growth rate signal on large scales and contamination on small and intermediate scales), and galaxy bias (the difference between galaxy and matter distributions). Co-I Padmanabhan is a leading expert in BAO/RSD data analysis [**? ? ? ? ?**]; he will lead our work in GRS modeling/interpretation methods, with participation from the PI and Co-Is Bean, Ho, Samushia, Spergel, Wang, and Weinberg [**? ? ? ? ? ? ? ?**].

BAO measurement is now a mature field, but WFIRST probes new regimes of precision and redshift using different instrumental choices and different classes of galaxy tracers from previous surveys. Effects of non-linear clustering and galaxy bias are expected to influence BAO measurements at the $\sim 0.5\%$ level [**?**], which is significant compared to WFIRST statistical errors. Reconstruction methods [**? ? ?**], which attempt to reverse the nonlinear evolution of the BAO feature, appear to remove most of this effect while simultaneously improving the precision of BAO measurements. Current observational studies use very simple reconstruction algorithms. We will explore more sophisticated reconstruction methods, building on low redshift work [**?**], and test their performance on simulations of WFIRST galaxy redshift catalogs, including realistic treatments of survey geometry, redshift evolution, source space density, and variable completeness. Building on current work by Co-Is [**? ? ? ?**], we will also investigate improved clustering estimators that can sharpen the precision and improve the robustness of BAO measurements.

In sharp contrast to BAO measurements, cosmological inference from RSD measurements is already limited mainly by uncertainties in theoretical modeling, with application of different models to the same underlying data yielding differences at the $\sim 10\%$ level. Furthermore, the statistical signal-to-noise ratio of RSD measurements increases rapidly with decreasing scale, so there are potentially large gains from modeling that extends into the fully non-linear regime. We will pursue a variety of approaches to improving and testing RSD models, including the efficient method of computing predictions numerically by populating the halos of N-body simulations with galaxies. Co-I Spergel has expertise in combining imaging and spectroscopic data to predict the relationship between galaxies and halos [**?**]. One can think of this method as producing "emulators" [**?**] that predict galaxy clustering statistics as a function of cosmological parameters and parameters that describe the relation between galaxies and dark matter halos. The approach shows promise (e.g., [**?**]), but it relies on parameterized models for populating halos, and the accuracy of these needs to be tested against galaxy catalogs constructed in ways that do not share the same assumptions (e.g., by semi-analytic models or abundance-age matching). We will use similar techniques to investigate the impact of non-linear evolution and bias on the broadband galaxy power spectrum, and thus improve our ability to extract cosmological information from this measurement.

Galaxy bias is likely scale-dependent; its testing will require realistic ELG mocks, and its mitigation will require the successful measurement of the higher-order statistics of galaxy clustering, which in turn requires a sufficiently high galaxy number density for the GRS. An essential difference between the WFIRST and Euclid spectroscopic surveys is that due to the much smaller pixel scale ($0.11''$ for WFIRST versus $0.3''$ for Euclid) and larger telescope aperture, WFIRST is capable of carrying out a significantly deeper GRS, which can result in a much higher space density of the WFIRST sample over most of its redshift range. This high sampling density represents a significant science opportunity for WFIRST; in particular it will boost the significance of higher-order correlations. Building on previous work by members of our team [**? ? ? ?**], we will investigate techniques that use the galaxy bispectrum and other higher-order statistics to sharpen cosmological constraints, by directly probing the matter density and velocity fields and by improving knowledge of "nuisance parameters" that describe galaxy bias. We will examine ways that RSD measurement precision can be improved by cross-correlating multiple tracer populations with different clustering bias to suppress cosmic variance [**? ? ?**], weighting galaxies by mass to suppress shot noise [**?**], and building group catalogs to collapse fingers-of-God [**?**]. We will

investigate potential gains from cross-correlating the WFIRST galaxy redshift catalogs with H I 21 cm "intensity mapping" measurements, or with CMB measurements, or (at $z > 2$) the Lyα forest. In all of these studies, we will pay particular attention to the influence of the sampling density, as this directly informs the trade between depth and area in the GRS (see §6.2).

Cosmological simulations. The cosmological simulations described in §3.5 will also be useful for the methodology development outlined above. However, the optimal simulations for BAO and RSD studies will typically be larger volume and lower resolution than those for weak lensing: large volumes are needed for good statistics and to eliminate finite box effects, but we do not need to model the small scale matter distribution or baryonic effects (which are encoded in the models used to populate halos with galaxies). Furthermore, simulations tuned to the WFIRST GRS need only be evolved to $z = 1$. Co-I Ho will lead our cosmological simulations for the GRS, with participation from team members Benson, Heitmann, Kiessling, Wang, and postdocs.

We will leverage the participation of several of our team members in Euclid and DESI to produce large simulated ELG catalogs, building on work we have already begun for these projects. To model emission-line selection, we will use both semi-analytic galaxy formation models (SAM) and HOD models that are tuned to produce observed number densities and clustering. We expect to be able to provide ELG mocks for WFIRST similar to those used by Euclid on a short time scale. We will incorporate these into the early pixel-level simulations described in §4.1, which will in turn be used to assess impacts of incompleteness, contamination, and redshift errors on the galaxy distribution.

We will base our first cosmological volume ELG catalogs on two very large simulations that are already available to us through collaborator Heitmann: the Outer Rim simulation, covering a volume of $4.225\,\mathrm{Gpc}^3$ with a particle mass of $\sim 2 \times 10^9 M_\odot$, and the Q Continuum simulation, covering a volume of $1.3\,\mathrm{Gpc}^3$ with particle mass of $\sim 10^8 M_\odot$. The Outer Rim simulation was used to create the simulated ELG catalog for DESI. We will combine the halo populations from these simulations with the Galacticus SAM code [**?**] to create clustered ELG catalogs, some in fixed-redshift cubes for methodology tests and some in light cones with realistic survey geometry and redshift evolution. Our most ambitious simulation efforts, later in the investigation period, will take full light-cone ELG catalogs, create pixel-level simulations that span the entire HLS area, and analyze these simulations with the proto-type pipeline to produce "observed" galaxy catalogs. We can then apply the full clustering measurement machinery (including corrections for varying completeness) to these catalogs and apply our cosmological inference tools to understand the impact of observational systematics on cosmology from the WFIRST GRS. As with the weak lensing investigation, we will investigate computational requirements beyond FY20 and techniques to reduce them. We will also make our simulations publicly available so that others can develop and test their own methods, with some of them released in the form of blind data challenges.

4.4   Systematics Testing and Mitigation (D8)

The galaxy clustering measurements are susceptible to observational and astrophysical systematic effects. We discussed the astrophysical systematic effects and their mitigation in §4.3. We now focus on the observational systematic effects.

As we discussed in the §4.3, we will take full light cone ELG catalogs, create pixel-level simulations that span the entire HLS area, and produce the observed galaxy and corresponding random catalogs. We envision that multiple LSS catalogs will be produced by extracting the galaxies (and their corresponding randoms) according to their specific continuum levels and/or line-fluxes (or other selection criteria). For each of these catalogs, we will test for systematics such as effects of stellar density, dependencies on line luminosity, continuum luminosity and varying exposure number. Co-I Ho led investigations in effects of observational systematics on galaxy over-density in BOSS [**?**]; we will apply the same techniques in detecting galaxy over-density variations due to the various potential systematic sources. Co-I Ho will lead our work to design and perform inter-

nal empirical tests that involve dividing the galaxies (and corresponding randoms) into subsets of different continuum luminosity, line luminosity, galaxy environment, exposure number and other relevant parameters. We will pay particular attention to potential systematic effects caused by the complex structure of the completeness function, as a function of redshift. This is particularly important as the number of exposures can vary significantly, from 0 to 10 with a median of 7 in the SDT15 observing strategy.

Team Co-Is have experience in using template projection method and cross-correlation method in photometric clustering to mitigate observational systematics such as stellar density, PSF variations, magnitude error fluctuations [? ?]. We will adapt these methodologies to remove systematics in 3D clustering. We will test our 3D systematics detection and mitigation methodology on our simulated catalogs and check whether we achieve unbiased results in BAO distances and the growth rate of large scale structure. These potential systematics will also affect the photometric clustering and the full shape of the galaxy power spectrum, which can be powerful in constraining the sum of neutrino masses.

5   Cosmological Forecasts [Oli: Tim, Elisabeth, Olivier, 10 pages]

6   Operations Model for the HLS and Evaluation of Trades [Oli: Chris, 5 pages]

The optimization of the WFIRST HLS will be tightly linked to operations simulations, which inform the possible range of footprint area and location, depth in each filter (or grism), redundancy, and temporal distribution of exposures. We propose a highly integrated approach, with the operations simulations at the core, but with links to pixel-level simulations to assess required redundancy, cosmological forecasting tools (`CosmoLike`) to assess science reach, and comparison to the observing regions of other surveys and telescopes to maximize synergies and meet requirements for deep fields and photo-z calibration.

6.1   Operations model (D7)

Co-I Hirata will lead the development of the HLS observing plan, extending his previous tools used for the SDT. These tools are already highly advanced, incorporating observing constraints in the candidate orbits (Geostationary Earth Orbit (GEO) or at the Lagrange point L2), an exposure-by-exposure observing sequence optimized with detailed model of overheads, and tiling/coverage maps including field distortions and curved sky effects. These tools treat both imaging and spectroscopy with unified functions and scripts, and are well suited to joint survey optimization when both hardware parameters (e.g., reaction wheel orientations) and the observing program (e.g., depth vs. area) are considered. Hirata's operations tools were used by the SDT for "proof of principle" studies, but we will now extend them to (i) evaluate survey performance at intermediate stages (e.g., what information is available after 1, 2, or 4 years); (ii) include a pilot survey in the first few months of the mission to validate HLS performance and make any "early changes" necessary; (iii) include dedicated calibration observations as needed; and (iv) export field overlap statistics as needed to assess internal calibration strategies such as "uber-calibration" [?]. This work will be performed in consultation with other science groups on the FSWG and with the WSC.

The SDT coverage maps included a study of the overlap with the LSST footprint, but with our links to other projects we will understand the overlap with the other cosmology surveys as well as the accessibility of our fields from ground-based telescopes. This information, combined with the community workshops (§7), will be used to maximize synergies with other facilities as well as ensure that potentially conflicting requirements (e.g., overlapping LSST photometry and photo-z

calibration including Northern telescopes such as Subaru or Keck) can be met.

6.2 Survey Optimization (D10, D11)

The key constraint in survey optimization is the limited amount of observational time available, since WFIRST is life-time limited and has multiple science focus areas. For fixed instrumental capabilities and observing time, the primary decisions on survey strategy are the trade between depth and total area, and the balance between imaging and spectroscopy. These decisions are driven by two major considerations: (i) *precision* – to maximize the DE science return of WFIRST, taking advantage of synergies with other surveys; and (ii) *accuracy* – tight control of systematic uncertainties, to ensure correct DE measurements. The combined expertise of our team in WL and GRS will enable us to rapidly evaluate these trades, informed by results from previous and concurrent surveys such as DES/DESI/Euclid. Furthermore, we plan on developing pipelines to rapidly analyse "pilot" data from WFIRST to quantify the telescope and instrument performance, as well as developing well defined metrics to guide the survey trades.

WL Survey. The statistical power of a WL survey scales with the total number of galaxy shape measurements, the product of the survey area and the effective surface density. The SDT2015 report adopts an HLS imaging exposure time that yields roughly equal contributions from read noise and sky noise in the most sensitive filters, which approximately maximizes the total number of shape measurements. This is a compromise between minimizing overheads and read noise (which favors a "deep" mode), versus the shallow number counts of resolved WL sources (shallower than $N \propto F^{-2}$, which favors a "wide" mode). However, we will re-examine the depth vs. area trade using higher-fidelity tools (e.g., incorporation of shape measurement noise from realistic simulations, and updated detector properties) and propagating the trades all the way to cosmological parameters (using `CosmoLike`). We will also investigate the performance of hybrid strategies in which a four-filter survey over part of the HLS area is combined with a one- or two-filter survey over a larger area. This approach would yield more independent shape measurements at the cost of greater photo-z uncertainties and complicating cross-checks of systematics, so assessing it requires attention to the entire path from data to scientific conclusions, including the impacts on scientific investigations outside of dark energy and cosmological parameters.

GRS Survey. The depth vs. area trade for the GRS is driven by two competing factors: for deep surveys, where the number density of galaxies n is large ($nP \gg 1$, where P is the power spectrum at a given scale), the information per unit area saturates; whereas for wide surveys, overheads reduce inefficiency and galaxy shot noise inflates the statistical errors. In the SDT15 survey design, the GRS covers the same area as the HLS imaging survey ($\approx 2,200\,\mathrm{deg}^2$), to a 7σ limiting line flux of $\sim 10^{-16}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ over most of the grism bandpass. This yields approximately the largest number of galaxy redshifts for a fixed total observing time. SDT15 found that doubling the survey area at fixed observing time (even without additional imaging) *reduces* the precision of BAO measurements because of the rapid increase in galaxy shot noise with decreasing spectroscopic depth. However, this conclusion is sensitive to the luminosity function of Hα emitters at $z = 1 - 2$, which remains uncertain [**? ?**]; Co-Is Teplitz and Wang will lead our efforts to reduce this uncertainty and feed the results into optimization of the GRS.

We will study the link between the observing strategy and systematic error validation. For example, pixel-level simulations will predict the variation of galaxy density with respect to the diversity of roll angles, use of subsets of the exposures in the data reduction, etc.; our team has extensive experience with the measurement of such dependences in imaging surveys (Co-Is Ho, Hirata, Padmanabhan) and will extend this to the GRS. The hardest part of the GRS error validation will be galaxy biasing: while it is possible to "test" bias models against a range of simulations with plausible prescriptions for galaxy placements (HODs, SAMs, etc.), we will ultimately need to validate galaxy biasing models against the data itself. The most promising avenues are higher-order statistics (such as 3-point correlations) and the consistency of results obtained from subsets

of the galaxy sample (e.g., split by magnitude or color). Both of these demand larger nP and drive the survey deeper than the power spectrum measurement. We will develop the tools to assess the precision of these validation tests and will incorporate them into the survey optimization.

A key consideration in the depth vs. area trade is the relation to the Euclid and DESI redshift surveys; the optimization of WFIRST GRS should avoid portions of this space that will already be well-explored by these other projects and focus on the unique capabilities provided by a large telescope in the low-background space environment.

WFIRST HLS Survey. Optimization of the WFIRST HLS requires the end-to-end approach outlined in this proposal, considering everything from the read noise of the detectors and contamination of redshift catalogs to the theoretical modeling systematics for higher order clustering statistics. It also requires joint consideration of the imaging and spectroscopic surveys: multi-filter photometry (as planned in SDT2015) is required to suppress contaminants in the GRS [**?**]. We will provide, in addition to our own recommendations, publicly available tools that can be used by future science teams, the WFIRST Project, and the community at large to evaluate a variety of HLS strategies, including changes to the balance between imaging and spectroscopy. These tools will enable us and others to incorporate updated information about instrument and software performance, empirical data on the galaxy population, progress in modeling techniques, the expected contributions of other projects like DESI, LSST, and Euclid, and changes in the experimental dark energy landscape that highlight the importance of particular classes of measurements.

7 Community Engagement and External Data-sets (D11, D12) [Oli: Olivier, 5 pages]

The coming decade will be an exciting time for cosmology. Before WFIRST launch, major cosmological imaging surveys (KiDS, HSC, DES) and the DESI and PFS spectroscopic surveys will significantly advance our current understanding. WFIRST, Euclid and LSST will then go further and survey the sky at optical and infrared wavelengths, the James Webb Space Telescope (JWST) and the Extremely Large Telescopes (ELTs) will make very deep maps of the sky; eROSITA will survey the X-ray sky; CMB-S4 will make a deep map of the millimeter sky; and the Canadian Hydrogen Intensity Mapping Experiment (CHIME) and other radio surveys will map the large-scale distribution of H I. A goal of the pre-CDR activities will be to determine the analysis infrastructure and observations needed to achieve the full potential of WFIRST in combination with these surveys. This is best done through a broad community effort that brings together scientists from these complementary projects. Our SIT team is extremely well-placed to do this, through team members' significant roles in a number of these projects and ties to the other major surveys. We will engage the community to identify and pursue the key areas where WFIRST and the concurrent projects will provide new opportunities to mitigate systematics and enhance the combined cosmological science return.

Through organizing a number of open workshops (with some modest travel support budgeted for the team and community members) our SIT will incorporate the interplay between major planned surveys and WFIRST into the WFIRST strategy, to identify: (i) pre-launch observations, (ii) how these external data sets affect the WFIRST observing strategy (e.g., deep fields) and the instrument, and (iii) the software needed (to be built post-CDR) for combining these data sets.

Themes we have identified for these multi-experiment workshops include: (i) Statistical methods for next generation cosmological analyses (ii) Strong lensing: synergies between ground and space (iii) Photometric redshift accuracy for next generation surveys (iv) LSST and WFIRST: data simulations and joint analyses to tackle systematics and reveal the dark sector.

Furthermore, through our team's international scientists in Canada and Japan, we plan to also engage these communities in WFIRST science. If these countries become partners, we hope that

our international co-Investigators will receive support from their agencies and that they will serve as points of contact with their scientific communities and with the CHIME project (Smith) and the Subaru HSC/PFS project (Takada/Yoshida).

8   Future Workplan [Oli: Olivier, all, 5 pages]

Team Management. We have structured our team to cover all the areas of expertise required for the proposed work, and to maximize the synergies between WFIRST and the cosmology community. Scientific decisions including the allocation of funds will be made by the PI in consultation with a Steering Committee consisting of the PI, the two topic leads (Hirata and Wang), the topic sub-lead Weinberg plus Spergel, who have extensive organizational experience through WMAP, ACT, SDSS, National Research Council (NRC) and NASA committees, and other activities. While the PI will have final authority on these decisions, this Committee has the mission experience and breadth of knowledge needed to advise the PI. Monthly Steering Committee telecons will review the budget and priorities so that our effort remains focused on the scientific success of WFIRST. Each topic lead will report to the PI on a monthly basis, to facilitate monitoring of all team work, enable advice on budget priorities, and ensure a regular review of work effort. The PI will ensure proper communication with the WFIRST program office and will maximize collaborations with other WFIRST SITs. For example, we anticipate jointly developing and sharing low-level image simulation tools with SN-focused SITs. Topic leads will closely monitor the progress of the technical tasks (listed in §§ 2,3,4) in association with the deliverables and will adjust the scope of the work according to developments in the project and in the field.

    Risk Management. Inefficient collaboration/coordination presents the most significant risk for accomplishing our program. Within our SIT, this is mitigated partly by most Co-Is having experience working together in existing (or past) projects. The PI will coordinate the team's effort, and ensure it is well-integrated into and aligned with the broader WFIRST effort. The PI will lead the following: (i) Monthly telecons with the Steering Committee, to optimize the team effort; (ii) Regular telecons with all team members, with status updates from the deliverables leads (Fig. ??) so that progress, issues, and solutions are broadly visible across the team. (iii) Yearly face-to-face meetings and more focused working meetings will be added as needed (the associated travel costs have been budgeted). Monthly, the Steering Committee will evaluate progress reports and redistribute responsibilities within the team if necessary. Finally, team members will present recent progress, results, and goals for the upcoming year to an external review panel at this annual meeting. This will also serve as a NASA management review with representatives of NASA HQ and the JPL WFIRST Project Office invited to attend. This follows the successful model used by the US Planck and US Euclid teams with which the PI and many Co-Is have experience. The PI will also set-up wikis and repositories for efficient sharing and discussion of software, documents, plans, progress and issues. We plan to share with the community the software developed for this program. Key results will be published in peer review papers.

    Program Milestones. In Fig. 2 we outline milestones for our effort in conjunction with the project timeline.

    Postdocs and Students. The largest component of our budget is support for post-doctoral researchers who will work under the supervision of the PI and Co-Is to carry out
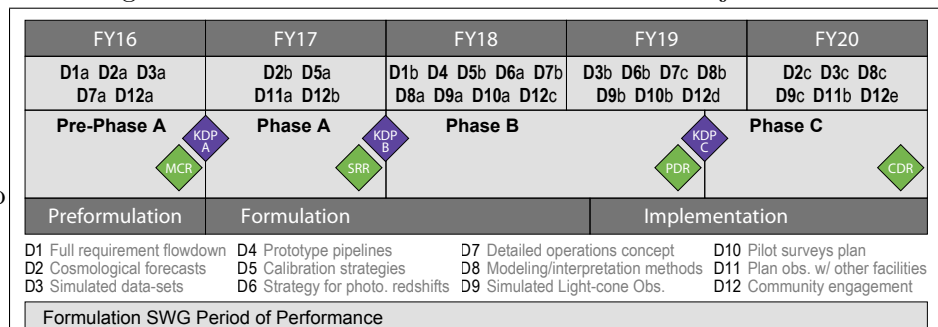

Figure 2: Our deliverable schedule in concordance with the WFIRST project timeline as displayed in the WFIRST SIT call §3.2. Deliverables that are made in multiple stages are labeled (a, b and c) and appear in multiple years.

the many calculations,
simulations, and tests
needed to accomplish
our tasks. Our budget
incorporates an initial
plan for which postdocs will be located at which institutions in which years to work on which tasks.

However, the optimal division of labor may shift over time, and the Steering Committee will consider reallocations of work annually based on internal proposals from team members. Where appropriate, we will consider raising collaborators to funded Co-Is, with NASA approval, if they are best positioned to carry out particular tasks. Many of the methodology development efforts, and some of the technical tasks, are well suited to graduate students, supported by other sources or by this Investigation if their work clearly falls within its scope. In addition to accomplishing the work of this proposal, the involvement of many postdocs and students will build a cadre of young researchers who are ideally positioned to exploit the scientific opportunities of WFIRST.

A Unified Team. As is evident from our previous discussion, jointly addressing the weak lensing and redshift-space clustering elements of the WFIRST dark energy program leads to an ambitious task list that offers critical advantages relative to studying them separately. First is the effective use of expertise; many members of our team have expertise in both of these program elements and will contribute to both during the course of this investigation. Second is the commonality of the hardware; HLS Imaging and Spectroscopy use the same telescope, detectors, and data systems, and it makes sense to develop associated requirements through joint consideration of the two surveys. Third is the need to develop a unified operations concept for these two interleaved surveys; in overall footprint and in details of dither patterns and roll angles, choices made for one survey affect the performance of the other.

Finally, a joint investigation allows us to take a broad perspective on the WFIRST dark energy program, including a common forecasting framework that can account for complementary information and the ability of one survey to mitigate systematics in the other, common sets of cosmological simulations and models of galaxy bias that are relevant to both topics, and investigation of trades between HLS Imaging and Spectroscopy. We have carefully constructed a team capable of addressing all of the essential tasks for elements A and C of the WFIRST SIT NRA, and we have requested resources that will enable us to accomplish this work. If NASA decides to select additional teams in these areas we will be glad to collaborate with them. As emphasized in §7, WFIRST is a national and international resource, and it is valuable to engage as much of the astronomical community as possible in its formulation.

9   List of Acronyms and Abbreviations and References

σ_m	– rms amplitude of matter fluctuations	LSS	– large scale structure
Ω_m	– dimensionless density of the Universe	LSST	– Large Synoptic Survey Telescope
a	– scale-factor of the Universe	NICMOS	– HST Near Infrared Camera and Multi-Object Spectrometer
ACT	– Atacama Cosmology Telescope		
AdvACT	– Advanced ACT	NIR	– near-infrared
AFTA	– Astrophysics Focused Telescope Asset	NRA	– NASA Research Announcement
BAO	– baryon acoustic oscillations	NRC	– National Research Council
BOSS	– Baryon Oscillation Spectroscopic Survey	NWNH	– New Worlds, New Horizons
		P_m	– matter power spectrum
CDR	– critical design review	PFS	– Subaru Prime Focus Spectrograph
CFHT	– Canada-France-Hawaii Telescope	photo-z	– photometric redshift
CGL	– cluster-galaxy lensing	PSF	– point spread function
CHIME	– Canadian Hydrogen Intensity Mapping Experiment	RSD	– redshift-space distortions
		SAM	– semi-analytic galaxy formation models
CL	– galaxy clusters / cluster growth	SDSS	– Sloan Digital Sky Survey
CMB	– cosmic microwave background	SDT	– Science Definition Team
CMB-S4	– CMB stage 4 experiment	SDT13	– 2013 WFIRST SDT report
$D(z)$	– distance-redshift relation	SDT15	– 2015 WFIRST SDT report
$D_A(z)$	– angular-diameter distance	SIT	– Science Investigation Team
DE	– dark energy	SMEX	– NASA Small Explorer
DES	– Dark Energy Survey	SN	– supernovae
DESC	– Dark Energy Science Collaboration	S/N	– signal-to-noise
DESI	– Dark Energy Spectroscopic Instrument	SPHEREx	– Spectrophotometer for the History of the Universe, Epoch of Reionization, and Ices Explorer
ELG	– emission line galaxies		
ELTs	– Extremely Large Telescopes		
eROSITA	– extended Roentgen Survey with an Imaging Telescope Array	SPT-3G	– South Pole Telescope Third-Generation Camera Survey
ETC	– Exposure Time Calculator	STEP	– Shear Testing Program
f_g	– fluctuation growth rate	STIS	HST Space Telescope Imaging Spectrograph
FSWG	– Formulation Science Working Group		
GEO	– geostationary earth orbit	SZ	– Sunyaev-Zeldovich
GGL	– galaxy-galaxy lensing	$w(z)$	– dark energy equation-of-state
GREAT	– Gravitational Lensing Accuracy Test	WFC3	HST Wide-Field Camera 3
GREAT3	– The third GREAT challenge	WFIRST	– Wide-Field Infrared Survey Telescope
GRS	– Galaxy Redshift Survey	WISPs	– HST WFC3 IR Spectroscopic Parallel survey
$H(z)$	– Hubble parameter		
HLS	– High Latitude Survey	WL	– weak lensing
HOD	– halo occupation distribution	WMAP	– Wilkinson Microwave Anisotropy Probe
HSC	– Subaru Hyper Suprime-Cam	WPS	– WFIRST Preparatory Science
HST	– Hubble Space Telescope	WSC	– WFIRST Science Centers
IA	– intrinsic galaxy alignments	z	– redshift
IPAC	– Infrared Processing and Analysis Center	z_p	– pivot redshift
IPC	– inter-pixel capacitance		
IR	– infrared		
JDEM	– Joint Dark Energy Mission		
JWST	– James Webb Space Telescope		
KiDS	– Kilo Degree Survey		
L2	– Lagrange point 2 orbit		
LF	– luminosity function		

We highlighted in bold the team members in the following bibliography.