

Cosmology with the WFIRST High Latitude Survey



Annual Report, May 2017

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 (LBL), 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: Ivano Baronchelli (Caltech/IPAC), Andrew Benson (Carnegie Obs.), Ami Choi (OSU), Katrin Heitmann (ANL), George Helou (IPAC), Shoubaneh Hemmati (IPAC/Caltech), Michael Hudson (Univ. of Waterloo), Albert Izard (JPL), Elisabeth Krause (Stanford), Alexie Leauthaud (UCSC), Niall MacCrann (OSU), Elena Massara (LBL), Alex Merson (Caltech/IPAC), Hironao Miyatake (JPL), Andres Plazas Malagon (JPL/Caltech), Eduardo Rozo (U. of Arizona), Mike Seiffert (JPL), Chaz Shapiro (JPL), Kendrick Smith (Perimeter Insti.), Melanie Simet (UCR/JPL), Masahiro Takada (Univ. of Tokyo), Michael Troxel (OSU), Anja von der Linden (Stony Brook Univ.), Naoki Yoshida (Univ. of Tokyo), Ying Zu (OSU)

TABLE OF CONTENTS

1 Executive Summary [Oli: Olivier, 2 pages]	1
2 Proposed Deliverables [Oli: Olivier, 4 pages]	2
3 Weak Lensing and Cluster Growth Investigation [Oli: Chris, Rachel M, Mike, David W., 15 pages]	4
3.1 Requirements (D1, D3, D4, D5)	4
3.1.1 Polarization effects	6
3.1.2 Interpixel capacitance requirements	6
3.2 Photometric Calibration	6
3.2.1 Dark filter	8
3.2.2 Calibration plan	12
3.2.3 Detector characterization	14
3.3 Photometric Redshifts	16
3.4 Cosmological Forecasting, Simulations, and Methodology Development (D2, D8, D9)	17
3.5 Systematics Testing and Mitigation (D8)	19
4 Galaxy Redshift Survey Investigation [Oli: Yun, Lado, Shirley et al., 15 pages]	20
4.1 Requirements	20
4.2 Simulations	23
4.3 Cosmological Forecasting, Modeling, and Simulations (D2, D8, D9)	23
4.4 Systematics Testing and Mitigation (D8)	25
5 Cosmological Forecasts [Oli: Tim, Elisabeth, Olivier, 10 pages]	26
5.1 HLSS Forecasts	26
5.2 HLIS Forecasts	27
5.3 Shear tomography power spectra	27
5.4 Shear covariances	27
5.5 Expanding the Science Case - Multi-Probe Forecasts	38
6 Operations Model for the HLS and Evaluation of Trades	38
6.1 Snapshot of the HLS observing plan	38
6.2 Further optimizations	39
7 Community Engagement and External Data-sets (D11, D12) [Oli: Olivier, 5 pages]	39
8 Future Workplan [Oli: Olivier, all, 5 pages]	42
9 List of Acronyms and Abbreviations and References	46

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 $T\ 1\text{--}6$ 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 (T_1 , T_3 , T_4).

- (**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.

Below is some new text from Rachel M. It needs to be put in the right place. If more information or detail is needed, please let me know.

3.1.1 Polarization effects

During early 2017, work was carried out to assess the approximate level of an effect that could cause weak lensing systematics, but that had never been previously considered by the weak lensing community. This effect is polarization-dependent quantum efficiency (due to e.g. different reflectivity of various coatings for different polarizations of light). Since the light from edge-on disk galaxies typically has some low level polarization perpendicular to the disk, any polarization-dependence of the QE could result in a preferential selection of such galaxies based on their orientation in the focal plane. This would violate the baseline assumption in a weak lensing analysis, which is that all coherent galaxy alignments are due to gravitational lensing.

A student at CMU, Brent Tan, worked with Rachel Mandelbaum and Chris Hirata on a simple toy model for this effect. The toy model had two parameters: the fraction of the disk galaxy light that is polarized, and the relative attenuation of that perpendicular polarization component (both numbers in the range $[0, 1]$). For each point in that parameter space, the coherent shear due to selection bias was calculated; see results in Figure 1. Finally, the results were modified to account for the fact that not all disk galaxies are viewed edge-on and that not all galaxies are disks, giving a net coherent shear due to this selection bias of $\sim 3 \times 10^{-4}$. The results are still quite uncertain because our fiducial values for the disk polarization fraction were based on observations of nearby galaxies, not $z \sim 1$ disks. However, this is large enough to be relevant for WFIRST, so this systematic needs to be evaluated more carefully and requirements placed in future. A publication on this topic will be prepared during summer 2017.

Another possible polarization-related systematic is a polarization-dependent PSF. That will be the subject of future work.

3.1.2 Interpixel capacitance requirements

The WFIRST detectors will suffer from electrical crosstalk between the pixels, unlike the optical detectors that are based on CCDs. This effect, known as the *interpixel capacitance* (IPC), appears as a systematic effect in the weak lensing shear measurements and causes a bias in the measurements if not properly taken into account. The effect of IPC on the point-spread function (PSF) was already studied by members of our SIT in [?], and requirements were placed on the level of uncertainty in the IPC based on how that uncertainty affects the PSF.

More recently, in late 2016, members of our SIT (Mandelbaum and student Kannawadi) carried out and analyzed simulations to determine whether additional requirements on IPC are needed to ensure that weak lensing shear estimation is not biased beyond our tolerances. To calibrate the shear multiplicative bias to an accuracy of 2×10^{-3} , we find that the requirements on the IPC placed by the PSF requirements are sufficient, so no new requirement is needed. A paper on this result is in preparation.

3.2 Photometric Calibration

[Authors: Nikhil, Chris]

Our SIT has been instrumental in the Calibration Working Group, since precision cosmology measurements depend sensitively on calibration; subtle effects that might not be noticeable in other

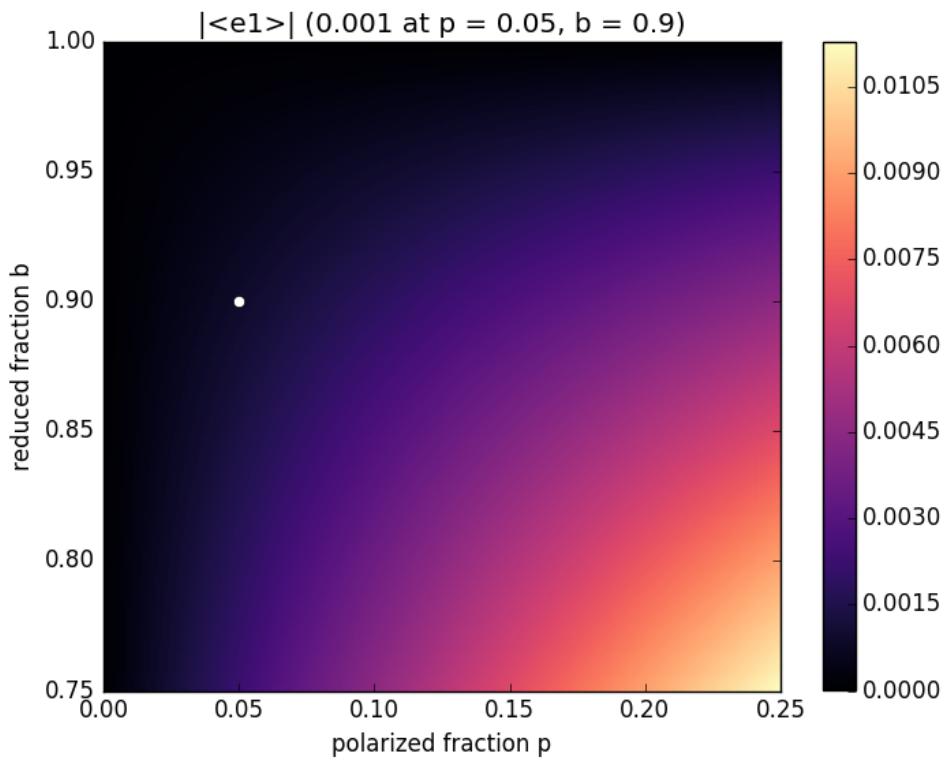


Figure 1: The average shear due to weak lensing selection biases due to polarization-dependent quantum efficiency, as a function of the fraction of polarized light from edge-on disks (horizontal axis) and the fractional attenuation of the perpendicular polarization (vertical axis). The dot at $(0.05, 0.9)$ is our fiducial point in parameter space, and has $\langle e \rangle \approx 0.001$.

areas of astrophysics can become important when trying to measure galaxy shapes to $< 0.1\%$. Activities over the past year have included:

- *Dark filter*: Co-I's Wang, Capak, and Hirata participated extensively in the analyses and discussions that led the FSWG to recommend a dark position in the element wheel on WFIRST.
- *Calibration plan*: Our SIT has contributed extensively to the WFIRST WFI Calibration Plan, including detailed quantitative assessments of calibration approaches and their ability to meet requirements. In some areas, such as dark current and the point spread function, our contributions to the calibration plan are now traceable all the way from science measurements (WL shear) down to the specific calibration approaches and the hardware stability requirements needed for them to work. A major area of work leading up to SRR/MDR is to complete this flow-down for the other areas of calibration.
- *Detector characterization*: We have made use of the H4RG data provided by the Detector Characterization Laboratory to measure some of the non-linear effects relevant to weak lensing in real H4RG detectors. This is an important practice step toward building calibration pipelines that will support WL science.

In what follows, we provide some highlights from our calibration activities. The list is not exhaustive.

3.2.1 Dark filter

In the summer and fall of 2016, the FSWG was tasked with determining whether a dark filter was needed for WFIRST calibration. This required the FSWG to enumerate the list of calibration tests that might use the dark filter, and establish whether alternative options were possible. We led the effort to assemble this list of tests based on input from the SITs (both ours and others), the SOC, and Project personnel. The list¹ included 14 items: (i) the dark current (including internal instrument backgrounds); (ii) unstable pixels; (iii) post-reset transients; (iv) read noise correlations; (v) inter-pixel capacitance; (vi) gain measurement; (vii) the high spatial frequency flat; (viii) the low spatial frequency flat; (ix) persistence from previous observations; (x) persistence from slews; (xi) classical linearity; (xii) count rate dependent non-linearity; (xiii) the brighter-fatter effect; and (xiv) persistence re-activation.

The problem of persistence from slews (i.e. streaks across the detector following a slew from one observation to another) is of particular importance to weak lensing, because it leads to a coherent, highly directional pattern on the detector that has the correct symmetry to induce a coherent systematic error in the galaxy ellipticities. This is a concern without a dark capability, or even with a dark capability if it is not (or cannot be) used during every slew. Our group identified two budgets in WL that flow down into slew persistence requirements. First is the total systematic shear error budget of 2.7×10^{-4} . Second is the masked pixel budget.

The details of the slew persistence study are provided in the Calibration Plan. It consisted of several stages: first, assessing the magnitude distribution of the stars that would be encountered in the High Latitude Survey; then assessing the probability of stimulus levels in a slew, given the distribution of slews from our operations model (§6); and then folding this through a persistence model (based on DCL data for the development H4RG detectors) to predict the probability distribution of persistent pixels in the HLS imaging survey. The stimulus distribution (x in e: the well depth to which a pixel is filled during a slew) from the Calibration Plan is shown in Figure 2, and the persistence signal distribution (y in e: the persistence signal in a pixel over the course of an exposure) is shown in Figure 3.

¹DarkAlternativesMatrix_161030.docx

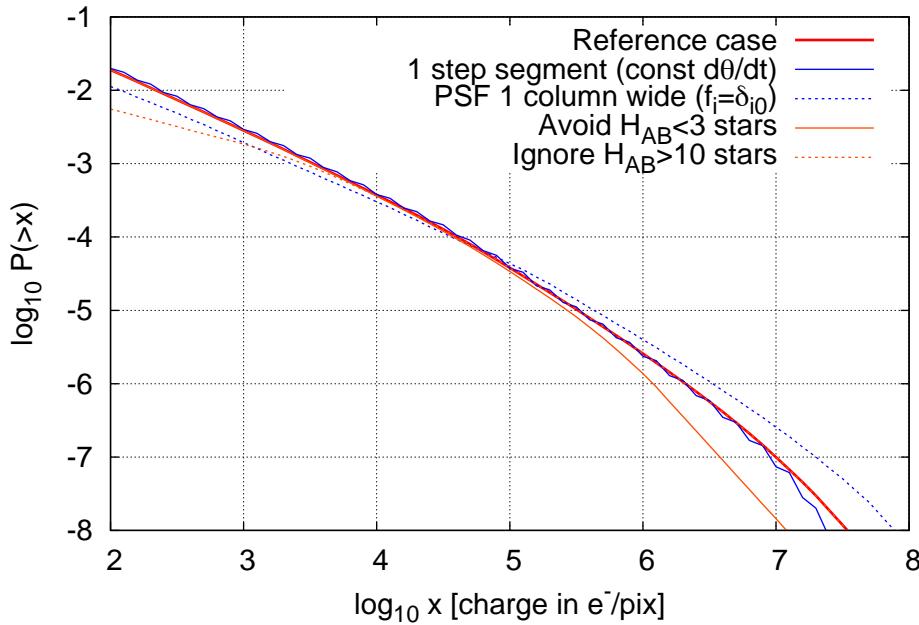


Figure 2: Comparison of stimulus levels predicted under different assumptions and approximations. The vertical axis shows the log probability to exceed a given stimulus level during a slew of 0.4 degrees (a step along the short axis of the field, executed frequently during the HLS). The thick red line indicates reference assumptions. The solid blue line treats the slews as being at constant $\dot{\theta}$. The dashed blue line approximates the PSF as 1 column wide (all the flux from the star is concentrated in the central column). The orange lines show what happens if bright ($H_{AB} < 3$) or faint ($H_{AB} > 10$) stars are excluded from the model.

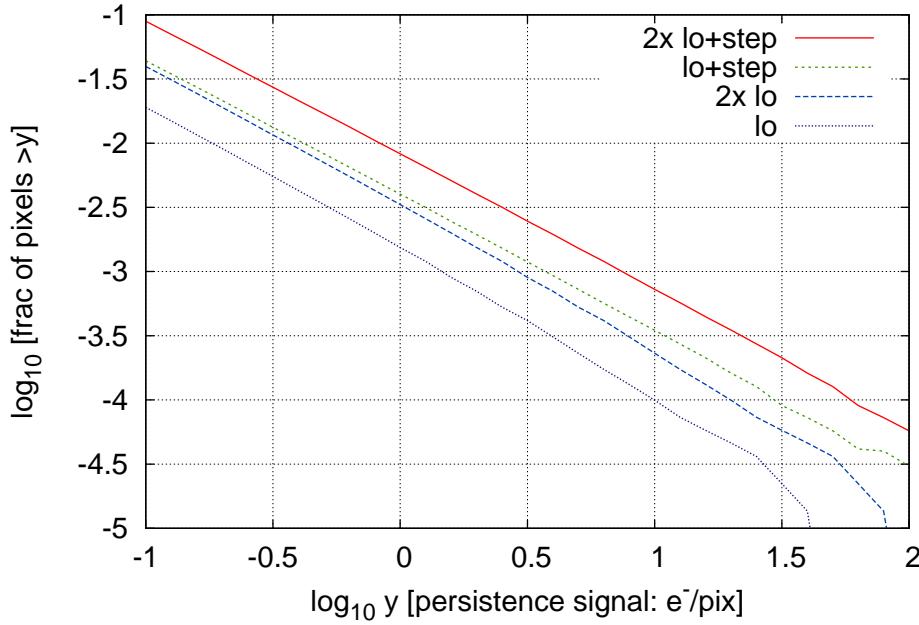


Figure 3: The cumulative distribution function of slew-induced persistence in the HLS imaging survey, $P(>y)$. The persistence signal y is estimated in electrons per pixel; the decaying persistence curve is integrated over a 160 s. Several persistence models are shown, including the “lo” case (typical of the development detectors), and a “lo+step” case (including an order of magnitude step at saturation, as seen in portions of some detectors). This figure has not been updated yet to go from the 6-year to the 5-year observing plan, although we expect only minor differences.

After negotiating with the Project, we settled on a mitigation strategy for slew persistence that involved saving the spacecraft orientation information from the Attitude Control System (ACS), using this to predict the locations of persistence from bright star streaks, and masking $\pm 2\sigma$ on either side of these streaks. Unmasked streaks are simply accepted as part of the systematic error budget. Their impact on shape measurement is based on an analytic result derived by our SIT and tested against Monte Carlo simulations:

$$\Delta\gamma_1 + i\Delta\gamma_2 = \frac{M\Omega_{\max}\sigma_n^2 R^4}{2F^2 N_{\text{ind}} \text{Res}} f_{\text{scale}} f_{\text{aniso}}, \quad (1)$$

where $\Delta\gamma_{1,2}$ are the two components of spurious shear; M is a margin factor; $\Omega_{\max} = 421.3$; σ_n^2 is the variance of the persistence image; R is the radius of the galaxy in pixels; F is the signal from the galaxy in electrons per exposure; N_{ind} is the number of *independent* exposures of the galaxy²; Res is the galaxy resolution factor [?]; f_{scale} and f_{aniso} are factors ≤ 1 describing the scale dependence and anisotropy of the persistence power spectrum (defined to be 1 in the worst case).

The results of this study – shown in Figure 4 – are promising, given the top-level systematic shear budget of 2.7×10^{-4} and that the modern detectors typically show “lo” or (in some regions) “lo+step”-like behavior, rather than the much larger persistence characteristic of the WFC3-IR model (third column). The masking algorithm will continue to be revisited as part of the mission optimization. However, the small number of masked pixels led the FSWG to conclude that a dark shutter that operated during every slew was not required for the WFIRST HLS.

We carried out a related study, also using Eq. (1) and related machinery, to assess how well we need to know the dark current for WFIRST. Dark current measurements without a dark filter are possible, e.g. via median algorithms that combine many exposures from a survey, but are subject to: (i) a degeneracy in which the “true” sky brightness is unknown and hence the zero level of the dark current cannot be established, and (ii) possible correlated errors from imprinted celestial sources. The requirements, as derived in the appendix to the calibration plan, are:

- The error in the dark current + bias determination in a 140 s HLS imaging exposure shall be no more than $0.0096 f_{\text{corr}}^{-1/2}$ e/p/s (uncorrelated part) or $0.0017 f_{\text{corr}}^{-1/2}$ e/p/s (imprinted celestial sources).
- The error in the dark current + bias determination in a 297 s HLS spectroscopy exposure shall be no more than $0.0059 f_{\text{corr}}^{-1/2}$ e/p/s (uncorrelated part) or $0.00072 f_{\text{corr}}^{-1/2}$ e/p/s (imprinted celestial sources).

Here “ f_{corr} ” denotes the factor by which we plan to correct biases induced by errors in the dark current map (we normally choose $f_{\text{corr}} = 1$ to be conservative). The requirements are traceable to additive shear biases from non-circular imprinted celestial sources; multiplicative shear biases as the noise in the dark current map results in e.g. galaxy centroids getting “pulled” toward pixels whose measured dark current fluctuates below the true dark current of that pixel; and Eddington-like biases for sources detected in the GRS. While the semi-analytic estimates in the calibration plan based on source counts suggest that the HLS imaging requirement can be met without a dark filter, our SIT and the Calibration Working Group had concerns about possible degeneracies in the self-calibration procedure that can only be addressed by a detailed simulation. Moreover, the approach requires empty space in the images, which we will not have in the case of grism spectroscopy. As the imaging exposures are shorter than the spectroscopy exposures, this would require dedicated long imaging exposures (of HLS spectroscopy exposure length) just for the purpose of self-calibrating

²This may be less than the total number of exposures of the galaxy, since slew persistence from successive exposures will be correlated.

Masked pixels & systematic shear results for ACS-based flagging

Example parameters:

- Mask trails of $H_{AB} < 9$ stars, $\pm 2\sigma$ on either side of track, if expected persistence is > 8 e.
- For fainter stars, assume masking of science data for > 64 e outliers (4.3σ).

	H4RG-lo	H4RG-lo + step		WFC3-IR		
ACS error (arcsec rms per axis)	Masked pixel fraction	Systematic shear per component	Masked pixel fraction	Systematic shear per component	Masked pixel fraction	Systematic shear per component
1.0	0.16%	3.4E-5	0.68%	8.6E-5	2.85%	6.6E-4
2.0	0.32%	3.4E-5	1.31%	8.6E-5	5.50%	6.6E-4
4.0	0.62%	3.4E-5	2.58%	8.6E-5	10.78%	6.6E-4

Note that further tuning of the parameters may yield some improvement, or enable different trades between masked pixels and systematics in the unmasked data.

Figure 4: The outcome of the October 2016 slew persistence study. This shows the masked pixel fraction and the predicted systematic shear due to unmasked streaks as a function of both the persistence model and the accuracy of pointing information.

the dark. Due to sky Poisson noise, we would need many of these images – our February 2017 estimate was for $N = 73$ exposures, which, if done every week, would consume 4% of the wall clock time. In light of these and other issues, the Calibration Working Group recommended that WFIRST maintain the dark filter.

3.2.2 Calibration plan

Our SIT has contributed extensively to the WFIRST WFI Calibration Plan. This includes extensive quantitative analysis of proposed calibration techniques, as detailed in the appendix to the plan. Some highlights follow.

The requirement on knowledge of the dark current and the calibration approaches are fully defined, based on analysis done during the dark filter trade (October 2016 – February 2017).

Weak lensing was found to place demanding requirements on measurement of the count rate-dependent non-linearity (CRNL). The weak lensing program is sensitive to CRNL because it enhances the bright center of a PSF star relative to its wings, thereby making the star appear slightly smaller, but does not have a similar effect on the faint galaxies used for shape measurement. The PSF second moment is biased by a factor of $1 - \alpha$ (where α is the CRNL exponent), and has a top-line systematic error budget of 7.2×10^{-4} . This means that if α is measured to $\pm 3 \times 10^{-4}$ (the requirement from the supernova SITs), then CRNL consumes 17% of the PSF size error budget, in an RSS sense. Given that CRNL is a pernicious bias for two of the dark energy probes, we recommended a multi-faceted approach to CRNL calibration, including a lamp-on/lamp-off capability for WFIRST (this was not available on WFC3-IR).

Our team has revisited the wavefront stability requirements for weak lensing, using a set of

codes and scripts on the team’s GitHub site. This begins with a Fisher matrix analysis of the uncertainties in the shear power spectrum, and our top-line requirement that the systematic errors be equivalent to the statistical errors even if the survey is extended to 10,000 deg² (i.e. in an RSS sense, the systematic errors should be 20% of the statistical errors in the nominal 2,000 deg² survey). Requirements are assessed using the significance, defined by

$$Z = \sqrt{\Delta\mathbf{C} \cdot \boldsymbol{\Sigma}^{-1} \Delta\mathbf{C}}, \quad (2)$$

which is the number of sigmas at which one could distinguish the correct power spectrum from the power spectrum containing a systematic error. We built sub-allocations for multiplicative (shear calibration) errors, and for additive (spurious shear) errors in each angular bin. An early discovery was that this process depends on the redshift dependence of the shear error: some redshift dependences are “worse” than others by the Z -metric. The worst possibility is *not* for the error to be redshift-independent, but rather for it to change sign, as this can mimic a change in redshift evolution of the growth of structure.

In our current formalism, for each angular template, we introduce a limiting amplitude $A_0^{\text{flat}}(\alpha)$, defined to be the RMS spurious shear per component A_0 at which we would saturate the requirement on $Z(\alpha)$ for angular bin α in the case of a redshift-independent systematic $w_i = 1 \forall i$ (here α denotes an angular bin and i a redshift bin). That is, if the additive systematics did not depend on redshift, we could tolerate a total additive systematic shear of A_0^{flat} (RMS per component) in band α . We also introduce a scaling factor $S[\mathbf{w}, \alpha]$ for a systematic error

$$S[\mathbf{w}, \alpha] = \frac{Z(\alpha) \text{ for this } w_i}{Z(\alpha) \text{ for all } w_i = 1} \quad (3)$$

that depends on the redshift dependence w_i . An additive systematic error that is independent of redshift will have $S = 1$. A systematic that is “made worse” by its redshift dependence will have $S > 1$, and a systematic that is “made less serious” by its redshift dependence will have $S < 1$. The requirement that the (linear) sum of Z s not exceed $Z(\alpha)$ thus translates into

$$\sum_{\text{systematics}} [A(\alpha)]^2 \times S[\mathbf{w}, \alpha] \leq [A_0^{\text{flat}}(\alpha)]^2, \quad (4)$$

where $A(\alpha)$ is the RMS additive shear per component due to that systematic. We take the “reference” additive shear to be the additive shear in the most contaminated redshift slice; in this case, $w_i = 1$ for that slice, and $|w_i| \leq 1$ for the others. Under such circumstances, we can determine a *worst-case scaling factor* $S_{\max, \pm}(\alpha)$, which is the largest value of $S[\mathbf{w}, \alpha]$ for any weights satisfying the above inequality. We may also determine a worst-case scaling factor $S_{\max,+}(\alpha)$ conditioned on $0 \leq w_i \leq 1$, i.e. for sources of additive shear that have the same sign in all redshift bins. In most cases, however, something is known about the redshift dependence of the systematic error (e.g. for PSF errors the error scales with the size of the galaxy, and hence has a redshift dependence tied to the measured redshift evolution of galaxy sizes). In these cases, we use the correct redshift weighting factor S . This approach has been critical in order to set stability requirements that are consistent with the Project’s integrated modeling results.

We have begun incorporating the HLS observing strategy (§6) in studies of self-calibration of time-dependent drifts in the response of the system (i.e. time dependence of the conversion from μJy on the sky to DN/s in the digitized detector system outputs). This model is in a state of flux as we add parameters to it, but here we show a current snapshot allowing for time-dependent drifts of the response of each of the 18 SCAs making up the focal plane, with time dependence parameterized in calibration periods of Δt (assessed down to a period of 3 hours) each. Both individual-SCA drifts and common-mode drifts are allowed, with an assumed intrinsic variation (calibration prior) of 1%

RMS drift in each e -fold of timescales. A network of randomly distributed stars with a density of 500 stars/deg² and $S/N = 50$ was assumed; in self-calibration, the magnitudes of these stars are *not* known a priori, but are assumed to be stable across multiple repeated observations of the same field. These are preliminary parameters being used to test our tools and are not currently held as requirements. The stellar density model is very conservative since the Trilegal model predicts star counts of 572, 803, 990, and 1137 stars/deg² at $H_{AB} = 18 - 19$, $19 - 20$, $20 - 21$, and $21 - 22$ at the SGP, and even an $H_{AB} = 22$ star will have $S/N > 50$. The temporal stability of the system needs further study and will be varied as an input parameter in future versions of this model. The current model uses the April 19, 2017 update to the HLS observing strategy. The number of calibration parameters varies depending on the filter, since there are no parameters for periods of time when the instrument is not observing in that filter; the current version has 17262 parameters for the H band.

Despite the intrinsic stability assumed, in which each SCA can have its response fluctuate by 1.67% RMS from one time interval to the next, the repeated observations do an excellent job of tracking these changes and reducing the posterior uncertainty. Even for $\Delta t = 0.125$ days = 3 hours, the posterior calibration errors are at the level of 0.14–0.17% RMS (here “RMS” is weighted by number of observations), depending on the filter. An example of the model output (predicted uncertainties in the calibration parameters for each SCA at each epoch) is shown in Figure 5. It must be remembered that this analysis is overly simplistic in some ways – particularly that we have not yet allowed for shorter-timescale variations (i.e. on timescales $< \Delta t$), nor have we allowed for separate gain drifts among the different readout channels. These will have to be included in a future version of the model. On the other hand, the stellar density and S/N assumptions were extremely conservative (e.g. the full range of stellar magnitudes 18–22 should have 7 times more stars than were assumed, even at the Galactic pole), so there is margin to absorb these additional degrees of freedom. The next iteration of the model for time-dependent calibration drifts will include additional parameters, as well as updated priors reflecting expected detector system stability rather than the place-holder requirements shown here.

3.2.3 Detector characterization

The WFIRST dark energy analyses will place enormous demands on our understanding of the detectors. Some aspects of this problem can be anticipated in advance – for example, we know that effects such as inter-pixel capacitance, count-rate-dependent non-linearity, etc. will need to be carefully characterized, and we are working as part of the Calibration Working Group to build these measurements into the mission. However, with systematic error budgets at the level of a few $\times 10^{-4}$, it is likely that WFIRST analyses will turn up new effects that were not apparent in past missions. Therefore a key task for our SIT is to analyze the data from development detectors and identify these new effects early enough to inform the calibration plan.

In ground-based weak lensing projects using thick CCDs (e.g. DES), one of the key detector issues has been the *brighter-fatter effect* (BFE). This is an electrostatic effect in which as a pixel fills up with collected charge, it changes the electric field geometry and new charges generated are more likely to be deflected into neighboring pixels. This has the effect of making bright stars appear larger than faint stars, as the repulsion effect is non-linear and increases with signal level. The field geometry is very different in a NIR detector, but a brighter-fatter effect is still possible.

We have searched for the brighter-fatter effect in the H4RG detector arrays using the flat fields for two devices H4RG-17940 and H4RG-18237, provided to us by the DCL. The BFE imprints a signature in the auto-correlation function of a flat field; using the correlations in multiple non-destructive reads in a flat field, one can separate linear IPC from the BFE. Preliminary brighter-fatter effect results for H4RG-17940 are shown in Figure 6. The BFE coefficients are $a'_{\Delta i, \Delta j}$, which is the fractional change in effective area of pixel (i, j) when an electron is placed in pixel

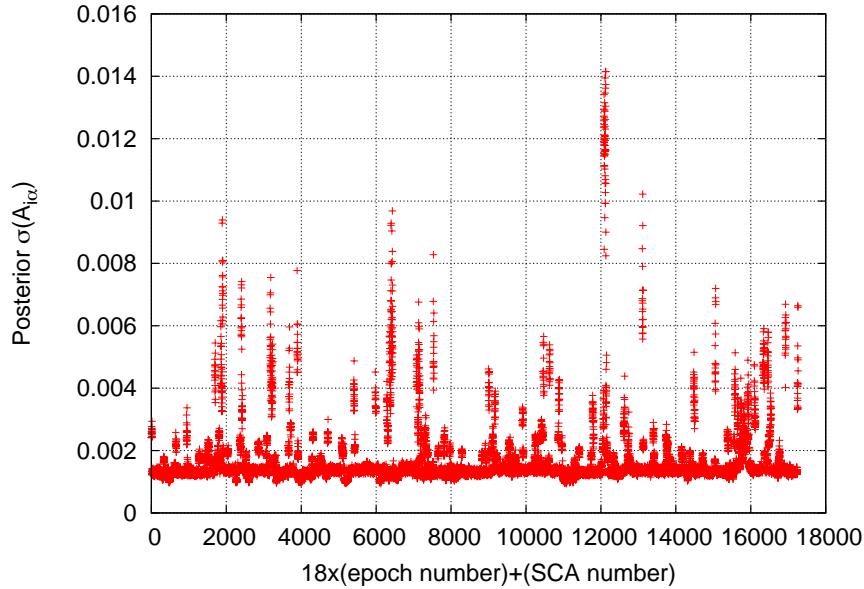


Figure 5: An example of the posterior calibration error from an HLS self-calibration calculation. The horizontal scale displays the time intervals (left-to-right in time order), with 18 points per time interval indicating the various SCAs. The vertical axis shows the standard deviation of the calibration solution for that SCA at that time, $\sigma(A_{i\alpha})$, relative to the survey mean. The figure shows the case of H band with $\Delta t = 0.125$ days. This example had 17262 calibration parameters. A few epochs, mostly containing only a few observations, are poorly constrained due to minimal overlap observations. It is subject to refinement and the input parameters of the model will be varied as we work toward requirements on the stability of the detector system.

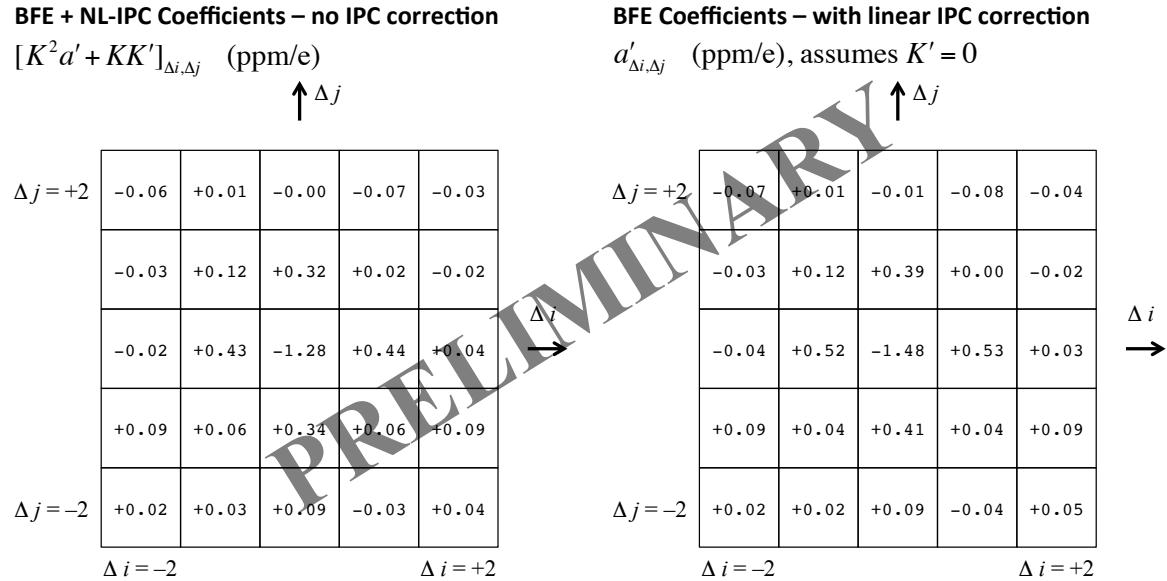


Figure 6: The BFE + NL-IPC coefficients $[K^2 a' + KK']_{\Delta i, \Delta j}$ (left panel) and IPC-corrected coefficients $a'_{\Delta i, \Delta j}$ (right panel), for H4RG-17940. Note that the IPC-corrected coefficients assume that the IPC is linear, i.e. the non-overlapping correlations are ascribable entirely to the BFE and not NL-IPC. The 1σ uncertainty in each pixel is 0.07 ppm/e.

$(i + \Delta i, j + \Delta j)$; they have units of parts per million per electron (ppm/e). The flat auto-correlations are sensitive to both the brighter-fatter effect and non-linear inter-pixel capacitance (NL-IPC); we are currently working on distinguishing the two effects.

3.3 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.

3.4 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.

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

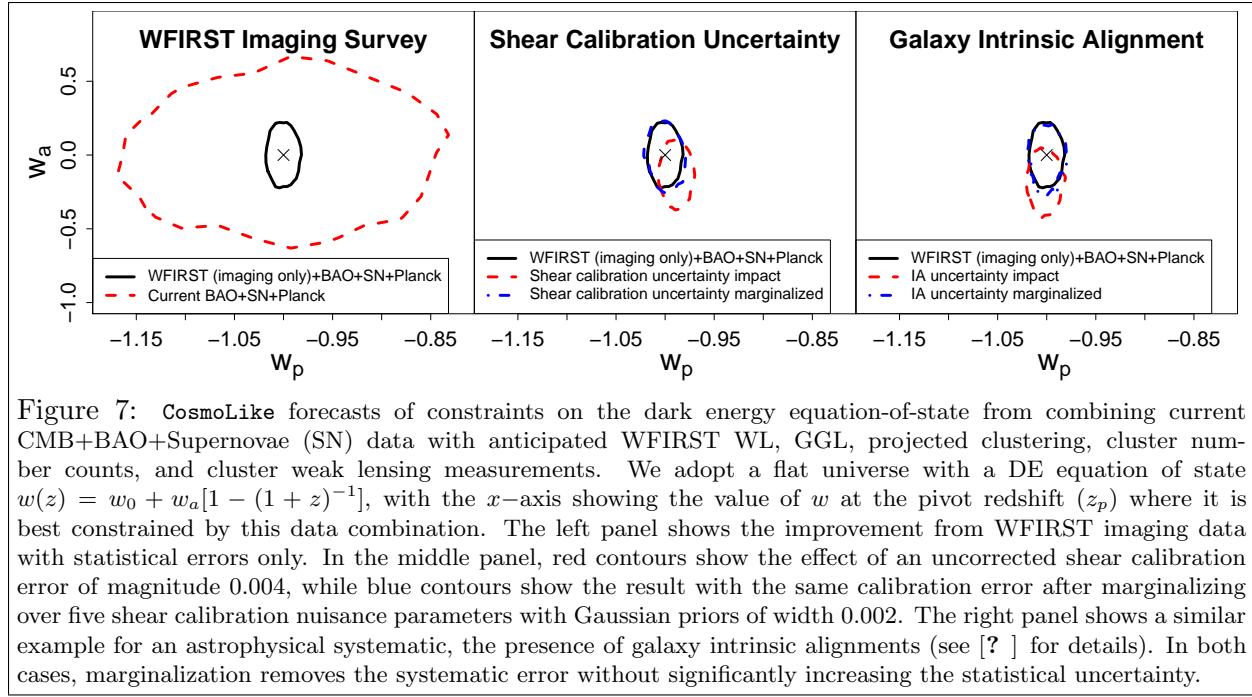


Figure 7: *CosmoLike* forecasts of constraints on the dark energy equation-of-state from combining current CMB+BAO+Supernovae (SN) data with anticipated WFIRST WL, GGL, projected clustering, cluster number counts, and cluster weak lensing measurements. We adopt a flat universe with a DE equation of state $w(z) = w_0 + w_a[1 - (1 + z)^{-1}]$, 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.

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 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).

3.5 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]

With true corners
Test

4.1 Requirements (D1)

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. Accordingly, we have made this task our main focus over the past year. Our SIT made three deliveries of updates to the WFIRST GRS requirements to the WFIRST Project on July 1, 2016, December 1, 2016, and March 2, 2017. These provide progressively precise definitions of the GRS requirements. We describe the main requirements and their science drivers below. This work was carried out by Wang, Eifler, Hirata, Ho, Kiessling, Krause, Merson, Padmanabhan, Pearson, Samushia, Benson, Capak, Doré, Heitmann, Spergel, Teplitz, and Weinberg.

4.1.1 Science Requirements (Level 2a)

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 should be contiguous and large enough to reduce edge effects in the BAO/RSD measurements. The $> 90\%$ overlap with the HLIS enables joint analysis of 90% of WL and GRS data, which maximizes the dark energy science from WFIRST. Imaging also provides undispersed galaxy positions, improving redshift determination. The statistical precision of the dark energy constraints is sensitive to the survey area as well as the survey depth; a trade study of depth

versus area will need to be carried out to optimize both, in the context of Euclid and LSST. We also need to investigate the impact of dividing the area into two equal patches near the NEP and SEP respectively, to take advantage of potential ground-based telescope resources. A survey of one or two large, contiguous areas has smaller edge effects and better window functions than a survey comprised of many smaller areas.

We have carried out trade studies of the HLSS survey design. We note that it will be important to conducting these trade studies in the context of the joint science return of HLSS and HLIS that properly accounts for correlations among spectroscopic and imaging observables and accounts for their correlated systematics. We are in the progress of implementing a corresponding forecasting effort. Here, we have carried out a trade study of area versus depth for the HLSS only, starting from a baseline survey of 2227 deg^2 and a wavelength range of 1.05-1.85 microns. We consider two alternative scenarios, i.e. a survey twice as wide and shallower and a survey half as wide but correspondingly deeper. The galaxy redshift distributions were computed using the WFIRST Exposure Time Calculator ETC v14. The $\text{H}\alpha$ forecasts are based on the average of the 3 models in Pozzetti et al. (2016), and the [O III] forecasts are based on the Mehta et al. (2015) luminosity function. We extend the CosmoLike framework (Eifler et al 2014, Krause & Eifler 2016) to compute the constraining power of all scenarios on cosmic acceleration, closely following Wang et al (2013). We run 500,000 step MCMC simulated likelihood analysis in a 23 dimensional parameter space. We simultaneously vary 7 cosmological parameters and 16 “nuisance” parameters describing uncertainties due to the linear galaxy bias model, the non-linear smearing of the BAO feature, peculiar velocity dispersion, power spectrum shot noise, and redshift errors. We assume priors on cosmological parameters from the current state of the art experiments, i.e. the Planck mission, the Baryon Oscillation Spectroscopic Survey (BOSS), the Joint Lightcurve Analysis (JLA) supernovae, as described in Aubourg et al (2015).

The information gain is quantified using the standard Dark Energy Task Force FOM and an extended cosmology FOM, which measures the enclosed volume in the full 7-dimensional cosmological parameter space, not just in the 2 dark energy parameters. We will refer to these FOMs as DE-FOM and Cosmo-FOM. Compared to the baseline survey, we find a decreased DE-FOM of 32% and a decreased Cosmo-FOM of 45% for the shallow/large area survey. For the deep/small area survey we find an increased DE-FOM of 5% and an increased Cosmo-FOM of 2%. While our trade study validates the design of the baseline survey, we note that these findings are model and prior dependent and will carry out further studies varying the input parameters. In particular, the [OIII] galaxy number density will be updated pending inclusion of the results from the latest observational data from HST grism observations.

We also investigated whether the survey area needs to be contiguous. We constructed two identical sets of Gaussian simulations one set covering contiguous 2000 deg^2 and a second set consisting of two 1000 deg^2 disjoint fields. The BAO signal was then measured in the 2D power spectrum using the most recent techniques applied to the BOSS DR12 data. The BAO positions measured in disjoint fields were biased by 1% on average compared to the contiguous field in both line-of-sight and transverse directions. This bias persists even after properly correcting for the window effects and is unlikely to be coming from the sample variance since our sets consisted of close to one thousand independent simulations. This bias could be a result of either bigger than the box-size modes or various edge effects. The window of the real data will be more involved than we considered in our test case and the biases may be larger. This investigation is ongoing but our preliminary results seem to support the conclusion that a contiguous area is preferable for the standard BAO analysis.

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 $n P_{0.2} \sim 1$ at $z = 1.6$, with 20% margin. Requiring $n P_{0.2} \sim 1$ 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 $z = 1.8$. Given the Hirata

forecast of H α ELG counts (Model 3 in Pozzetti et al. 2016), $nP_{0.2} \sim 0.6$ at $z = 1.8$, and $nP_{0.2} > 2$ at $z = 1.3$. 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, $z = 1.6$, at which it is impossible for a ground-based survey to obtain spectra for a large number of galaxies. There remain large uncertainties in the H α LF due to the limited availability of uniform data. It is likely that the actual number of H α ELGs is higher than assumed here; thus we have additional margins for this requirement. We have assumed a bias for H α ELGs of $b(z) = 1 + 0.5z$. The bias relation has been rescaled to agree with Geach et al. (2012) measurement of $b = 2.4$ at $z = 2.23$ for $f > 5 \times 10^{-17}$ erg s $^{-1}$ cm $^{-2}$.

This is significantly deeper than the Euclid GRS survey, which ensures that the WFIRST GRS 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. This number density requirement could in principle be met using either H α or [OIII] ELGs, depending on the survey strategy. There is no need to set a separate requirement for [OIII] ELGs; this depth ensures high number densities for both [OIII] and H α ELGs.

Galaxy number density is a key input in the dark energy Figure-of-Merit. It is very sensitive to the H α LF, which still has large uncertainties but will become better determined as more data become available and more comprehensive analyses are done. The flow down of the galaxy number density requirement here to the minimum survey depth depends on the LF of ELGs. Co-I Teplitz is a key member of the WISP team. He is supervising a postdoc, Ivano Baronchelli, in deriving more precise LFs for H α and [OIII] ELGs using WISP data.

HLSS 3: The wavelength range of the HLSS will allow measurement of H α emission line redshifts over the redshift range $1.1 < z < 1.9$.

The corresponding wavelength range is 1.38 μm to 1.9 μm . This wavelength coverage also allows measurements of [OIII] emission line redshifts over the range $1.8 < z < 2.8$. A wider wavelength range that allows H α emission line detection over a wider redshift range is desirable, as it increases the survey volume and therefore adds margin for meeting other baseline requirements. It is also critical that the WFIRST GRS redshift range is complementary to that of Euclid, with its red cutoff at 1.85 μm , or $z < 1.8$.

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 GRS can reach $z \sim 1$ without great difficulty, thus we should focus on $z > 1$. Euclid GRS can only reach $z \sim 2$; its shallow depth does not enable a high enough number density of observed [OIII] ELGs. WFIRST GRS is deep enough to observe both H α (656.3nm) and [OIII] (500nm) ELGs, with the number density of the latter sensitive to the survey depth (the deeper the survey the higher their number density).

For the nominal wavelength range of 1-2 microns, WFIRST GRS covers $0.52 < z < 2$ using H α ELGs, and $1 < z < 3$ using [OIII] ELGs. Thus the redshift range requirement is met including both types of ELGs.

In addition to the trade studies in HLSS 1 we examine the impact of an extended wavelength range on the DE-FOM and the Cosmo-FOM. We follow the same methodology as detailed in the HLSS 1 description in extending the wavelength range from 1.05-1.85 microns for the baseline model to 1.00-1.89 for the extended model. We find a decreased DE-FOM of 2% and a decreased Cosmo-FOM of 11% for the extended wavelength survey with respect to our baseline scenario. While the FOM trade study seems to favor a narrower redshift range, we emphasize that these findings are model and prior dependent and will conduct further studies varying the input parameters. The reduction in the telescope temperature to 260K will have a major impact on this trade study. In addition, the FoM comparison is quantitative but simplistic; it does not reflect how the various future surveys will complement each other. Euclid GRS covers the wavelength range of 0.92-1.85 microns using the same BAO/RSD tracers as WFIRST, thus there is unique scientific value in

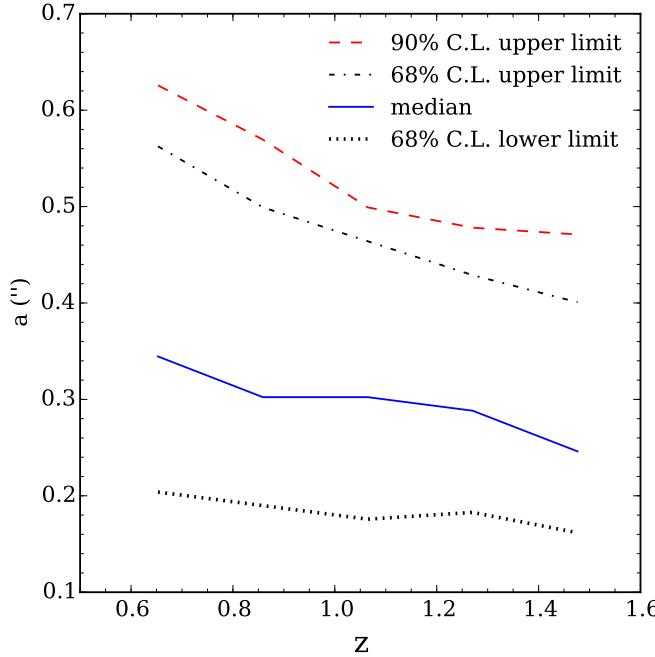


Figure 8: The 90% upper limit of the semi-major axis continuum size of H α galaxies with H α line flux $> 10^{-16}$ erg/s/cm 2 , $0.55 < z < 1.85$), based on 1773 galaxies from WISP (WISP team, private communication).

WFIRST having a wavelength cutoff longer than 1.85 microns.

HLSS 4: Redshift measurement errors σ_z shall satisfy $\sigma_z < 0.001(1 + z)$, excluding outliers, for galaxies smaller than 0.54'' in radius. The fraction of outliers with $|z_{\text{obs}} - z_{\text{true}}|/(1 + z_{\text{true}}) > 0.003$ shall be less than 10%.

This is a requirement on the rms error of the redshift measurements, and not on every redshift measurement.

We justify the requirement on the knowledge of the outlier fraction as follows. If you add a contaminant to the galaxy power spectrum with contamination fraction α , then you leak in an amount of power from the “wrong” line with amplitude α^2 , but you dilute the power spectrum of the “real” signal by a factor of $(1 - \alpha)^2$. For BAO the dilution is a minor issue (it reduces S/N), but for RSD it is a problem because it reduces the galaxy bias by $(1 - \alpha)$ without changing the linear redshift-space distortion parameter $f\sigma_8$. So your inferred rate of growth of structure $f\sigma_8(z)$ is reduced by a factor of $1 - \alpha$. This leads to a stringent requirement on knowledge of α – if it is 9.8% but you think it is 10% you have a 0.2% systematic error, which is a reasonable budget for this contribution.

Larger size galaxies have larger redshift errors. To assess redshift accuracy, a realistic pixel level grism simulation covering at least 2 square degrees needs to be carried out and processed. The current data from the HST grism survey WISP finds that 90% of galaxies that would be observed by WFIRST (H α flux $> 10^{-16}$ erg/s/cm 2 , $0.55 < z < 1.85$) have a size less than 0.54'' (semi-major axis continuum size). Since the maximum redshift of the WISP sample is ~ 1.6 , WFIRST H α ELGs will likely have smaller sizes on average, see Figure ??.

The [OIII] ELGs are more compact, so this requirement is also sufficient for the redshift precision for [OIII] ELGs.

HLSS 5: Relative position measurement uncertainties shall be less than 3.4'' over the entire survey area.

We need to measure galaxy positions to better than $\sim 0.1\text{Mpc}/h$ (which corresponds to 3.4''),

assuming that 105 Mpc/h subtends 1 degree), in order to measure galaxy clustering accurately. This should be met easily if HLIS 26 is met, which makes systematic errors in the astrometry negligible. 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 50% (TBC), and the redshift purity shall be 90% (i.e., the outlier fraction is less than 10%). Completeness is defined as the fraction of H α 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 H α ELG number counts predicted by the H α LF to the galaxy number density that can be used to measure BAO/RSD by WFIRST GRS. The completeness of 50% and redshift purity of 90% are put in as crude estimates based on extrapolations from Euclid. Since WFIRST has a higher spatial and spectral resolution compared to Euclid, and more rolls (4 versus 3) per field, we expect a higher completeness and purity for WFIRST. The actual requirements will need to be validated by grism simulations, since these are determined by what are feasible given the instrumentation and the true universe. The requirement on the knowledge on the contamination fraction is set by HLSS 4.

4.1.2 Implementation Requirements (Level 2b)

HLSS 7: The observatory shall provide a slitless spectroscopy mode, with a spectral dispersion no larger than 10.85 Å/pix.

Gratings tend to give constant dispersion in linear space, rather than R-space. The above dispersion would give point-source spectral resolution $R = \lambda/\Delta\lambda$ in the range $550 < R < 800$, for a 2-pixel resolution element.

The grism resolution requirement is set by requiring the redshift precision to be 0.1% (set by BAO/RSD science), thus is not sensitive to which ELGs (H α vs [OIII]) we use as tracers. Going to lower spectral resolution would degrade the redshift precision and put BAO/RSD science goals at risk. The number density of [OIII] ELGs may be significantly higher than previously assumed; this gives some margin in the spectral resolution requirement due to the smaller sizes and less line blending of [OIII] ELGs.

Given the margin from a likely higher [OIII] ELG number density than previously assumed, we have removed the requirement on resolving H α and NII for all galaxies of radius 0.3'', and 90% of galaxies of radius 0.54'', which would drive the grism resolution higher. The blending of H α (6563Å) and NII (6584Å) leads to a metallicity-dependent shift in line centroid for larger sources this would lead to a systematic bias in the measured redshifts, which can propagate into the BAO/RSD measurements. Having a higher grism resolution would alleviate this problem, at the cost of a reduction in survey depth, and more overlapping of spectra for galaxies. This is a trade study that we will carry out as the required grism simulations become available.

HLSS 8: Spectra shall achieve $S/N \geq 5$ for $r_{\text{eff}} = 300\text{mas}$ for an emission line flux $1.0 \times 10^{-16} \text{ erg/cm}^2/\text{s}$, from a source at 1.8 μm.

This sensitivity is sufficient to meet the comoving space density requirement HLSS 2 with some margin given best estimates of the H α luminosity function (Pozzetti et al. 2016) at these redshifts. The use of a $S/N \geq 5$ threshold for an arbitrary spectrum pre-spectral-decontamination gives margin for detection of sources whose spectra overlap others, or for loss of some exposures to cosmic ray hits or other artifacts, as $S/N \geq 5$ post-decontamination is expected to be sufficient for meeting the redshift accuracy requirement HLSS 4, and the post-decontamination S/N should be significantly higher than the pre-decontamination S/N for a given spectrum. Current calculations of observatory performance indicate that the sensitivity specified here is achieved in a total exposure time of ~ 1200 seconds per field.

The median continuum size (semi-major axis) of H α ELGs is 0.3'' (see Figure ??). This sensitivity requirement is phrased in parallel with the sensitivity requirement of the WL survey. This depth is a factor of two to three deeper than the Euclid GRS. The depth is sufficient to give the required galaxy number density in HLSS 2.

HLSS 9: The uncertainty of the wavelength measurement λ shall satisfy $\Delta\lambda/\lambda \leq 0.001$.

Although this is redundant since it is essentially the same as HLSS 4, it is necessary to keep it since it flows HLSS 4 into a dataset requirement.

HLSS 10: The spectroscopic bandpass shall satify $\lambda_{\text{max}} \geq 1.9 \mu\text{m}$, and $\lambda_{\text{max}}/\lambda_{\text{min}} < 1.82$.

We need $\lambda_{\text{max}} > 1.9 \mu\text{m}$ for redshift reach, in order to be complementary to Euclid and ground-based surveys. Furthermore, we need $\lambda_{\text{max}}/\lambda_{\text{min}} < 1.82$ for line identification using multiple lines: to ensure that we cannot have [OII] (373nm) falling off the blue end of our coverage while H α (656.28nm) falls off the red end. We have assumed that the actual bandpass extends 1.5% from either end, since it is problematic to use emission lines that fall within 1.5% of the bandpass edges.

HLSS 11: 50% of the energy (excluding diffraction spikes and non-1st order light) shall be enclosed in a circle of radius $< 0.21''$ over 95% of the field.

This limit of 0.21'' is required by source separation in the input catalog for spectral extraction, and is enabled by the addition of the phase mask corrector, and leaves some margin on the wavefront error.

HLSS 12: The filter used to define the bandpass of the grism shall have cutoff transition widths $\sigma < 1\%$ (0.7% goal) after including the effects of broadening by the range of incident ray angles at each position in the FoV, where σ is defined by $\sigma = (\lambda(T = 0.90) - \lambda(T = 0.10))/\lambda(T = 0.50)$. T is the transmission of the grism bandpass.

This is based on the grism guiding considerations; the assessment of grism guiding (and its positive outcome) assumed $\sigma < 1\%$.

4.1.3 Implementation (Operations Concept) Requirements

We did not update Requirements HLSS 13-15, but include them here for completeness. Co-I Hirata is the co-lead for the WFIRST Operations Working Group.

HLSS 13: Exposures of each field shall be obtained at a minimum of 3 dispersion directions, with two being nearly opposed.

HLSS 14: The observatory shall be able to place the WFC at a commanded orientation with an accuracy of 0.64'' (3 σ) in pitch and yaw, and 87'' (3 σ) in roll (TBR these were arbitrary values that give a net 3 σ position uncertainty of 10 pixels. For the HLSS, the primary driver is that the position uncertainty is small with respect to chip gaps, which gives larger uncertainties than specified above. The smaller values quoted here are consistent with efficient target acquisitions, which would flow down from an observing efficiency spec.)

HLSS 15: The observatory pointing jitter and drift shall not exceed 100 mas in the spectral direction on the WFC focal plane (goal of 60 mas) and 50 mas in the cross-dispersion direction (TBR).

HLSS 16: Imaging observations shall be obtained of the fields in the HLSS that reach JAB=24.0, HAB=23.5, and F184AB=23.1 for an $r_{\text{eff}} = 0.3''$ source at 10 σ to achieve a reference image position, in 3 filters.

Provided the HLSS covers area already observed in the HLIS, this requirement will be met automatically. This requirement applies to any HLSS fields that are counted toward the minimum survey area requirement but are not covered by the HLIS. Imaging in at least three filters is required to build a minimal spectral template for grism spectral decontamination.

HLSS 17: There shall be 40 observations of two deep fields, each 11 deg² in area,

sufficient to characterize the completeness and purity of the overall galaxy redshift sample. The 40 observations repeat the HLSS observing sequence of 4 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 degrees.

To calibrate the HLS GRS, we need a spectroscopic subsample, with the same selection criteria as that of the HLS GRS, containing more than 160,000 galaxies that have a redshift purity > 99%. We need 160,000 galaxies to know the redshift purity to 1% (which requires 10,000 objects, assuming noise of $1/\sqrt{2N}$ from Poisson statistics) in at least four categories (low z, high z, faint, luminous).

Based on the estimated galaxy number density of > 7273 per deg² at the flux limit for the GRS, 10^{-16} erg s⁻¹ cm⁻², we need a total area for the deep fields of $160,000/7273=22$ deg². These can be split into two subfields of 11 deg² each. Smaller subfields prevent the testing of galaxy clustering statistics in each subfield. Each deep field should be part of the HLS footprint, so they are representative of the GRS as a whole.

The visits to the deep field should consist of 10 sets of HLS-GRS-like visits, matching the integration time, dither pattern, and observational time-sequence of the HLS-GRS strategy, with each set of HLS-GRS-like visits covering the same areas of 22 deg². Assuming a completeness of 50% and uncorrelated sets, the completeness after 10 sets of visits is $(1-0.5)10=0.001$, leading to a 99.9% complete sample for calibrating the GRS. Since each set of observation consists of 4 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 degrees, in order to map out possible sources of systematic errors due to inhomogeneity.

HLSS 18: The observing efficiency of the HLSS, defined as the total science exposure time divided by the total time allocated to the survey, shall be TBD%.

The total time includes slew, settle, target acquisition, and calibration observations that are specific to the HLSS, including the extra-depth observations of the deep fields described in HLSS 17. This minimum observing efficiency, together with a 0.67 year total allocation of observing time, allows science exposures of 1600 deg² (TBC) with the exposure time indicated in the comment to HLSS 8; this provides a 7% (TBC) margin over the 1500 deg² requirement (HLSS 1) to allow for data that may be unusable because of instrumental artifacts, bright sky objects, etc. This is a high level requirement that will need to be revisited as the mission implementation details become more solid; it should be set such that the core science goals for the GRS are achieved without putting mission success at risk.

4.1.4 Calibration Requirements

HLSS 19: The relative spectrophotometric flux calibration shall be known to 2 percent 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 spectroscopic sample for the GRS is expected to be defined by a line flux limit of 10^{-16} erg s⁻¹ 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.

We start by setting a requirement on the spatial uniformity of the mean number density as a function of physical scale. We require that the non-cosmological fluctuations in the mean number

density (or the selection function of the survey) be $< 1\%$ (sqrt variance) when averaged over spatial scales between $10 \text{ Mpc}/h$ to $200 \text{ Mpc}/h$. At small scales, this is \sim two orders of magnitude smaller than the cosmological signal, while at the \sim BAO scale of $100 \text{ Mpc}/h$, this is \sim one order of magnitude smaller than the cosmological signal. These fluctuations equal the cosmological signal at $\sim 400 \text{ Mpc}/h$. These physical scales correspond to ~ 0.5 degrees to 6 degrees at a redshift of 1.5.

We convert the above requirement to a requirement on the spectrophotometric calibration accuracy, assuming the Model I luminosity function of Pozzetti et al. At the flux limit of WFIRST, this yields a requirement of 1% relative spectrophotometric calibration, averaged over angular scales of 0.5 degrees to 6 degrees.

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.

We plan to make this requirement more precise, in the form of "The relative spectrophotometric flux shall be known to 2% relative accuracy in TBD (probably the spectral resolution) wavelength bins with a goal of 1% on scales larger than TBD (per pointing, 0.3 deg) and TBD% on scales smaller than 0.3 deg. We are working on deriving and justifying these numbers. Co-I's Capak, Hirata, and Padmanabhan are members of the WFIRST Calibration Working Group, working on a detailed calibration strategy for WFIRST.

HLSS 20: The uncertainty in the wavelength calibration shall not introduce biases in the wavelength measurement by amounts greater than $\Delta\lambda/\lambda = 10^{-4}$ on any angular scales exceeding 0.064 degrees within a field, and $\Delta\lambda/\lambda = 2 \times 10^{-5}$ from field to field.

Variations in the wavelength calibration within a field, and from field to field on large scales, wash out the clustering signal by de-correlating the projected component of the clustering signal on those angular scales.

Within a field, the acceptable level of wavelength error is $\Delta\lambda/\lambda \sim 10^{-4}$, which is 10% of the errors on individual redshift measurements (0.001), to avoid increasing the overall redshift error by a significant factor. The angular scale is set by the optimal smoothing scale for BAO reconstruction, $\sim 5 \text{ Mpc}/h$. At $z = 3$, this subtends 0.064 degrees for a flat universe with $\Omega_m = 0.3$ and a cosmological constant.

For field to field, the acceptable level of wavelength error is 2×10^{-5} , which comes from comparing two adjacent fields. Since we expect ~ 104 galaxies per deg^2 , we have ~ 2810 galaxies per FOV of 0.281 deg^2 . If the galaxies have a redshift error of 10^{-3} each, then one can measure systematic offsets between fields (statistically) at the $10^{-3}/\sqrt{2810}$ level, which is 1.9×10^{-5} . At that level the power from the systematics is sub-dominant to the power from the redshift error.

4.1.5 Requirements on Science Data Products:

We are in the process of studying HLSS 21-25. These depend on the structure and responsibilities of the SOCs and the SITs. Co-Is Teplitz and Capak have extensive experience in data processing for space missions, and have provided detailed comments on these requirements to the WFIRST Project Office.

HLSS 21: The raw data for each grism exposure shall be available through the archive, with each dataset including identifying information such as time of exposure, observatory pointing orientation, a unique dataset identifier, and any engineering information needed for subsequent processing. Each detector readout for a given exposure shall be included in the dataset.

HLSS 22: Calibrated data for each grism exposure shall be available through the archive. Each detector readout shall be calibrated at the appropriate level, and the individual calibrated readouts will be combined to produce a net spectral image.

These datasets shall include information on the effective PSF as a function of position and incorporate any World Coordinate System information needed for subsequent stages of processing. As sources are not yet identified, association of a pixel with a source position and wavelength is not yet possible.

HLSS 23: Source catalogs of the same field derived from WFC imaging data shall be combined with observatory pointing information for each grism exposure to produce a segmentation map that associates each catalog source with a range of spectral image pixels. The spectral images of bright stars in each detector shall be used to refine the astrometric solution. These segmentation maps shall be used to extract 1D spectra for each source, and to flag pixels that may contain flux from multiple sources. The extracted spectra shall include information on the effective exposure time for each pixel, effective PSF as a function of position, data quality flags, and any other information needed to interpret the data.

HLSS 24: Extracted spectra of each source from multiple roll angles shall be combined to produce a single net spectrum of each source. For sources that are spatially resolved, the result shall be provided as a data cube of position and wavelength. The spectra obtained at nearly opposing roll angles shall be used to account for possible offsets of the emitting region from the center of the broad-band image. The data from all roll angles shall be used, to the extent possible, to resolve ambiguities in the proper source to associate with pixels illuminated by overlapping spectra. These net spectra shall include information on the effective exposure time for each pixel, statistical and systematic uncertainties in the measured fluxes and wavelengths, effective PSF as a function of position, data quality flags, and any other information needed to interpret the data.

HLSS 25: The data processing system shall have the capability of inserting fake sources into the spectral image data and re-executing the generation of high-level science products. These tests are essential for verifying the proper operation of the tools that generate high level science products and for understanding the sensitivity of the survey and systematic effects that may be present in the survey sample.

HLSS 26: 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 degrees, and less than 20% on larger angular scales.

This requirement is only meaningful in terms of the contribution to the total error budget by the uncertainties in the 3D selection function. The BAO scale is less than 2 degrees in the redshift range for the HLSS.

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 due to effects of target selection, data reduction and observing conditions. Previous surveys were able to separate the selection function in two independent parts: the radial selection function and the angular selection function. It is likely that the WFIRST selection function will not be separable in this way, i.e. different parts of the sky will have different radial profiles. For now we will assume that this type of separation is possible. This assumption is reasonable for preliminary investigation since 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 catalogues. When computing clustering statistics, the random catalogues remove the systematic effects of varying mean number density (due to target selection, data reduction or observing conditions).

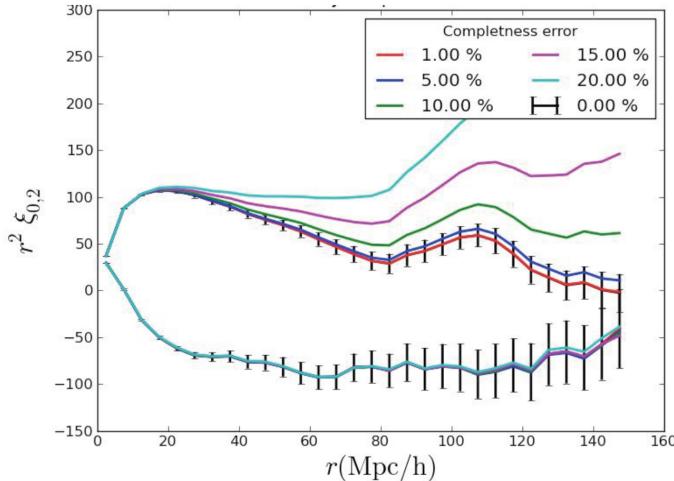


Figure 9: The monopole (upper curves) and quadrupole (lower curves) of a single mock survey and how they respond to various completeness errors in the scenario described in scenario. We can see from the level of two point correlation function, even a 5% error on the completeness can change the monopole and quadrupole significantly that we expect RSD to be affected, while BAO is not significantly affected, as this mostly changes the amplitude of the correlation function.

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 WFIRST data will vary pixel to pixel on the infrared 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 $3 h^{-1}\text{Mpc}$ 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 TBD per cent of statistical errors between the scales of 10 and 150 $h^{-1}\text{Mpc}$ in clustering signals (either in correlation function multipoles or power-?spectrum). For the galaxy sample expected from WFIRST, this corresponds to TBD per cent uncertainty in the knowledge of the radial distribution and the angular mask.

To further quantify the effect of systematics offset in the angular mask on clustering measurements we have performed tests on mock catalogues representing BOSS CMASS sample. This is justified by the fact that the BAO and growth rate measurements from WFIRST GRS in redshift bins of $z \sim 0.1$ are expected to be roughly equal to the CMASS constraints with $z \sim 0.2$.

The mock surveys are generated from N-body simulations, with a median redshift of 0.6, with galaxies of halo mass range about $7 \times 10^{13} M_\odot$. The mock surveys have proper BOSS 3D selection function (which we take as truth here). Now we distort the selection function in the following scenarios:

(a) The survey region is divided into two equal area along RA and one area has true completeness whereas the other one has $1 - x$ completeness where x is a given completeness error (as shown in the legend of Figure ??). We show in Figure ?? the resulting monopole and quadrupole with varying completeness error. This is an interesting limiting case as surveys can sometimes be affected by large scale systematics generated by either calibration of two parts of the sky, or due to large scale effects caused by the galactic foregrounds.

(b) The true survey completeness is multiplied with a gaussian function. The gaussian function has mean 0 and variance as the denoted completeness error. We then fold both positive and

negative side of the Gaussian to the negative side and hence allow the completeness to be only smaller than its true value. We vary the scale at which we change the 3D selection function, starting with 1 degree to 4 degrees (4 deg is approximately the BAO scale at this redshift).

From the preliminary analysis shown here, we expect that we will need to accurately model the 3D completeness function down to a few % level. At 5% the effects can already be very detrimental to our large scale structure analyses using BAO and RSD. Scaling from these, we arrive at the requirement that artificial correlations due to inaccuracies in the 3D selection function are less than 10% of the statistical error bars on scales smaller than 2 degrees, and less than 20% on larger angular scales.

4.1.6 Requirement on Cosmological Volume Simulations:

This is a new category of requirements for a space mission. These are not needed for mission success (data acquisition by the spacecraft), but for meeting the high level science requirements (level 1). We summarize these as follows:

- (a) a few accurate mocks with galaxies included using semi-analytical galaxy formation model, to verify and validate WFIRST GRS pipeline;
- (b) ~ 100 mocks with high mass resolution of 109 solar masses, to inform theoretical modeling of the data;
- (c) $\sim 10,000$ mocks with low resolution, to derive the covariance matrices for the WFIRST data.

To quantify the requirement on cosmological volume simulations, we consider only the case for galaxy clustering science (which includes BAO and RSD) for now for simplicity. For WFIRST, simulations are required for the following three objectives: a. establishing the basic correctness of the pipeline; b. informing the theoretical modeling of small scale clustering as a function of tracer properties; c. calculating the covariance matrix for each GRS probe and across many probes.

In order to establish the basic correctness of the WFIRST pipelines and predictions, sophisticated synthetic mock galaxy catalogs are essential. These catalogs, which must realistically emulate WFIRST both in sky area and depth, are typically constructed by running large gravity-only simulations and then “painting” realistic galaxies on top. Populating a simulation with galaxies can be done in several ways: (i) empirically, using statistics such as the “halo occupation distribution”, (ii) by placing the normal, baryonic matter in the simulation ab initio and explicitly solving the hydro-dynamical equations, or (iii) by using a semi-analytical galaxy formation model (SAM), whereby the astrophysical processes and formation histories of galaxies are described using physically motivated, parameterized equations. The advantage of SAMs over alternative methods is their ability to meet the demands from next generation cosmological surveys for large (suites of) galaxy mock catalogues that are both accurate and can be constructed rapidly. In contrast, full hydro-dynamical simulations are far too slow and empirical methods are limited by the availability of existing high redshift observations, which are necessary for the calibration of these methods. SAMs also require some observations for calibration but, once tuned to fit observations at low redshift, they are able to make predictions out to high redshift without the need for further observational input. Furthermore, empirical methods are often limited in that they are calibrated in one or two photometric bands, whilst SAMs are designed to model the star formation history of a galaxy and so have the ability to make predictions for a wide variety of multi-wavelength data simultaneously. This feature of SAMs is vital to ensure that we can examine cross-correlations between the spectroscopically-selected dataset for galaxy clustering analysis and the photometrically-selected dataset for weak lensing analysis. Besides testing the pipeline and making (limited) cosmological

forecasts, these galaxy mock catalogs would also be a valuable resource for science working groups focusing on legacy science (e.g. galaxy evolution, active galactic nuclei). Note that, compared to the large number of approximate mock catalogs necessary for covariance estimation, only very few accurate galaxy mocks are required to verify and validate the WFIRST pipeline.

To inform the theoretical modeling of clustering especially at non-linear scales as a function of the tracer properties would require a significant number of simulations that have relatively realistic modeling of the tracer properties at the relevant redshift. For WFIRST, we can take the current number density of emission line galaxies (for H α galaxies only) from our baseline calculation and used the Tinker et al. 2008 halo mass function, along with the Giocoli et al. (2008) subhalo mass functions to compute the total number of halos and subhalos above some mass threshold and then match that to the baseline GRS number densities. This maps back to approximately 1012 solar masses from $z=1$ to $z=2$. Assuming that we need to have at least 100 particles to resolve halos at 1012 solar masses, and another factor of 10 particles to resolve properties of the halo progenitors, we will need dark matter particle mass resolution of approximately 109 solar masses. The extra factor of 10 is due to the galaxy formation model that depends on the properties of the progenitors which is an approximation that may change as we understand the galaxy properties better and as more observations of the tracers arrive. We expect to require of order 100 simulations to reduce the shot noise of the correlation function in order to compare the theoretical modeling to the simulated correlation function. These realistic mock surveys may also require the modeling of non-standard cosmological models, such as extensions to non-zero total neutrino masses, or modified gravity models.

Finally, we will need to calculate the covariance matrices of the main probes of clustering, namely BAO and RSD, and the cross-covariances among these probes (or across different methods as in recent BOSS analyses). We can approach the calculation of the covariance matrices through multiple avenues. One can generate (in principle) a large number of approximate mock surveys using relatively fast approximate methods (eg. PTHalos, QPM, FastPM, etc), and apply the relevant survey properties onto these mock surveys. The small scale modeling of the clustering may not be 100% accurate, but is likely to be adequate for the linear RSD modeling and BAO analyses where medium to large scales are most important. The number of approximate simulations required can be on the order of $O(10,000)$ depending on the number of parameters we will be estimating using these covariance matrices, but the time requirement of these approximate mocks is relatively modest. One can also envision using more theoretical approaches (such as O'Connell et al. 2015, Padmanabhan et al. 2015), which only require a relatively modest number of realistic mock surveys which are required for (b).

4.2 Simulations

The KSU group has produced a suit of few thousand fast “enhanced log-normal simulations” for the WFIRST GRS expected samples. While these simulations do not correctly reproduce the small scale structure and higher order statistics of the field, they can be used for studying various large scale effects and implement light-cone effects. The simulations have so far been used to study the effect of splitting the WFIRST footprint into two non-contiguous areas. We plan to use this simulations in the future to study systematic effects in the measurements (e.g. window effect correction) and to validate the BAO/RSD proto-pipeline. These simulations are very well-suited for such tasks since their input two-point signal is known exactly.

4.3 Cosmological Forecasting and Data Analysis Algorithms

The key dark energy constraints from the WFIRST GRS will result from the BAO and RSD measurements from the two-point statistics of the observed galaxy field. Similar measurements from the higher order statistics are weaker and currently are considered less robust. Cosmologi-

cal constraints from higher order statistics scale very steeply with the number density and since WFIRST GRS will provide very dense galaxy samples they may significantly enhance the yield from the standard two-point BAO/RSD analysis. We also expect the methods of analysing higher order statistics to become more robust and standardized by the time of WFIRST launch. Because of this considerations it would be helpful to have a higher order statistics forecasting tool. We have developed software package to make forecasts on basic dark energy parameters from higher order statistics. The main assumptions are similar to the ones made in the standard power-spectrum forecasting tool used for baseline WFIRST predictions. We will work on integrating this software with the standard forecasting tool developed by our SIT. While the key design decisions will still be based on the two-point statistics forecasts, knowing how different choices will effect higher order analysis will be very informative.

The KSU group has assembled a fast and lightweight set of tools for analysing the WFIRST GRS data. Currently this toolset starts from the redshift catalogue and the visibility cube and produces the measurements of power spectrum multipoles. The multipoles are then analyzed to extract the BAO and RSD signal from them. The BAO extraction algorithms replicate the analysis of the final BOSS DR12 sample. The RSD analysis is currently simplistic and uses the linear model. We will update this toolset by implementing more realistic RSD models. The toolset will eventually be linked to the redshift catalogue and visibility cube producing software. This software will provide the backbone of our BAO/RSD proto-pipeline and will be validated with high fidelity WFIRST simulations.

[Oli: SWITCHED HERE]

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.

4.4 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 (§3.1), 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 Teplitz 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 H α 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 H α luminosity function (LF) measurement. Through WISPs, Teplitz is also working to obtain significantly more HST data to improve the H α 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 H α 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.

4.5 Simulations

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

4.6 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 H α 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 Ly α 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 §??).

Cosmological simulations. The cosmological simulations described in §3.4 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 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 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 semi-analytical galaxy formation model [? ?] 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.

$H\alpha$ emitter number density forecasts. In work led by Merson, Wang, Benson, Kiessling and Rhodes the open source semi-analytical galaxy formation model, GALACTICUS [? ?], was used to predict the $H\alpha$ -emitter number counts and redshift distributions for the WFIRST GRS.

A four square degree lightcone catalogue was constructed by processing the dark matter merger trees of the Millennium Simulation [?] with the GALACTICUS model. Emission lines are modelled in GALACTICUS by interpolating over a library of emission line luminosities obtained from the CLOUDY [?] code and stored as a function of hydrogen (HI), helium (HeI) and oxygen (OII) ionising luminosities, as well as the hydrogen gas density and metallicity of the interstellar medium (ISM). The emission line luminosities are then processed to incorporate attenuation due to interstellar dust, which can be modelled using several different methods. Merson and collaborators consider three dust methods from Ferrara *et al.* (1999) [?], Charlot & Fall (2000) [?] and Calzetti *et al.* (2000) [?]. However, it is worth noting that any user-specified dust method can be used in conjunction with GALACTICUS.

First, the GALACTICUS predictions for the cumulative counts of $H\alpha$ -emitting galaxies over the redshift range $0.7 \leq z \leq 1.5$ obtained with each dust method are compared with the latest WISP counts from Mehta *et al.* (2015) [?]. The $H\alpha$ luminosities from GALACTICUS are corrected to introduce contamination due to NII by assuming that NII contributes 29% to the observed emission line luminosity. A fixed NII contamination was chosen to be consistent with the WISP analyses of Colbert *et al.* (2013) [?] and Mehta *et al.* [?]. Adjusting the strength of the attenuation in two of these dust models, the Charlot & Fall [?] method and the Calzetti *et al.* [?] dust law, leads to GALACTICUS yielding cumulative $H\alpha$ flux counts that are consistent with the observed

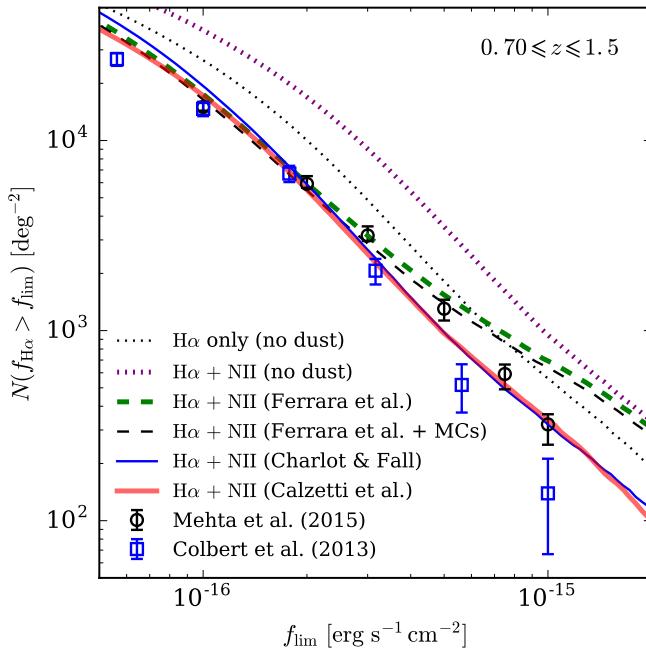


Figure 10: Predictions for the cumulative H α flux counts for the redshift range $0.7 < z < 1.5$ from a GALACTICUS lightcone mock catalogue. The various lines show the predictions for the pure H α fluxes and the H α fluxes when blended with NII (H α + NII), assuming either no dust attenuation or attenuation using one of three methods: Ferrara *et al.* (1999) [?], Charlot & Fall (2000) [?] or Calzetti *et al.* (2000) [?]. To be consistent with the observed counts from WISP [? ?], which are shown by the data points, contamination from NII is assumed to contribute 29 per cent to the galaxy luminosity.

Table 1: GALACTICUS predictions for the cumulative number of H α -emitting galaxies for the WFIRST GRS, assuming a redshift range of $0.5 \leq z \leq 2$. The upper half of the table shows the counts for NII contaminated fluxes, whilst the lower half of the table shows the counts when NII contamination is removed. Predicted counts are reported for three dust methods: Calzetti *et al.* (2000) [?], Charlot & Fall (2000) [?] and Ferrara *et al.* (1999) [?]. For the Ferrara *et al.* method we include attenuation due to molecular clouds (MCs). The strength of the attenuation for the various methods are the same as in Fig ???. Note that the flux limits assume contamination by NII, which is set at 29% of the total flux, as adopted by Colbert *et al.* (2013) [?]. The efficiency of each survey is instrumentation dependent, and has not been included.

Flux limit [erg s $^{-1}$ cm $^{-2}$]	Cumulative number of H α -emitting galaxies [deg $^{-2}$]		
	Calzetti <i>et al.</i> (2000)	Charlot & Fall (2000)	Ferrara <i>et al.</i> (1999, inc. MCs)
CUMULATIVE COUNTS INCLUDING NII CONTAMINATION			
1×10^{-16}	26344	28958	26730
2×10^{-16}	9169	10439	10571
3×10^{-16}	4675	5378	6013
CUMULATIVE COUNTS WITH NII REMOVED			
1×10^{-16}	15833	17660	16938
2×10^{-16}	5202	5971	6557
3×10^{-16}	2628	2978	3751

counts of Mehta *et al.* [?] down to a flux limit of 1×10^{-16} erg s $^{-1}$ cm $^{-2}$ (see Figure ??). Adopting the Ferrara *et al.* [?] dust method yields counts that agree with the observations for flux limits fainter than approximately 5×10^{-16} erg s $^{-1}$ cm $^{-2}$, but predict an excess in the number counts for the brightest fluxes. Overall, this work shows that down to the expected flux limits of the WFIRST mission, the GALACTICUS model is able to predict number densities that are consistent with existing observations from the WISP survey. Although the lightcone used in this analysis has a small area on the sky, it is expected that cosmic variance has little effect on the predicted number counts. Since the lightcone analysis is time-consuming and expensive in computing resources, the lightcone used in this study was limited in size to 4 sq deg, in order to provide timely input to WFIRST. However, significantly larger lightcones will be built in the future.

As a consistency check, the authors show that GALACTICUS, in combination with the revised dust attenuation strengths, is consistent with the observational estimates for the distribution of optical depths (at the observer-frame H α wavelength) as well as the H α line luminosity function, though at $z > 1.5$ the luminosity function from GALACTICUS progressively a worse fit to the observations. This could suggest that the dust methods may be lacking some redshift evolution or dependence on other galaxy properties, or that the GALACTICUS emission line luminosities are the incorrect strength. Investigating these possibilities requires rigorous calibration of the GALACTICUS model, which will be carried out in future work.

Finally, the GALACTICUS lightcone is used to present predictions for the redshift distribution and the differential and cumulative H α flux counts for the WFIRST GRS, as well as two surveys mimicking a Euclid-like selection. These predicted cumulative flux counts to are compared to forecasts from Mehta *et al.* [?] and empirical models originally presented by Pozzetti *et al.* (2016) [?]. The GALACTICUS forecasts have counts that are consistent with those of Pozzetti *et al.* [?] and about 30 per cent lower than those of Mehta *et al.* [?]. The deficit compared to the Mehta *et al.* forecasts could be due to Mehta *et al.* using the OIII line to extrapolate the number of H α emitters for $z \gtrsim 1.6$. For a WFIRST GRS with redshift range $0.5 \leq z \leq 2$ and flux limit of 1×10^{-16} erg s $^{-1}$ cm $^{-2}$ GALACTICUS predicts a number density between 26,300 and 29,000 galaxies

per square degree, prior to removal of NII contamination, and 29 per cent lower with NII removed. Note that all the H α -emitter counts, which are shown in Table ?? are expected number counts of target galaxies for spectroscopy, and *not* the counts of galaxies with redshift measurements. The latter will depend on the redshift purity and completeness for each survey, which in turn depends on instrumentation and noise parameters.

Future work in this area will involve exploiting the GALACTICUS model to examine a variety of other properties of emission line galaxies, including the distribution of OIII luminosities fluxes and the contamination from NII.

4.7 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 internal 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]

5.1 HLSS Forecasts

We have carried out a trade study of area versus depth for the HLSS only, starting from a baseline survey of 2227 deg² and a wavelength range of 1.05-1.85 microns. We consider two alternative scenarios, i.e. a survey twice as wide and shallower and a survey half as wide but correspondingly deeper.

The galaxy redshift distributions were computed using the WFIRST Exposure Time Calculator ETC v14. The H-alpha forecasts are based on the average of the 3 models in Pozzetti et al. (2016), and the [O III] forecasts are based on the Mehta et al. (2015) luminosity function. The resulting redshift distributions are visualized in Fig. 1. We extend the CosmoLike framework

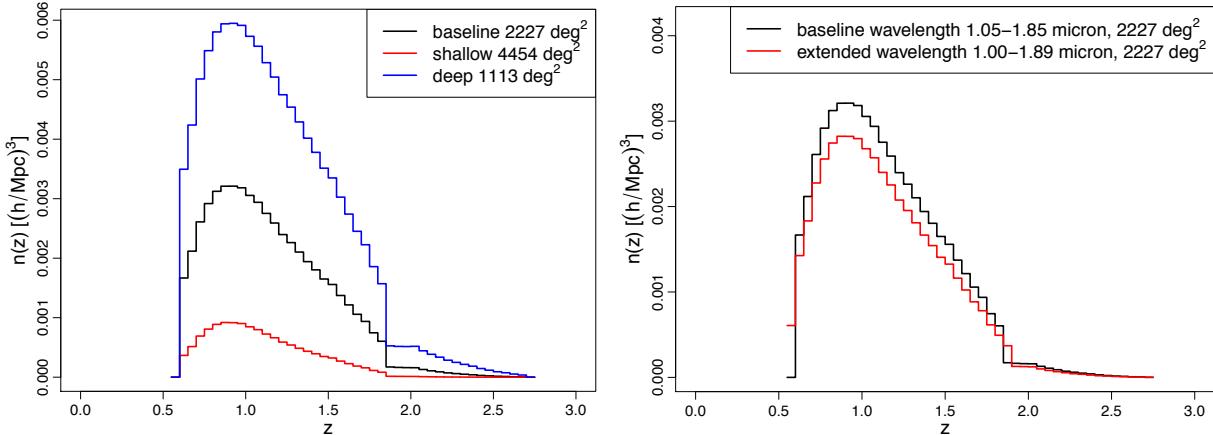


Figure 11: *left:* Redshift distribution of galaxies of the baseline, the shallow/large area and the deep/small area survey. *right:* Redshift distribution of galaxies of the baseline and extended wavelength survey.

(Eifler et al 2014, Krause&Eifler 2017) to compute the constraining power of all scenarios on cosmic acceleration, closely following Wang et al (2013). We run 500,000 step MCMC simulated likelihood analysis in a 23 dimensional parameter space. We simultaneously vary 7 cosmological parameters and 16 ?nuisance? parameters describing uncertainties due to the linear galaxy bias model, the non-linear smearing of the BAO feature, peculiar velocity dispersion, power spectrum shot noise, and redshift errors. We assume priors on cosmological parameters from the current state of the art experiments, i.e. the Planck mission, the Baryon Oscillation Spectroscopic Survey, the Joint Lightcurve Analysis, as described in Aubourg et al (2015).

The information gain is quantified using the standard Dark Energy Task Force FOM and an extended cosmology FOM, which measures the enclosed volume in the full 7-dimensional cosmological parameter space, not just in the 2 dark energy parameters. We will refer to these FOMs as DE-FOM and Cosmo-FOM. Compared to our baseline scenarios we find a decreased DE-FOM of 32% and a decreased Cosmo-FOM of 45% for the shallow/large area survey. For the deep/small area survey we find an increased DE-FOM of 5% and an increased Cosmo-FOM of 2%. We note that these findings are model and prior dependent and recommend further studies varying these input parameters. We also note that the [O III] numbers are pending a future update in part due to the reduction in the baseline telescope temperature to 260 K.

HLSS 3: The redshift range of galaxies surveyed shall encompass $1. \leq z \leq 2.7$ (3 desired, max z TBC depending on telescope temperature trade)

In addition to the trade studies in HLSS 1 we examine the impact of an extended wavelength range on the DE-FOM and the Cosmo-FOM. We follow the same procedure as detailed in the HLSS 1 paragraph extending the wavelength range from 1.05-1.85 microns for the baseline model to 1.00-1.89 for the extended model. The corresponding redshift distributions of the galaxy samples computed from the ETC v1.14 are depicted in Fig. 2. We find a decreased DE-FOM of 2% and a decreased Cosmo-FOM of 11% for the extended wavelength survey with respect to our baseline scenario. We iterate that these findings are model and prior dependent and recommend further studies varying these input parameters.

5.2 HLIS Forecasts

5.3 Shear tomography power spectra

We compute the linear power spectrum using the [?] transfer function and model the non-linear evolution of the density field as described in [?]. Time-dependent dark energy models ($w = w_0 + (1 - a)w_a$) are incorporated following the recipe of ICOSMO [?], which in the non-linear regime interpolates Halofit between flat and open cosmological models [also see ? , for more details].

Having obtained the density power spectra we calculate the shear power spectra as

$$C^{ij}(l) = \frac{9H_0^4\Omega_m^2}{4c^4} \int_0^{\chi_h} d\chi \frac{g^i(\chi)g^j(\chi)}{a^2(\chi)} P_\delta \left(\frac{l}{f_K(\chi)}, \chi \right), \quad (5)$$

with l being the 2D wave vector perpendicular to the line of sight, χ denoting the comoving coordinate, χ_h is the comoving coordinate of the horizon, $a(\chi)$ is the scale factor, and $f_K(\chi)$ the comoving angular diameter distance (throughout set to χ since we assume a flat Universe). The lens efficiency g^i is defined as an integral over the redshift distribution of source galaxies $n(\chi(z))$ (see Sect. ?? fro details) in the i^{th} tomographic interval

$$g^i(\chi) = \int_{\chi}^{\chi_h} d\chi' n^i(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')} . \quad (6)$$

Since we chose five tomographic bins, the resulting data vector which enters the likelihood analysis consists of 15 tomographic shear power spectra, each with 12 logarithmically spaced bins ($l \in [100; 5000]$), hence 180 data points overall. The limits of the tomographic z -bins are chosen such that each bin contains a similar number of galaxies.

5.4 Shear covariances

Under the assumption that the 4pt-function of the shear field can be expressed in terms of 2pt-functions (so-called Gaussian shear field) the covariance of projected shear power spectra can be calculated as in [?]

$$\text{Cov}_G(C^{ij}(l_1)C^{kl}(l_2)) = \langle \Delta C^{ij}(l_1) \Delta C^{kl}(l_2) \rangle = \frac{\delta_{l_1 l_2}}{2f_{\text{sky}}l_1\Delta l_1} [\bar{C}^{ik}(l_1)\bar{C}^{jl}(l_1) + \bar{C}^{il}(l_1)\bar{C}^{jk}(l_1)] , \quad (7)$$

with

$$\bar{C}^{ij}(l_1) = C^{ij}(l_1) + \delta_{ij} \frac{\sigma_\epsilon^2}{n^i} , \quad (8)$$

where the superscripts indicate the redshift bin; n^i is the density of source galaxies in the i^{th} redshift bin; and σ_ϵ is the RMS of the shape noise.

Since non-linear structure growth at late time induces significant non-Gaussianities in the shear field, using the covariance of Eq. (7) in a likelihood analysis results in underestimates of the errors on cosmological parameters. Therefore, the covariance must be amended by an additional term, i.e. $\text{Cov} = \text{Cov}_G + \text{Cov}_{NG}$. The non-Gaussian covariance is calculated from the convergence trispectrum T_κ [? ?], and we include a sample variance term $T_{\kappa,HSV}$ that describes scatter in power spectrum measurements due to large scale density modes [? ?],

$$\text{Cov}_{NG}(C^{ij}(l_1), C^{kl}(l_2)) = \int_{|\mathbf{l}| \in l_1} \frac{d^2\mathbf{l}}{A(l_1)} \int_{|\mathbf{l}'| \in l_2} \frac{d^2\mathbf{l}'}{A(l_2)} \left[\frac{1}{\Omega_s} T_{\kappa,0}^{ijkl}(\mathbf{l}, -\mathbf{l}, \mathbf{l}', -\mathbf{l}') + T_{\kappa,HSV}^{ijkl}(\mathbf{l}, -\mathbf{l}, \mathbf{l}', -\mathbf{l}') \right] , \quad (9)$$

with $A(l_i) = \int_{|l| \in l_i} d^2l \approx 2\pi l_i \Delta l_i$ the integration area associated with a power spectrum bin centered at l_i and width Δl_i .

The convergence trispectrum $T_{\kappa,0}^{ijkl}$ is, in the absence of finite volume effects, defined as

$$T_{\kappa,0}^{ijkl}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_4) = \left(\frac{3}{2} \frac{H_0^2}{c^2} \Omega_m \right)^4 \int_0^{\chi_h} d\chi \left(\frac{\chi}{a(\chi)} \right)^4 g^i g^j g^k g^l \times \chi^{-6} T_{\delta,0} \left(\frac{\mathbf{l}_1}{\chi}, \frac{\mathbf{l}_2}{\chi}, \frac{\mathbf{l}_3}{\chi}, \frac{\mathbf{l}_4}{\chi}, z(\chi) \right), \quad (10)$$

with $T_{\delta,0}$ the matter trispectrum (again, not including finite volume effects), and where we abbreviated $g^i = g^i(\chi)$.

We model the matter trispectrum using the halo model [? ?], which assumes that all matter is bound in virialized structures that are modeled as biased tracers of the density field. Within this model the statistics of the density field can be described by the dark matter distribution within halos on small scales, and is dominated by the clustering properties of halos and their abundance on large scales. In this model, the trispectrum splits into five terms describing the 4-point correlation within one halo (the *one-halo* term T^{1h}), between 2 to 4 halos (*two-, three-, four-halo* term), and a so-called halo sample variance term T_{HSV} , caused by fluctuations in the number of massive halos within the survey area,

$$T = T_0 + T_{HSV} = [T_{1h} + T_{2h} + T_{3h} + T_{4h}] + T_{HSV}. \quad (11)$$

The *two-halo* term is split into two parts, representing correlations between two or three points in the first halo and two or one point in the second halo. As halos are the building blocks of the density field in the halo approach, we need to choose models for their internal structure, abundance and clustering in order to build a model for the trispectrum.

Our implementation of the one-, two- and four-halo term contributions to the matter trispectrum follows [?], and we neglect the three-halo term as it is subdominant compared to the other terms at the scales of interest for this analysis. Specifically, we assume NFW halo profiles [?] with the [?] fitting formula for the halo mass-concentration relation $c(M, z)$, and the [?] fit functions for the halo mass function $\frac{dn}{dM}$ and linear halo bias $b(M)$ (all evaluated at $\Delta = 200$), neglecting terms involving higher order halo biasing.

Within the halo model framework, the halo sample variance term is described by the change of the number of massive halos within the survey area due to survey-scale density modes; following [?] it is calculated as

$$\begin{aligned} T_{\kappa,HSV}^{ijkl}(\mathbf{l}_1, -\mathbf{l}_1, \mathbf{l}_2, -\mathbf{l}_2) &= \left(\frac{3}{2} \frac{H_0^2}{c^2} \Omega_m \right)^4 \times \int_0^{\chi_h} d\chi \left(\frac{d^2V}{d\chi d\Omega} \right)^2 \left(\frac{\chi}{a(\chi)} \right)^4 g^i g^j g^k g^l \\ &\times \int dM \frac{dn}{dM} b(M) \left(\frac{M}{\bar{\rho}} \right)^2 |\tilde{u}(l_1/\chi, c(M, z(\chi)))|^2 \\ &\times \int dM' \frac{dn}{dM'} b(M') \left(\frac{M'}{\bar{\rho}} \right)^2 |\tilde{u}(l_2/\chi, c(M', z(\chi)))|^2 \\ &\times \int_0^\infty \frac{k dk}{2\pi} P_\delta^{\text{lin}}(k, z(\chi)) |\tilde{W}(k\chi\Theta_s)|^2. \end{aligned} \quad (12)$$

5.5 Expanding the Science Case - Multi-Probe Forecasts

6 Operations Model for the HLS and Evaluation of Trades

Co-I Hirata is leading the development of the HLS observing plan, extending his previous tools used for the SDT. These tools incorporate observing constraints in the chosen orbit, an exposure-by-exposure observing sequence optimized with detailed model of overheads, and tiling/coverage maps

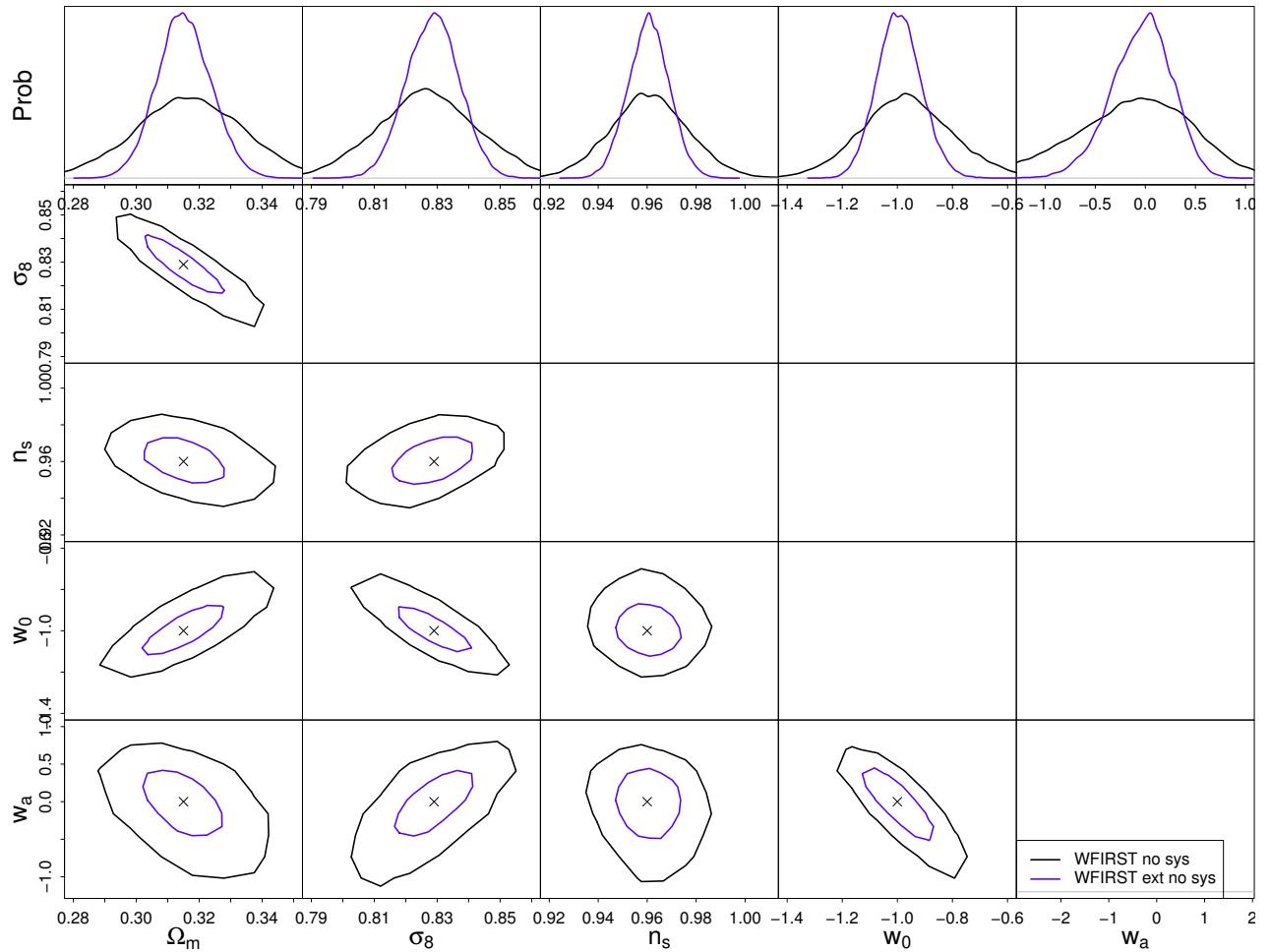


Figure 12: WFIRST forecasts statistical errors only. Extended Mission 10,000 deg 2 in blue, regular mission 2200 deg 2 in black

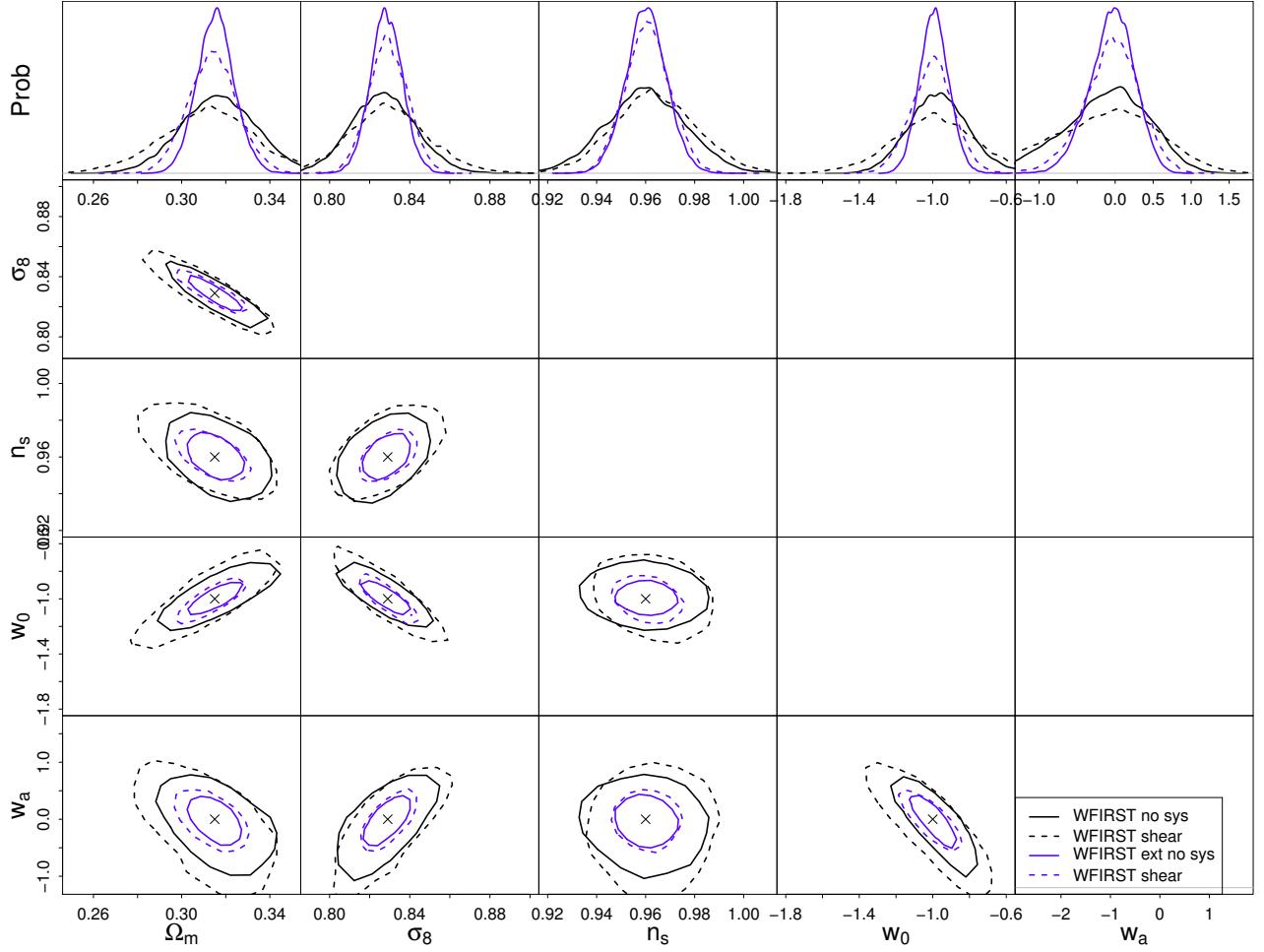


Figure 13: WFIRST forecasts statistical errors (solid) compared to errors when including uncertainties from multiplicative shear calibration errors (dashed). Shear calibration uncertainties are modeled as a Gaussian with $\sigma = 0.005$. Extended Mission 10,000 deg 2 in blue, regular mission 2200deg 2 in black

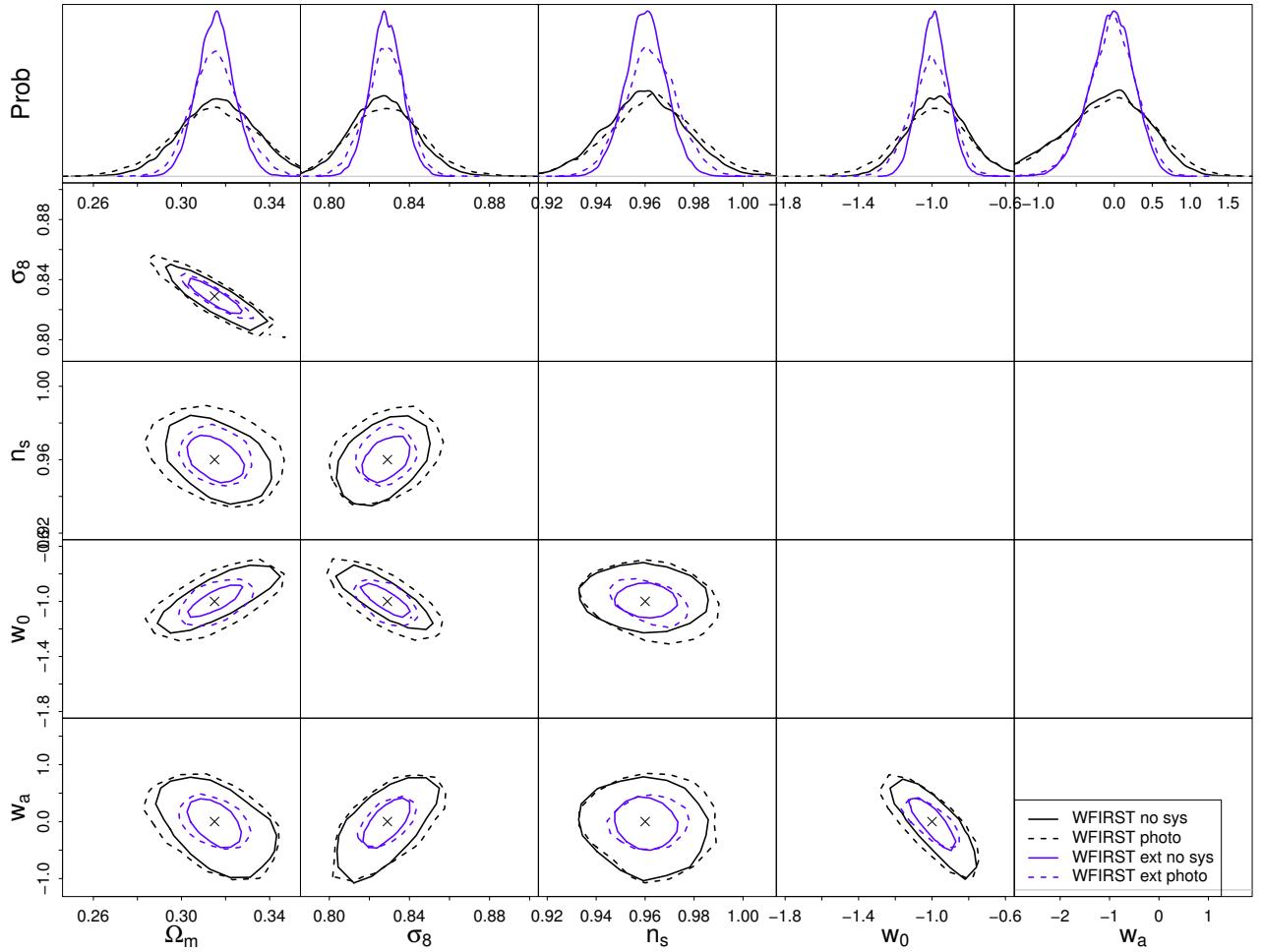


Figure 14: WFIRST contours when accounting for photo-z uncertainties. We model these as Gaussian photo-z errors with a bias around mean zero and $\sigma = 0.05$. These 2 parameters (bias and *sigma*) again have Gaussian priors, i.e. $\Delta\text{bias} = 0.005$ and $\Delta\sigma = 0.006$.

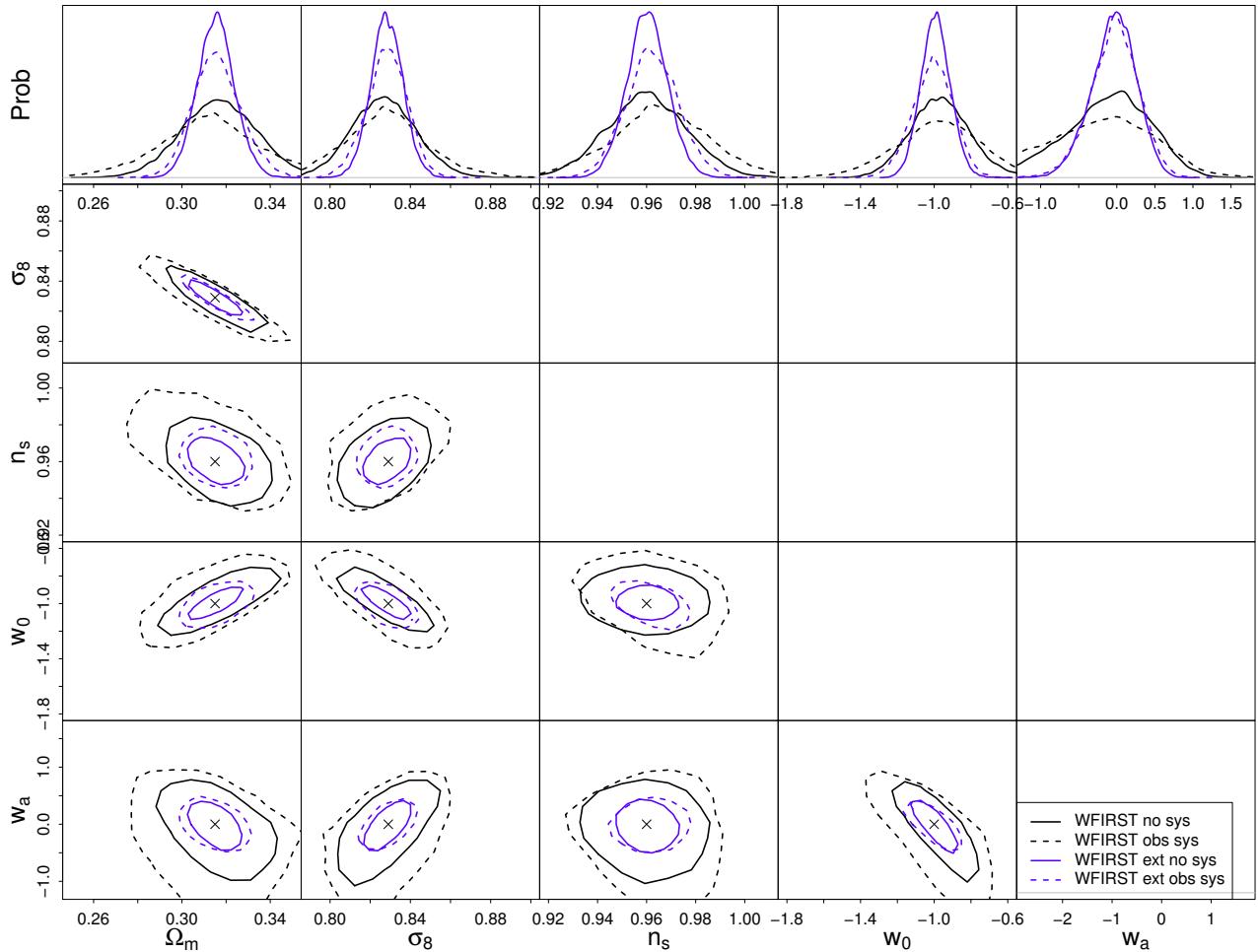


Figure 15: Broadening of WFIRST error bars accounting for shear calibration and photo-z errors

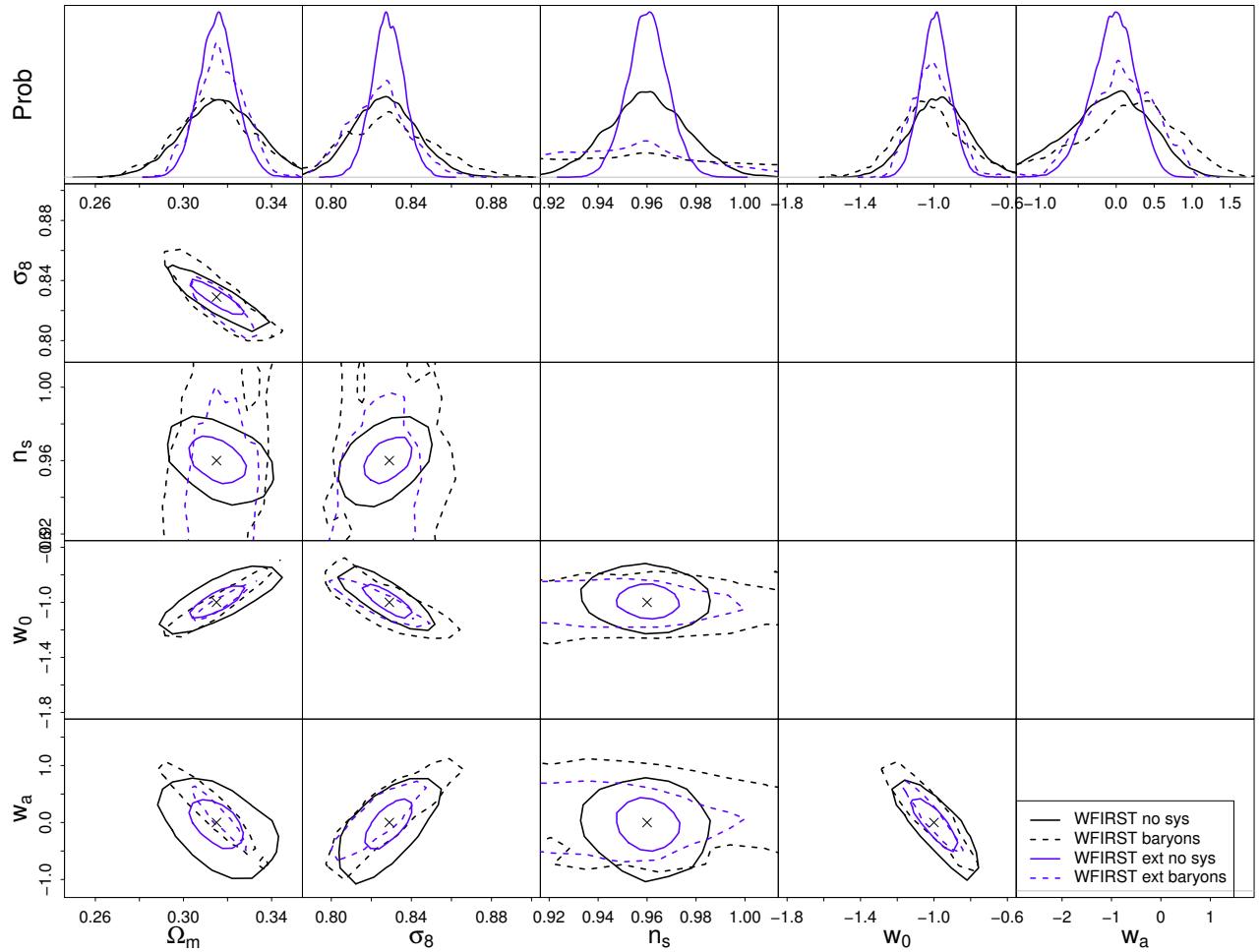


Figure 16: Broadening of WFIRST error bars when accounting for baryons.

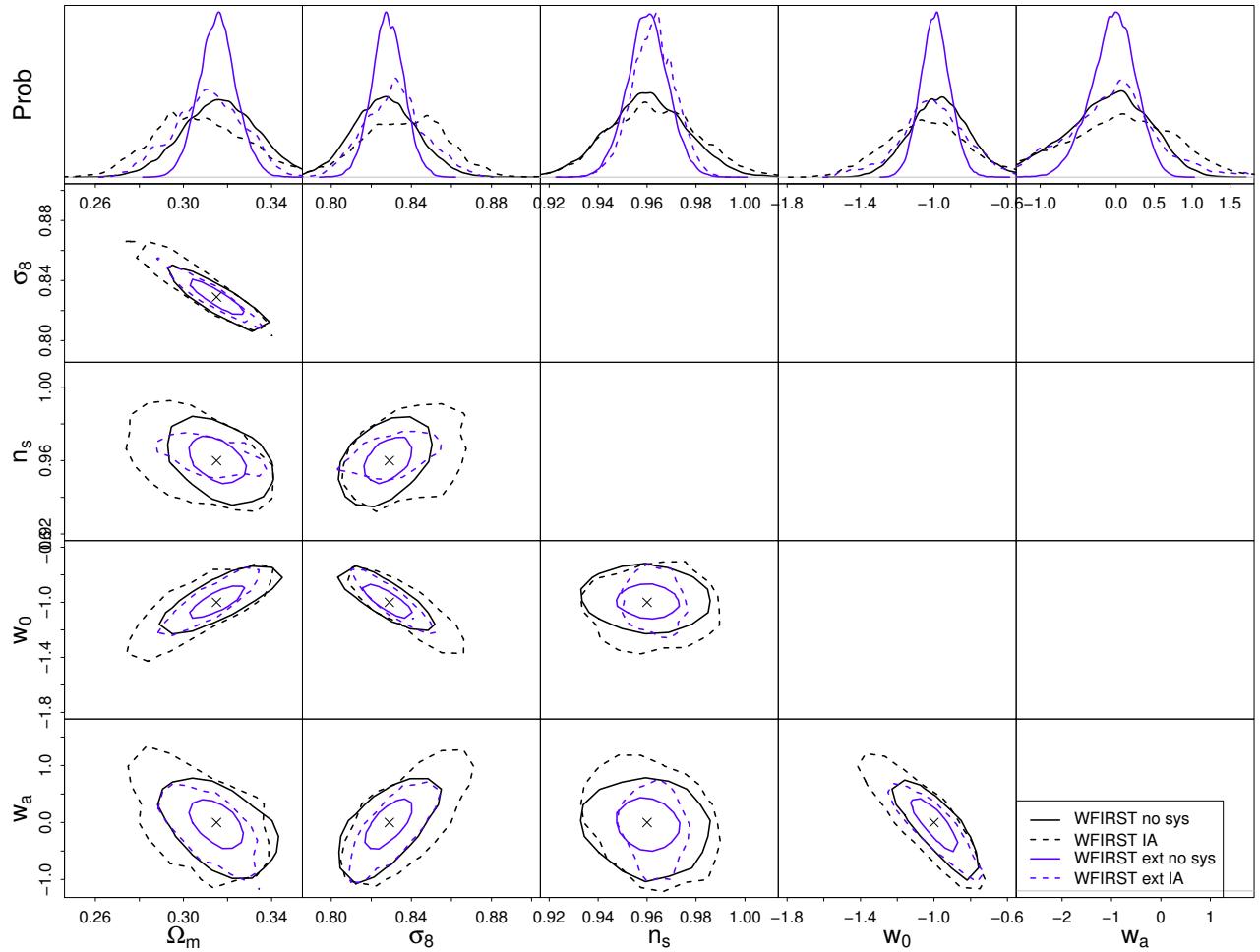


Figure 17: Broadening of WFIRST error bars when accounting for intrinsic alignment.

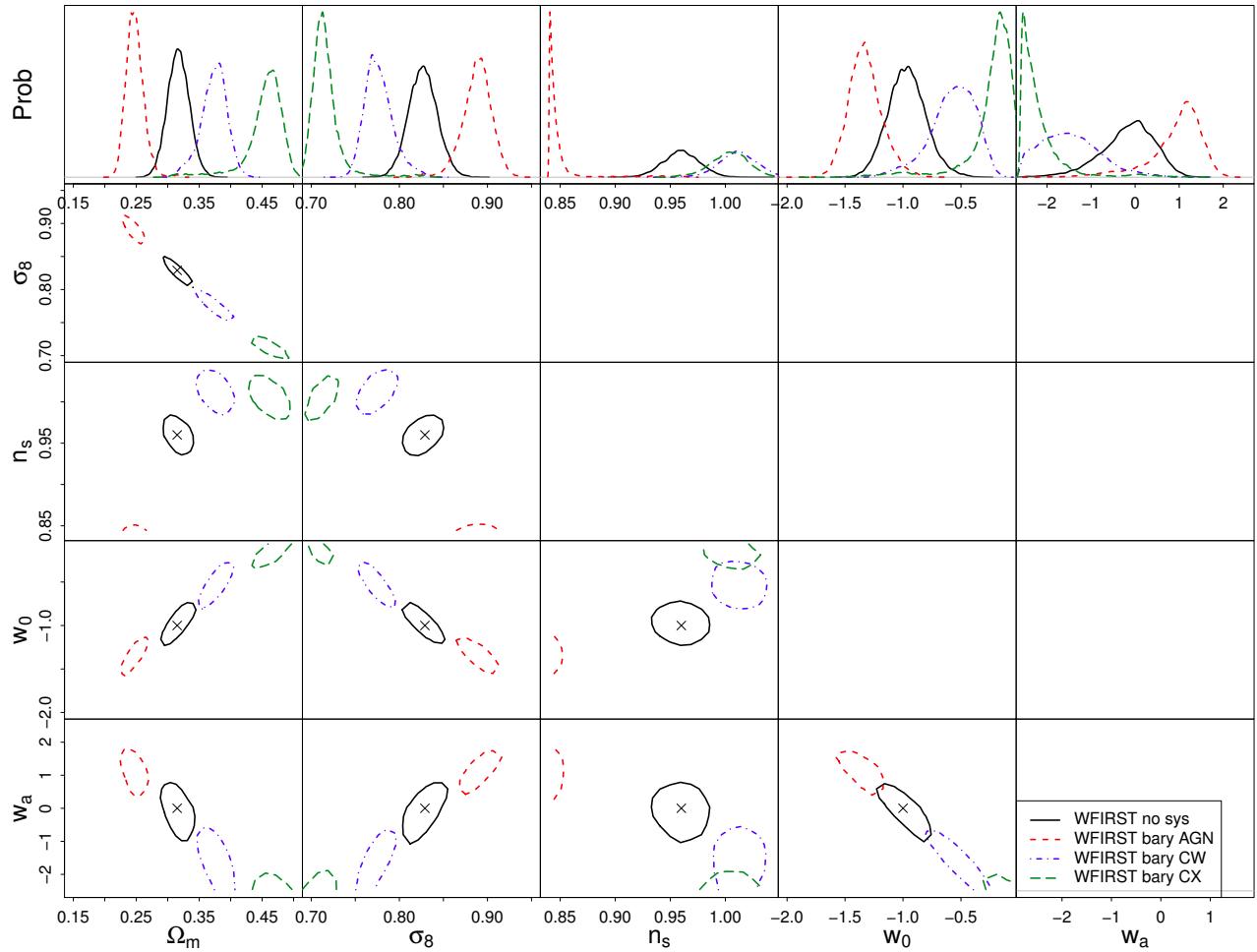


Figure 18: LSST IA marg

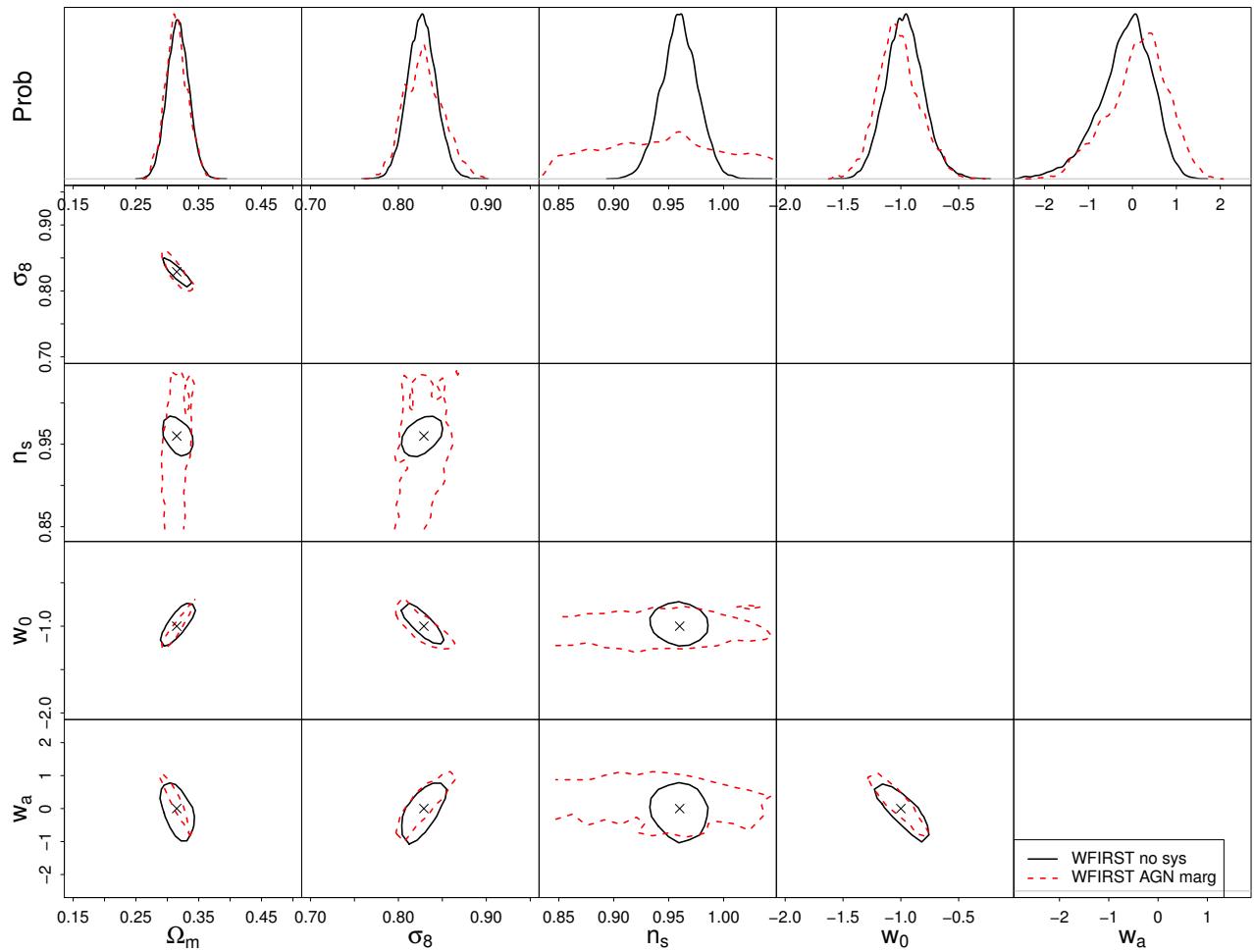


Figure 19: LSST IA marg

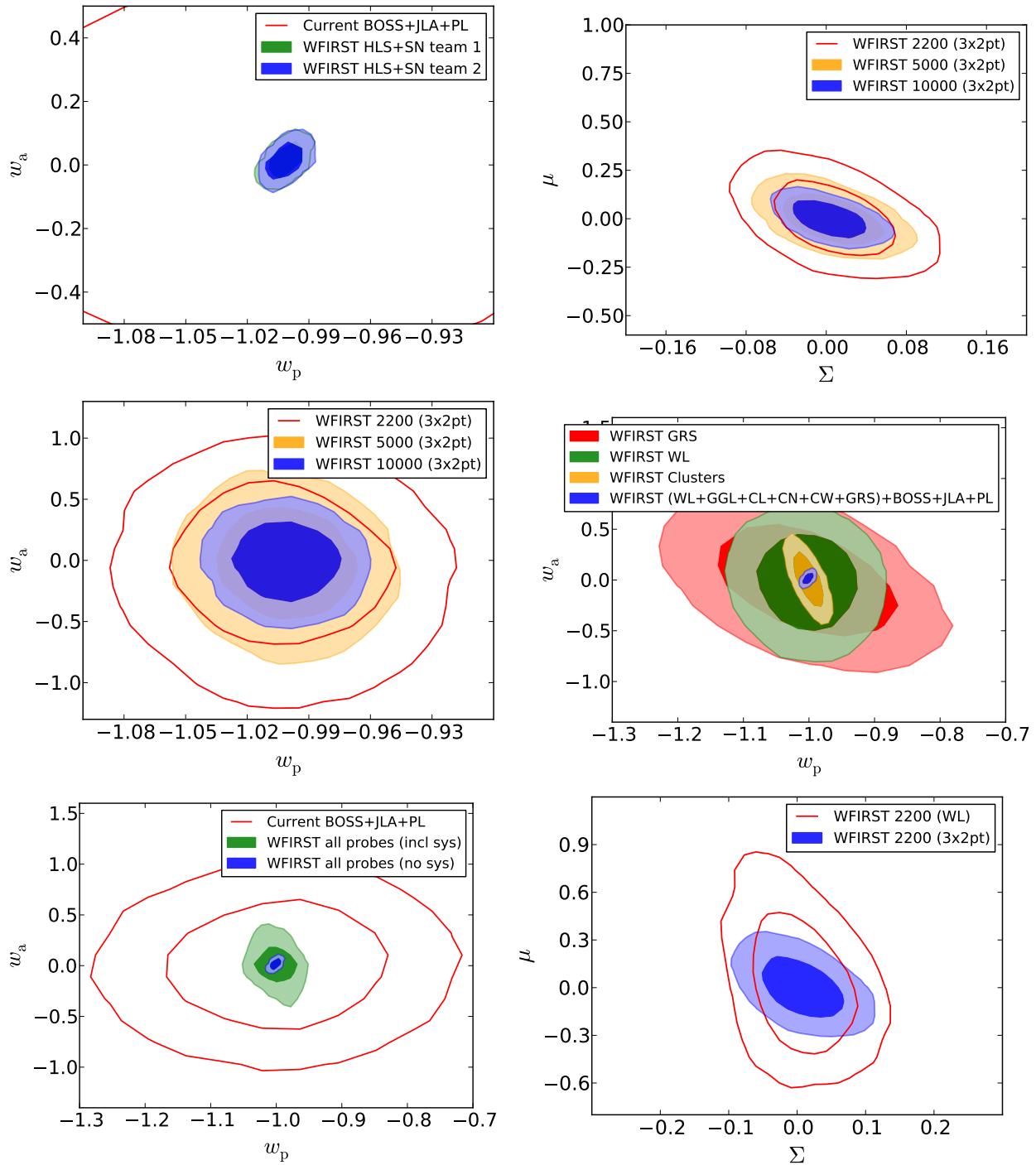


Figure 20: LSST IA marg

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. This effort is coordinated with the scheduling and Design Reference Mission Working Groups. We are both providing an example detailed plan for the HLS to the DRM working group, and cross-checking the Project’s spreadsheet-level survey calculators against our simulations. The HLS observing plan is also being transferred to the Calibration Working Group, since the HLS observing strategy feeds directly into the issue of self-calibration.

6.1 Snapshot of the HLS observing plan

Our team provided a “snapshot” of the HLS observing sequence to the full FSWG on April 19, 2017. This is by no means a final or even optimized version of the HLS, but is a work in progress as a result of trades in Phase A, as well as the recent decision to reduce the primary mission to 5 years.

Major updates relative to the SDT plan have included:

- A Lissajous orbit around L2. This is presently a place-holder, as the exact orbit has not yet been selected (and would depend on the launch date), but it gives a possible sampling of Sun, Earth, and Moon constraints.
- A rotated WFI (by 90° relative to the Cycle 6 design).
- Recommended slew and settle times provided by the Project.
- Faster detector readout (200 kHz instead of 100 kHz).
- Changes to the exposure time and dithering strategy to accommodate a 5-year baseline mission (as is to be presented to the WIETR). Specifically, we reduced exposure time to 140.2 s (imaging) and 297.0 s (spectroscopy); and changed the dither pattern in J band. (We are working on checking this strategy with image simulations, if it causes a problem we may have to revert.)
- Implemented bright star avoidance (observations are skipped if there is an $H_{AB} \leq 3$ star within 6 arcmin of any SCA).

Known current issues with the snapshot plan include:

- The SN and coronagraph programs in the code haven’t been updated since the SDT (except to cut the mission time by a factor of 5/6), even though they will likely change significantly. As in the SDT report, the coronagraph has blocks of time reserved. This will evolve in order to align the HLS plan with the other groups, as well as any changes to the scheduling architecture that we are directed to implement (e.g. block scheduling).
- We have begun putting the deep fields into this document, but right now they are (i) not fully specified, (ii) the tiling is not optimized, and (iii) some roll angles don’t align with WFIRST constraints (hence didn’t schedule). These issues will be solved in the next snapshot.

No policy decisions should be inferred from this sequence, as these will come from a higher level.

The survey bounding box is 2097 deg². The area covered with ≥ 3 exposures in every filter and the grism, including edge effects and holes around the bright stars, is 1947 deg². The time required for this version of the HLS is 394 days (imaging) + 215 days (spectroscopy).

The scheduling tools output a set of charts, included in this package:

- Fig. 18: Graphical display of the 5-year observing sequence.
- Fig. 19: HLS distribution of number of exposures in each filter.
- Fig. 20: HLS distribution of dust column [$E(B - V)$ in magnitudes]. Cosmological forecasts are based on a dust column of $E(B - V) = 0.035$ mag.
- Fig. 21: HLS distribution of zodiacal light (normalized to 1 at the ecliptic poles averaged over the year). Cosmological forecasts are based on a zodiacal brightness of 1.60 (except for 1.75 in the F184 filter, which is the least sensitive to zodiacal light and therefore was scheduled at inferior times of year).
- Fig. 22: The footprint of the HLS on the sky. This is an area of ongoing optimization, as we consider the needs of the deep fields, overlap with LSST, and the fraction of the survey footprint accessible from Northern observatories such as Subaru.

6.2 Further optimizations

Our team plans to study further optimizations to the HLS – including more drastic changes such as multi-tiered surveys, or a significant re-balancing of area vs. depth – in Phase B. However, in preparation for SRR/MDR, our main focus has been on demonstrating at least one survey configuration that meets requirements, and the construction of tools that link the observing strategy to calibration studies (§3.3) and image simulations (§??).

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.

6.3 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.

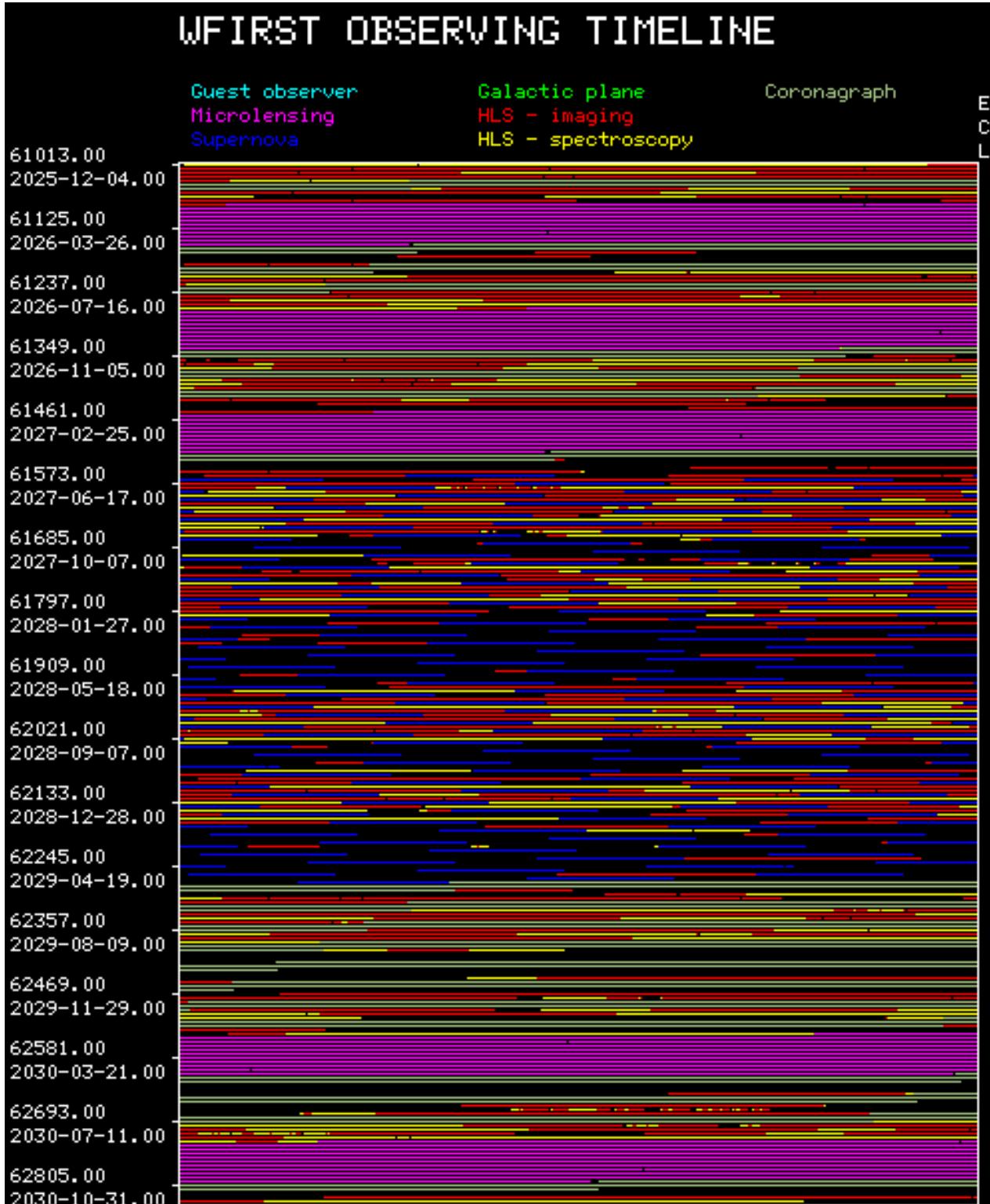


Figure 21: Observing timeline. Each row represents 7 days of observations, and is color-coded according to the observing program. Note the microlensing seasons (magenta), supernova survey (blue: \sim 5-day cadence), and HLS (red+yellow). Blank areas are not allocated. Labels on the left-hand side are shown every 16 weeks.

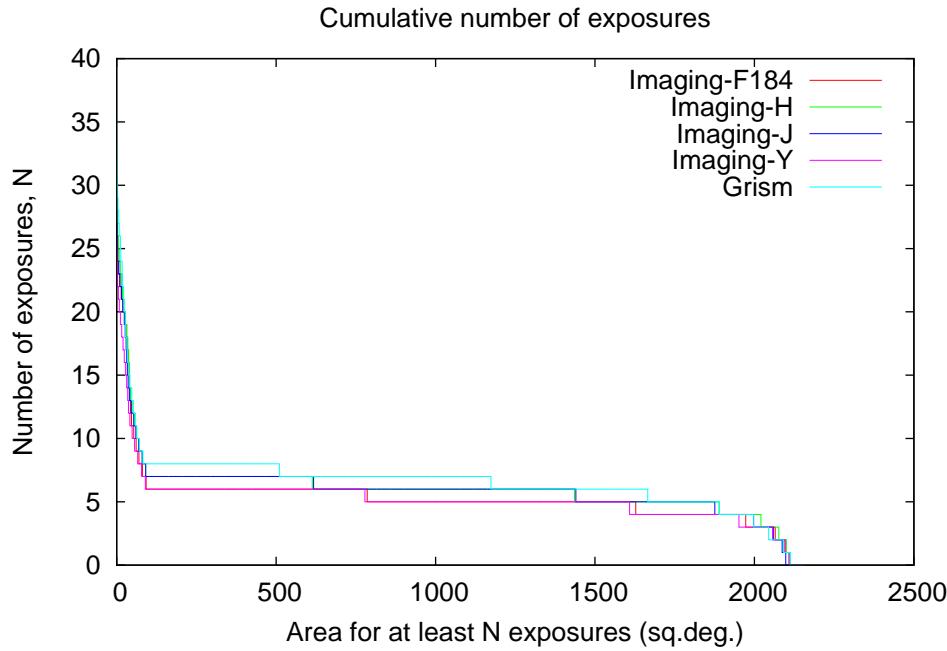


Figure 22: The cumulative distribution of HLS exposure depths above a certain area. The pile-up with many exposures at small area is the result of the deep fields.

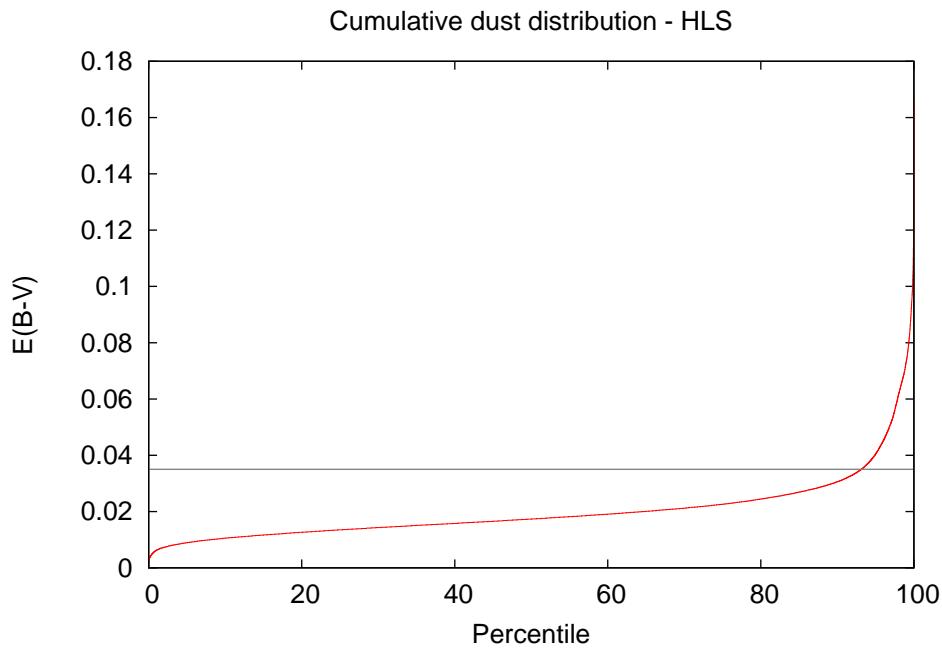


Figure 23: The cumulative distribution of Galactic dust in the HLS.

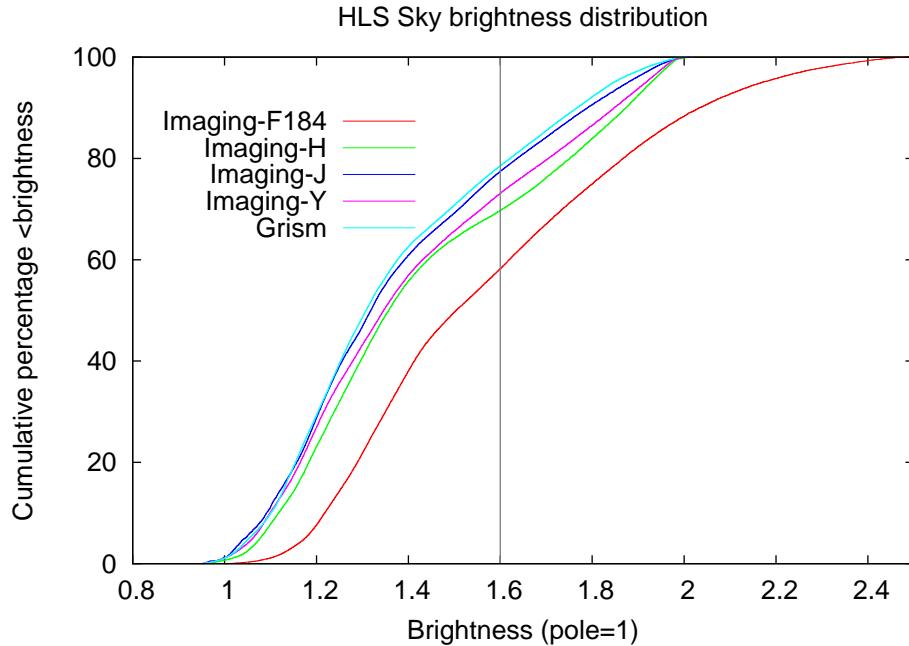


Figure 24: The cumulative distribution of zodiacal light in the HLS.

6.4 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

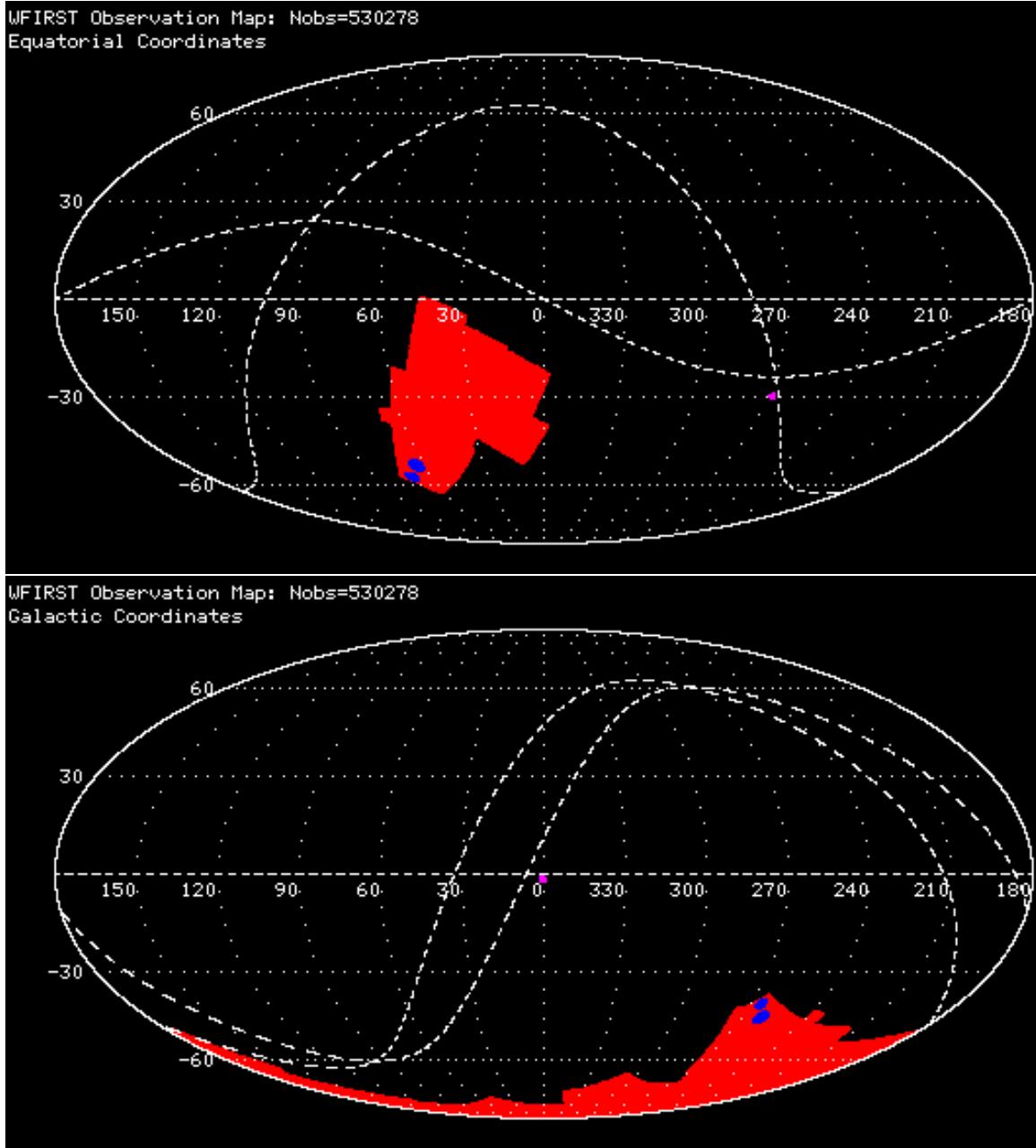


Figure 25: The footprint of the HLS (red) in Equatorial (top) and Galactic (bottom) coordinates.

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 ($\approx 2,200 \text{ deg}^2$), to a 7σ limiting line flux of $\sim 10^{-16} \text{ erg cm}^{-2} \text{ s}^{-1}$ 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 H α 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. 23 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 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 investi-

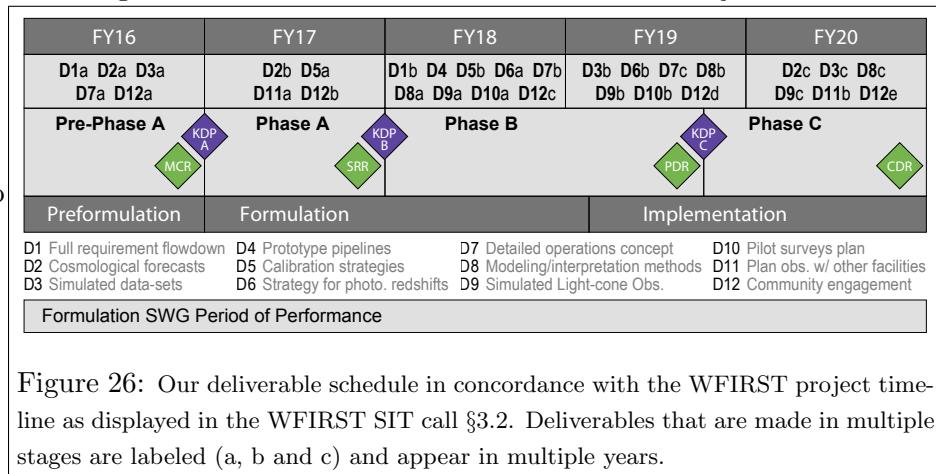


Figure 26: Our deliverable schedule in concordance with the WFIRST project timeline as displayed in the WFIRST SIT call §3.2. Deliverables that are made in multiple stages are labeled (a, b and c) and appear in multiple years.

gation 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	NIR	– near-infrared
AdvACT	– Advanced ACT	NRA	– NASA Research Announcement
AFTA	– Astrophysics Focused Telescope Asset	NRC	– National Research Council
BAO	– baryon acoustic oscillations	NWNH	– New Worlds, New Horizons
BOSS	– Baryon Oscillation Spectroscopic Survey	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
CL	– galaxy clusters / cluster growth	SAM	– semi-analytic galaxy formation models
CMB	– cosmic microwave background	SDSS	– Sloan Digital Sky Survey
CMB-S4	– CMB stage 4 experiment	SDT	– Science Definition Team
$D(z)$	– distance-redshift relation	SDT13	– 2013 WFIRST SDT report
$D_A(z)$	– angular-diameter distance	SDT15	– 2015 WFIRST SDT report
DE	– dark energy	SIT	– Science Investigation Team
DES	– Dark Energy Survey	SMEX	– NASA Small Explorer
DESC	– Dark Energy Science Collaboration	SN	– supernovae
DESI	– Dark Energy Spectroscopic Instrument	S/N	– signal-to-noise
ELG	– emission line galaxies	SPHEREx	– Spectrophotometer for the History of the Universe, Epoch of Reionization, and Ices Explorer
ELTs	– Extremely Large Telescopes	SPT-3G	– South Pole Telescope Third-Generation Camera Survey
eROSITA	– extended Roentgen Survey with an Imaging Telescope Array	STEP	– Shear Testing Program
ETC	– Exposure Time Calculator	STIS	– HST Space Telescope Imaging Spectrograph
f_g	– fluctuation growth rate	SZ	– Sunyaev-Zeldovich
FSWG	– Formulation Science Working Group	$w(z)$	– dark energy equation-of-state
GEO	– geostationary earth orbit	WFC3	– HST Wide-Field Camera 3
GGL	– galaxy-galaxy lensing	WFIRST	– Wide-Field Infrared Survey Telescope
GREAT	– Gravitational Lensing Accuracy Test	WISPs	– HST WFC3 IR Spectroscopic Parallel survey
GREAT3	– The third GREAT challenge	WL	– weak lensing
GRS	– Galaxy Redshift Survey	WMAP	– Wilkinson Microwave Anisotropy Probe
$H(z)$	– Hubble parameter	WPS	– WFIRST Preparatory Science
HLS	– High Latitude Survey	WSC	– WFIRST Science Centers
HOD	– halo occupation distribution	z	– redshift
HSC	– Subaru Hyper Suprime-Cam	z_p	– pivot redshift
HST	– Hubble Space Telescope		
IA	– intrinsic galaxy alignments		
IPAC	– Infrared Processing and Analysis Center		
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.