The Subaru-PFS/Roman (SuPR) Deep Survey: Redshifts for Roman Cosmology

White Paper Submission for
Roman/Subaru Synergistic Observations

Proposers:

Brett Andrews[1], Chihway Chang[2], Ami Choi[3], Olivier Doré[4,5], Chris Hirata[6], Arun Kannawadi[7], Alex Malz[8], Daniel C. Masters[5], Hironao Miyatake[9], Jeffrey A. Newman[1], Atsushi Nishizawa[10], Jennie Paine[3,11], Nikolina Sarcevic[7], Tomomi Sunayama[12], Chun-Hao To[2], Michael Troxel[7], Boyan Yin[7], for the Roman HLIS Cosmology Project Infrastructure Team[13], as well as Shun Saito[14,15,16], and Claudia Scarlata[16,17]

[1]University of Pittsburgh, [2]University of Chicago, [3]GSFC, [4]JPL, [5]California Institute of Technology, [6]The Ohio State University, [7]Duke University, [8]Carnegie Mellon University, [9]Nagoya, [10]Gifu Shotoku, Nagoya, [11]UMBC, [12]ASIAA/Nagoya, [13]Roman HLIS Cosmology PIT, [14]Missouri University of Science and Technology, [15]Kavli IPMU, WPI, [16]Roman HLSS Cosmology PIT, [17]University of Minnesota

March 14, 2025

1 Overview

Roman HLIS 3×2 pt and cluster cosmology require both well-characterized individual galaxy redshifts as well as precisely constrained redshift distributions of galaxies in tomographic shear bins. **Achieving both of these goals will depend critically on having representative spectroscopic redshift (spec-z) samples spanning the color/magnitude space of the photometric samples used**, both for (1) optimizing template-based or machine learning estimators for individual galaxies and (2) for precise characterization of redshift distributions of galaxies in weak lensing shear bins. We propose a PFS survey program focused on providing the necessary samples: **the Subaru-PFS/Roman (SuPR) Deep Survey.**

Due to its high multiplex and full sensitivity from the optical to the IR, Subaru-PFS will be by far the best instrument in the world for obtaining the secure redshifts needed to support Roman cosmology with galaxies to a depth limit of $H_{AB} \sim 24.5$. Here we propose a deep, H-limited survey over two 10 sq. deg. fields, COSMOS and SXDS/XMM-LSS, both of which are LSST deep-drilling fields (DDFs) also covered deeply by Roman and contiguous with the Roman High Latitude Imaging Survey (HLIS). Long exposures will be required to reach redshift success rates >90% for sample sizes of 10k–30k objects in total (with the total needed for this program depending on what spectroscopy will be obtained for other Stage IV imaging surveys).

The resulting deep spectroscopic sample would both improve photometric redshift (photo-z) estimates for individual objects and provide detailed characterization of results from photo-z algorithms. **All extragalactic science in the Roman High Latitude Wide Area Survey (HLWAS) region would benefit from this sample**, not just the static-sky probes of cosmology which are the focus of our team. For instance, galaxy evolution studies will depend critically on photo-z's to measure how galaxy populations change over time, identify unusual objects, etc. Better photo-z's will also help resolve redshift ambiguities for the Roman High-Latitude Spectroscopic Survey (HLSS) and identify transient sources of interest in deep drilling fields. The HLIS Cosmology Project Infrastructure Team will be producing photometric redshifts from joint processing of Roman and Rubin Observatory data, which should vastly outperform the Roman-only photo-z's which are slated to be produced by the Roman Science Operations Center, enabling all of this science for the community – **IF** we can properly train and calibrate photo-z algorithms. Furthermore, our proposed deep spec-z sample in regions with the most extensive photometric coverage on the sky also promises to provide significant legacy value to the community in its own right by providing a broad census of galaxy spectral properties and SEDs as well as their evolution over time.

2 Scientific Rationale

2.1 Photometric Redshifts: A vital tool for Roman extragalactic science

Roman will provide data with an unmatched combination of depth, area, wavelength coverage, and image resolution that will contain a wealth of information on cosmology and galaxy evolution. However, it is extraordinarily difficult to obtain secure redshift measurements for large samples of galaxies at the expected depths of Roman imaging; even exposure times of many tens of hours at Keck would yield highly incomplete samples (Newman et al. 2015a). As a result, studies that fully exploit the depth of Roman will depend critically upon our ability to determine redshifts from imaging data alone – i.e., *photometric redshifts* (photo-z's).

Photo-z's for Cosmology with Roman Roman's mission-critical large-scale structure and weak lensing analyses (as well as combined "3 × 2-point" studies of cosmology) will depend on photo-z's both to subdivide samples into z bins and to characterize redshift distributions with the exquisite accuracy needed for interpreting the observations. Photo-z's are also an important tool for finding galaxy clusters, characterizing strong lenses, and identifying and studying SNe Ia. This proposal is led by members of the High Latitude Imaging Survey Cosmology Project Infrastructure Team, which is producing the infrastructure to enable 3 × 2-point and cluster cosmology measurements. If both redshift and stellar masses can be predicted well, it is also possible to predict the foreground shear for strong lens systems and the lensing magnification for supernovae, increasing their constraining power for cosmology.

Photo-z's for Galaxy Evolution with Roman Photo-z's will also be critical for Roman studies of galaxy evolution. They enable us to estimate the look-back time for each object—necessary for detecting changes over time—as well as enabling the determination of key parameters such as luminosities, stellar masses, star formation rates, and rest-frame colors. These quantities are all important for a wide variety of galaxy evolution studies, as their distributions and correlations (e.g., the mass–size relation) can be readily measured at both low- and high-z once the redshift is known. High-precision photo-z's also enable the measurement of local environments (i.e., galaxy overdensities) and clustering statistics, revealing the connections between galaxy properties and the underlying web of dark matter.

2.2 The need for optical–NIR spectroscopy

The Importance of Spectroscopy for Photometric Redshifts Machine learning-based photo-z algorithms require a training set of objects for which the quantity of interest is well-known; this may come from moderate-resolution spectroscopy (which can provide highly-secure and precise measurements of redshifts) or many-band photometry/low-resolution spectroscopy (which provide lower precision estimates that are more frequently catastrophically incorrect). For template-based methods, spectroscopy is used to test and optimize algorithms (e.g., refine templates and zero points). Larger and more complete spectroscopic samples improve the *performance* of photo-z algorithms, reducing errors for individual objects and hence improving studies of galaxy clustering, cluster cosmology, and galaxy evolution (Newman & Gruen 2022).

However, the application of photo-z's for 3×2 point cosmology are even more dependent on the *characterization* of redshift distributions, or equivalently, the calibration of the photo-z's used. Spectroscopic redshifts are critical for providing this characterization at the accuracy levels demanded by cosmology, with uncertainties on both the mean z, $\langle z\rangle/(1+z)$, and the second moment, $\sigma(\langle z\rangle)/(1+z)$, required to be $< 0.3\%$ or less in each tomographic bin for photo-z errors to not dominate all others in Stage IV surveys like Roman; see, for instance, The LSST Dark Energy Science Collaboration et al. (2018).

For Roman specifically, the Roman Space Telescope Science Requirements Document (RST-SYS-REQ-0020, Revision D) requires uncertainty on the mean z to be $< 0.2\%$ (requirement HLIS 2.1.10), though it does not specify requirements for the second moment. Adapting the best current methods for redshift distribution calibration (see Fig. 1; cf. Buchs et al. 2019; Myles et al. 2021) for Roman will require spectroscopic redshifts across the full color–magnitude–redshift space populated by objects used in lensing analyses, within fields with much deeper photometry than the imaging used for cosmology. Thanks to its optical-through-infrared sensitivity, resolution high enough to split the [O II] doublet and characterize Lyman-α emission, and high multiplex, PFS should be the best instrument in the world for obtaining the redshifts needed to characterize the faint galaxies that will dominate the Roman cosmology sample. Roman lensing studies will extend to faint magnitudes, with many objects in these IR-limited samples being at $1.5 < z < 2.5$; this makes the infrared wavelength coverage that PFS provides critical.

In contrast, the Roman HLSS grism spectroscopy will provide IR coverage, but even the deepest parts of HLSS will be quite shallow (with the deep tier achieving similar depth/line sensitivity to 3D-HST or DESI; Brammer et al. 2012) and biased towards just the strongest emission-line galaxies, which will bias photo-z probability distributions (Gruen & Brimioulle 2017). Similarly, high-accuracy photometric redshifts from, e.g., medium- and narrow-band photometric surveys can be beneficial for training photo-z algorithms, but the rates of catastrophically incorrect redshifts in such surveys at faint magnitudes are still well above 1% (e.g,. $\sim 5\%$ in COSMOS2020; Weaver et al. 2022). Fig. 2 shows that such outlier rates are far too high to achieve the

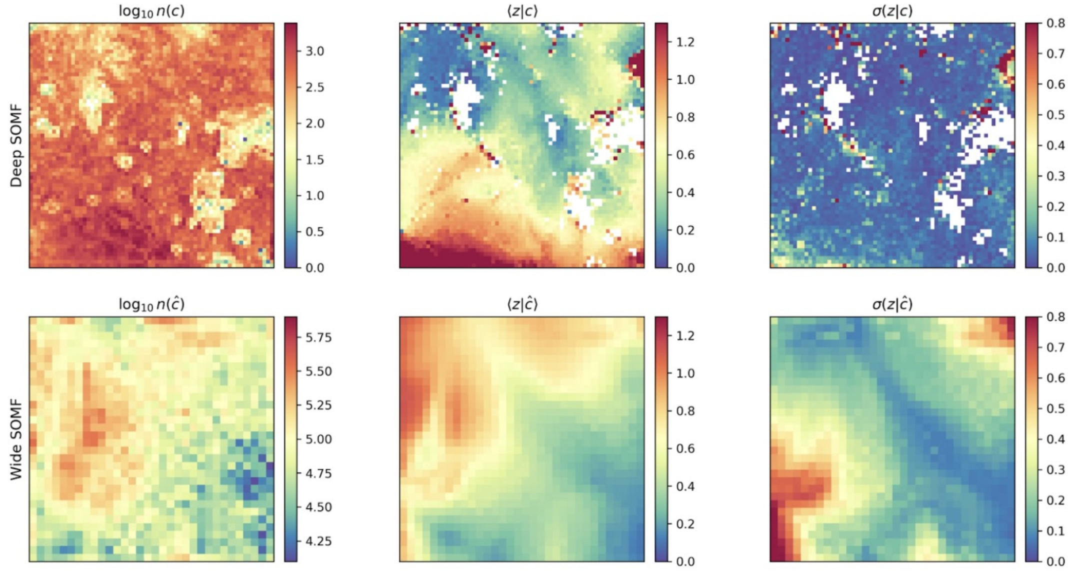

Figure 1: The most effective current methods for characterizing redshift distributions rely on redshift measurements over a deep sub-region with many bands, which are used to calibrate redshift distributions for shallower imaging in fewer bands across a wider field. The panels show 2-D representations of color space produced using a self-organizing map (SOM), where individual cells contain objects with similar spectral energy distributions (SEDs). The top and bottom rows show the $ugrizJHK$ and riz color spaces of the Dark Energy Survey Y3 deep- and wide-fields, respectively. The columns show the total number of galaxies (left), the mean redshift (middle), and the standard deviation of the redshift distribution (right) in each SOM cell. To accurately recover wide-field redshift distributions, these methods require spec-z coverage across the deep-field color space; ***for Roman samples, this can only be achieved with deep optical and NIR spectroscopy***. Reproduced from Campos et al. (2024).

precision calibration of redshift distributions needed for Roman weak lensing science (Newman & Gruen 2022). Additionally, Fig. 3 shows that we need to know the fraction of catastrophic outliers to 5–10% to avoid biasing cosmological parameter inference, which is much more difficult with medium-/narrow-band photometry. Thus, neither HLSS grism spectroscopy nor medium-/narrow-band imaging would be a sufficient substitute for the deep spectroscopy Subaru can provide.

As a result, PFS data will be urgently needed to train photo-z's and calibrate redshift distributions for Roman lensing analyses. Fig. 4 shows that even the best current deep field spec-z compilations (e.g., from the COSMOS field; Khostovan et al. 2025) severely lack coverage at faint NIR magnitudes ($K_s > 22.5$). These spec-z's are (1) biased towards high specific star formation rate galaxies and (2) primarily found in very small footprints where the deepest multi-band photometry exists (e.g., the CANDELS footprint within the COSMOS field), which exacerbates the effects of cosmic/sample variance in the training set. The SuPR-Deep survey would remedy this limitation in current deep spec-z samples by providing high-confidence spectroscopic redshifts for NIR-faint galaxies over a much larger sky area.

2.3 Sample size requirements

For Stage IV survey-like photometry, the uncertainties and catastrophic error rates in photometric redshift predictions from a wide variety of machine learning algorithms improve as the training sample gets larger, but this begins to reach the point of diminishing returns for training samples larger than $\sim 20,000$-$30,000$ objects, as illustrated in Fig. 5. Furthermore, a wide variety of theoretical attempts to estimate required spectroscopic sample sizes for characterizing redshift distributions for Stage IV weak lensing measurements have all ended up with estimated requirements in the same range (Ma & Bernstein 2008; Bernstein & Huterer 2010; Hearin et al. 2010). As a result, we aim to obtain a total spectroscopic sample of this size spanning the full range of properties of the objects that will be used in Roman lensing/3x2pt analyses, reaching a depth of $H \sim 24.5$. If we can cover parts of magnitude/color space with other instruments (cf. Section 4.7), we would endeavor to go deeper with PFS to ensure the highest possible redshift success for faint objects; if that is not feasible, we will use shorter exposure times and cover more objects with PFS alone.

2.4 Field selection

If redshifts are obtained over only a small area of sky, sample/cosmic variance will cause fluctuations in redshift distributions that can imprint correlated errors in estimated redshifts or distributions across the sky.

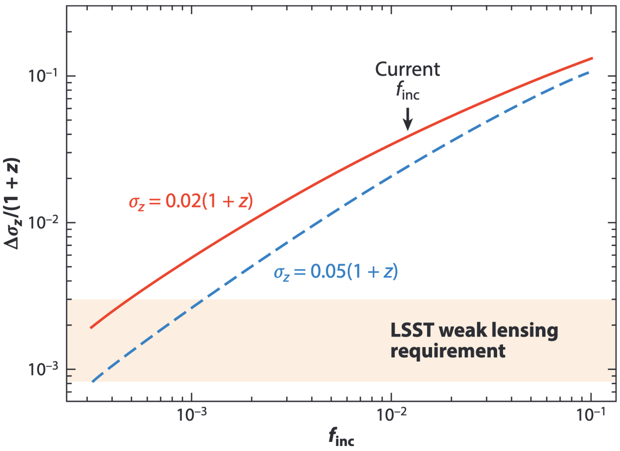

Figure 2: The bias in photo-z uncertainties ($\Delta\sigma_z/(1+z)$) increases with the fraction of objects in spectroscopic datasets with incorrect redshifts ($f_{\rm inc}$). The red and blue curves show how the bias scales with $f_{\rm inc}$ for two different levels of photo-z scatter. Current spectroscopic datasets have too many incorrect redshifts (arrow) to achieve Roman weak lensing requirements (which should be similar to LSST requirements; The LSST Dark Energy Science Collaboration et al. 2018). Medium-/narrow-band imaging will have even higher values of $f_{\rm inc}$, motivating our deep spectroscopic campaign in lieu of a many-band photo-z strategy. Adapted from Newman & Gruen (2022) Fig. 11.

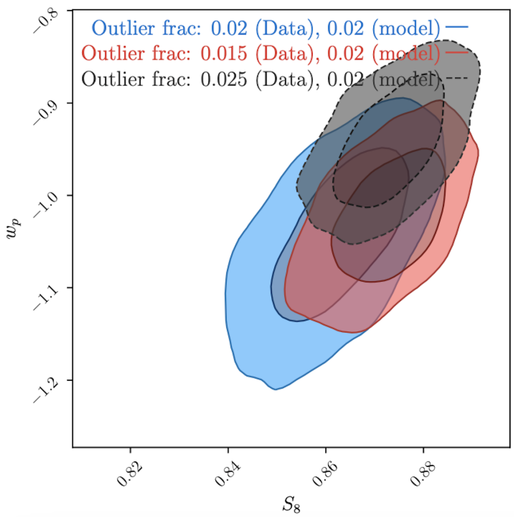

Figure 3: Forecasts of 3×2 point analyses for Stage IV experiments find that an incorrect determination of the fraction of catastrophic photo-z outliers (red and gray contours compared to the correct contours in blue) will significantly bias cosmological parameter inference, especially w_p–S_8 (shown here) and w_0-w_a (not shown). To determine the outlier fraction to within 5–10% (as would be needed to keep the bias introduced in w_0-w_a at $< 0.5\sigma$), ***we require high confidence spec-z's***. Adapted from Boruah et al. (2024) Fig. 8.

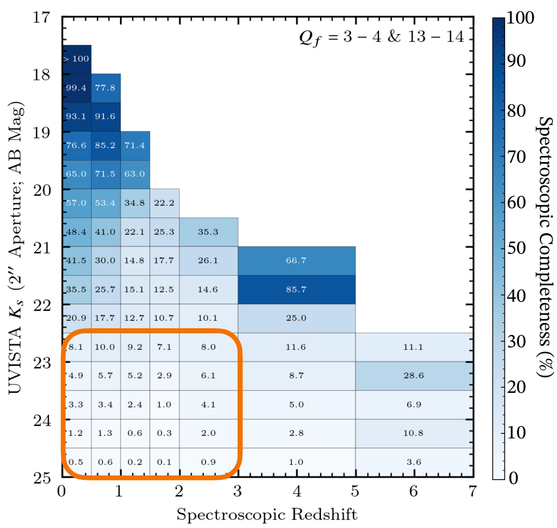

Figure 4: Existing spec-z datasets sparsely cover color–magnitude–redshift space, especially at faint NIR magnitudes and $z > 1$. This figure shows the spectroscopic completeness (the fraction of galaxies with spec-z's relative to the total population) of the COSMOS spec-z compilation as a function of redshift an dK_s mag for highly secure spec-z's. ***For the majority of Roman weak lensing sources (which will most commonly fall in the orange box corresponding to $K_s > 22.5$ and $z < 3$), the spectroscopic completeness is $\leq$10%.*** Adapted from Khostovan et al. (2025) Fig. 4.

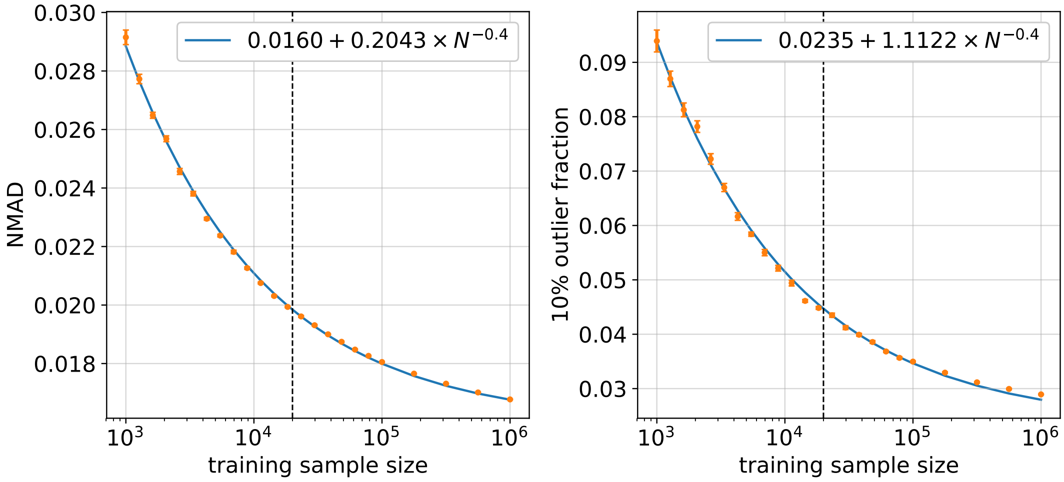

Figure 5: The scaling of photo-z errors (left) and outlier rates (right) as a function of the size of the spec-z training set (orange points with the best fit curves shown in blue), based on simulated galaxies from Graham et al. (2018). The vertical dashed lines indicate the point at which we are starting to reach diminishing returns, which corresponds to our minimum training set size of 20,000 spec-z's. This analysis focused on LSST photo-z's, though we expect the scaling to be similar for Roman samples. Reproduced from Newman et al. (2019).

Avoiding this issue requires obtaining redshifts over a significant volume (Newman et al. 2015b). In Fig. 6, we show the results of tests that indicate that obtaining samples spanning two 10 deg^2 contiguous patches would be sufficient to mitigate the impacts of these effects. This result is obtained by running the SOM-PZ (Campos et al. 2024) and error quantification pipeline from the Dark Energy Survey (DES) on the Cardinal simulation (To et al. 2024). We note that sparsely covering two fields of this size would have similar sample/cosmic variance to continuously covering the full contiguous field, so this could be sufficient for our needs.

Furthermore, we require that the fields are covered by deep photometry from both Roman and LSST in the same filters that will be used to infer redshifts for the Roman HLIS cosmology. Modern methods of calibrating redshift distributions for weak lensing are dependent upon the spectroscopic calibration sets used being obtained in fields with deep (essentially error-free) photometry, which allows the impact of photometric errors to be accurately modeled and assessed (Hildebrandt et al. 2017; Buchs et al. 2019; Myles et al. 2021).

The fields used also need to be contiguous with the HLIS lensing sky area, as the calibration in both the spectroscopic fields and the wider HLIS area need to be tied together to within 1 mmag. This holds because d(redshift)/d(color) is approximately 1 in the bands that span the 4000 Å break for a given object, so limiting redshift calibration systematics to $< 0.001(1+z)$ requires colors to be calibrated to $< 0.001(1+z)$ mag. It is very unlikely that this could be attained for non-contiguous fields; for DES, the lack of contiguous coverage between their deep and wide fields introduced a major source of uncertainty in the photometric calibration (Myles et al. 2021). This is even less likely to be possible if a different instrument from the LSST Camera were used for optical imaging (we further note that a very, very large amount of time on HSC would be needed to reach LSST deep-drilling depth over a 20 deg^2 area[1]).

Given these considerations, there are only two fields which are well-matched to our needs: COSMOS (10h) and XMM-LSS/SXDS (2h23m), both of which are equatorial LSST DDFs—and hence easily accessible from Subaru—that will also be Roman HLWAS DDFs. Unlike CDF-S/Fornax and the Euclid Deep Field South (EDFS), which are both planned LSST DDFs (with Roman imaging planned for the supernova survey for the latter), the COSMOS and XMM-LSS/SXDS fields are both contiguous with the planned Roman Medium/Wide tier imaging, allowing for robust photometric calibration. Furthermore, due to their position on the sky, it would be very costly in Subaru time to achieve a similar S/N in CDFS or especially EDFS as could be achieved in the equatorial LSST DDFs. The Northern Roman time domain field (ELAIS-N1) is much better positioned for Subaru, but is infeasible to target with LSST, making it very poorly suited for training and calibrating photo-z's for the Rubin+Roman photometric system.

3 Relevance to the Roman Core Community Survey

As described above, photometric redshifts will be an essential tool in essentially all extragalactic science with Roman. They will enable studies of how galaxies have evolved over time, help in the identification of transients

[1] Scaling according to etendue, a total of 170 nights on HSC would be required to reach 10-year LSST DDF depth over a 10 sq. deg. region (which is only half the total area needed), ignoring the inefficiency of HSC for u-band observations; even limiting to 4-year-DDF depth would still require 68 Subaru nights per field.

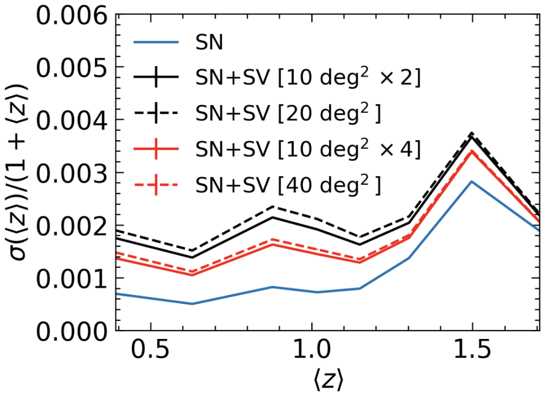

Figure 6: The uncertainty in the mean redshifts of tomographic bins is affected by shot noise (SN; blue line shows 20,000 objects) and the imprint of sample/cosmic variance (SV) due to calibration field area and number. Based on these tests, which utilize simulations from To, Yin, et al. (in prep.) that include both large-scale-structure and galaxy color information, two 10 deg^2 fields (black solid line) are nearly as good at eliminating the scatter caused by the imprint of SV as a single 40 deg^2 field (red dashed line).

of interest, disentangle contributions from galaxies at different redshifts in Roman grism spectroscopy, and – most relevant for the HLIS Cosmology Project Infrastructure Team – enable precise studies of cosmology using large-scale structure statistics, weak gravitational lensing, and galaxy cluster counts. Our proposed program would provide training samples that would reduce errors on individual photo-z's increasing the legacy value of Roman imaging, while simultaneously enabling the precision calibration of redshift distributions needed for photo-$z$based probes of cosmology to not be systematics-limited.

Synergies with the High Latitude Imaging Survey: Photometric redshift uncertainties and calibration needs for weak lensing and galaxy clustering cosmology were key drivers of the HLWAS design. In the interim report on the Core Community Surveys (CCS Definition Committees 2025), the HLWAS survey definition committee recommended a three-tier survey. The Medium tier (2400 deg^2 of YJH imaging down to $H = 26.4$ AB mag for 5σ point source) and Wide tier (2700 deg^2 of H-band imaging to $H = 26.2$ AB mag) will yield $\sim 6.2 \times 10^8$ galaxy shapes for weak lensing analyses; the additional bands in the Medium tier will provide better photo-z's (as well as characterization of wavelength-dependent PSF effects) that will enhance interpretation of the single-Roman-band Wide area.

Due to photo-z and weak lensing calibration requirements, the HLWAS also incorporates a Deep tier consisting of longer-exposure, multi-band Roman imaging (comprising 20 deg^2 of $WZYJHFK$ imaging to $H = 27.5$). The Deep-field-based studies will also require deep Rubin Observatory LSST imaging; in the end, the COSMOS and XMM-LSS LSST deep drilling fields were chosen for this tier because they are the only viable field choices that are both accessible to Subaru and are contiguous with the Medium/Wide tiers (cf. requirement HLIS 2.0.5 from the Roman Science Requirements Socument). **The best current photo-z calibration methods (see, e.g., Fig. 1) also require large samples of secure spectroscopic redshifts in the deep field that fully span the color space of photometrically-detected lensing galaxies** (Roman requirement HLIS 2.3.5 establishes $\geq 15,000$ spec-z's as a minimum). The combination of Deep tier HLWAS imaging + Subaru/PFS spectroscopy would also be a transformative dataset for galaxy evolution studies of distant and/or faint objects at (or surpassing) the depth of current state-of-the-art HST surveys but with 100–1000$\times$ the area and with spectroscopic redshift measurements for the full range of galaxies.

Synergies with the High Latitude Spectroscopic Survey: We have also been in communication with the HLSS team, who are interested in using the PFS spectroscopy we plan to obtain to characterize success rates in their spectroscopic samples. We have not attempted to optimize our procedures to maximize the utility for that application, but this may be possible. In any event, the improved photo-z's from our program would enable better identification of the true redshifts of objects with insecure grism z measurements directly over the full HLWAS area.

Synergies with the High Latitude Time Domain Survey: Transient studies with Roman should rely on host galaxy photo-z's to aid in transient classification and prioritize objects for follow-up. By training and calibrating combined Rubin + Roman photo-z's our sample should directly improve photo-z's in the Euclid Deep Field South time domain survey region. Additionally, if the optical imaging available in ELAIS-N1 can be tied to the Rubin system, photo-z's in that field should be improved as well. We note that the calibration requirements for this cross-photometric-system application are much looser than those for calibration of redshift distributions for weak lensing/LSS studies, since even photo-z's that are systematically at the 1% level would still clearly differentiate between transient classes. Stellar mass estimates and dark-to-luminous mass ratios enabled by the improved photo-z's plus galaxy-galaxy lensing measurments from the HLIS area can be used to map foreground magnification for time domain supernovae.

4 Required Observation Plans

4.1 Summary

In order both to train photo-z algorithms to improve weak lensing, galaxy clustering, and cluster cosmology and to characterize redshift distributions at the precision needed to make photo-z systematics subdominant in weak lensing/large-scale structure cosmology, we will need deep PFS spectroscopy for ∼10k–30k galaxies down to the depth of the planned weak lensing objects, $H_{AB} \sim 24.5$. However, due to the steepness of galaxy number counts, a sample that is simply magnitude-limited would be dominated by the faintest objects; as a result, we will select fainter objects at a lower rate to yield a flatter magnitude distribution in the final sample and more broadly and uniformly cover parameter space.

To achieve the required sample sizes, we will need a total of 5–15 PFS pointings, distributed over two ∼10 deg^2 patches to mitigate the impact of sample/cosmic variance (with the smaller number of pointings corresponding to scenarios where we would be able to exploit spectroscopic samples obtained in support of Rubin LSST or other Stage IV imaging surveys for part of our training/calibration spectroscopy). As described below, we will require exposure times in the range of 60h (if fewer pointings are used) to 20h (if we require more pointings) in order to achieve high rates of secure redshift measurements.

4.2 Instrument to be used

As described in Section 2.2, we require high-multiplex spectroscopy over wide fields covering both optical and infrared wavelengths in order to achieve high redshift success rates for large samples across the redshift range of Roman lensing samples. The PFS instrument is uniquely well-suited to this—not just amongst the instruments available at Subaru but *compared to all instruments available in the world today.*

4.3 Numbers of nights and exposure times

To make a meaningful contribution to the calibration, very high redshift success rates (in excess of 90%) are required. Tests have shown that DESI (with a PFS-like design but optical-only coverage) exhibits background-limited S/N scaling for faint objects in integration times ranging from 15 minutes to as long as 8 hours (Biprateep Dey et al. 2025, in prep.). With no reason to expect otherwise for PFS, this scaling would imply 20–60 hour integration times to achieve >90% redshift success rates with PFS at the weak lensing depth of Roman; we have obtained similar predictions based on simulated spectra and PFS errors. Without test observations of similar depth with optimal procedures it is difficult to be more definitive.

Since success rates should vary continuously with exposure time and magnitude, rather than requesting a total number of nights that would scale with the required exposure time which we are not certain of yet, we are requesting a fixed number of nights, and will split that time amongst different PFS pointings in order to a) take into account what other broad photo-z training/calibration spectroscopic campaigns are being undertaken (cf. Section 4.7) and b) take into account the actual instrument performance for our application. One limiting case applies if we need only obtain 10,000 spectra during Roman/Subaru time due to being able to take advantage of other similar (but optically-limited) programs; in this scenario, we will target 5 total PFS pointings sparsely covering our fields with 60 hours of exposure time each, equivalent to 50 nights total observing time assuming 6 good hours per night (given overheads and weather losses). In this case we would be able to maximize our redshift success rate all the way to $H = 24.5$. In the other limit, we will need to obtain a total of 30,000 spectra with PFS during Roman-Subaru time; in that case, we would target 15 pointings spanning our two fields with 20 hour exposure times each, again totaling 300 hours or 50 nights. The shallower spectroscopy may not be sufficient to ensure good success down to $H = 24.5$, in which case we would have to downgrade the magnitude limit for lensing samples (at least until further observations to enhance success at the faint end could be obtained).

If fewer than 50 nights are available for this program, the first descope we would undertake is scaling back from a total sample of 30,000 targets to 20,000, reducing the total exposure time needed by a third. This would lower the performance of photo-z algorithms due to having less training data and reduce the accuracy of our characterization of redshift distributions by $\sim 20\%$. In more extreme scenarios where even greater descoping is necessary, we would have to cut down exposure times. This would in turn limit the faintness of objects for which we have good photo-z training and calibration data, which would limit the depth that can be used securely for weak lensing analyses with Roman. We are thus very reluctant to cut total requests below 33 nights. As an example, if we were to cut exposure times by a factor of two, we would expect to reach the same redshift success rate at $H = 24.12$ that we would obtain in the full exposure time at $H = 24.5$. We presumably would then have to restrict lensing analyses to that shallower depth to ensure they are well-calibrated; however, doing so reduces the number of galaxies available by 19% (calculated upon a power-law fit to the F160W galaxy number counts compiled by Driver et al. (2016)). This far exceeds the 10% margin on effective number density of lensing sources allocated in the Roman Science Requirements Document (in requirement HLIS 2.0.1).

4.4 Required condition of nights

Due to the need for high sensitivity to detect spectral features from faint objects, dark time with $< 30\%$ moon is required. Observations should be conducted at airmasses below 1.5 and with better than $1''$ seeing; if conditions are not this good, additional observing time would be required to achieve our goals.

4.5 Time criticality

We anticipate that our earlier weak lensing analyses will be restricted to brighter objects and smaller sky areas, such that random errors are larger and photo-z calibration requirements are weaker. As a result, only shallower spectroscopy will be needed for earlier analyses (~ 2 years after launch). Only 40% as much PFS exposure time would be needed to reach a given level of completeness at $H = 24$ as at $H = 24.5$ in our full exposures, or 16% as much to reach that completeness at $H = 23.5$.

As a result, spectroscopic observations for our program can be somewhat back-loaded. We anticipate that our final analyses will take place roughly 5 years after launch, so having the full spectroscopic dataset in hand by 4 years after Roman launch would be ideal.

4.6 Survey flexibility

While our baseline plan would be to aim for very long integration times on all of the selected sources in the H_{AB}-limited survey (enhancing the value of the spectra of brighter objects for galaxy evolution studies), we could instead make the targeting of the survey more flexible by moving fibers off of sources that yield redshifts early in the total integration and moving to brighter targets, enlarging the total sample size and improving training and calibration for photo-z's at the bright end. We will explore in the future which of these scenarios will be best for our science.

4.7 Interactions with other spectroscopic programs

Significant amounts of spectroscopic time have been devoted to photometric redshift training for current and near-future imaging surveys, including the C3R2 survey (Masters et al. 2017, 2019; Stanford et al. 2021) and other projects inspired by it, such as the DC3R2 survey (McCullough et al. 2024) (which uses DESI spare fibers) and planned 4C3R2 survey(Gruen et al. 2023) (which will use 4MOST). However, the existing spectroscopy has substantial limitations that make it insufficient for Roman cosmology needs. Both DC3R2 and 4C3R2 can obtain spectra over large sky areas, essentially eliminating sample/cosmic variance, and should obtain large numbers of spectra in total. However, they are being observed in parallel to other, comparatively shallow spectroscopic programs on 4m telescopes, limiting the depth of the samples that can be obtained. Unfortunately, **the range of SEDs of bright galaxies does not cover the full range of colors of fainter galaxies** (as seen in both SDSS at $z \sim 0$ and DEEP2 at $z \sim 1$; cf. Blanton 2006), **so a substantial fraction of objects in deep H-limited Roman lensing samples will have no analogs in these surveys**. Furthermore, the spectral coverage of these programs is limited to optical wavelengths, limiting their utility at $z > 1.5$. Fig. 7 illustrates the substantial amounts of color space not covered by DC3R2, even for samples limited at comparatively shallow magnitudes compared to Roman lensing samples.

Conversely, C3R2 is obtaining spectra with larger telescopes and includes some infrared spectroscopy, but its total sample sizes are much smaller (due to the small multiplex of the instruments used) and the sample/cosmic variance is high due to the small sky areas covered. For C3R2, the errors due to sample/cosmic variance are too large to meet the requirements for calibration of Roman redshift distributions. **However, our planned survey geometry** (shown by the solid black line in Fig. 6) **can meet these requirements** due to its $> 3\times$ smaller errors.

The PFS/SuMIRe Galaxy Evolution survey will obtain spectra for galaxy samples that are color-selected to be at specific redshift ranges, limited in a different passband than Roman will use, and will utilize shorter exposure times than we expect to be required at the Roman lensing depth (Greene et al. 2022). Furthermore, its sky area (~ 12 sq. deg., or less for the deep sample) would be insufficient for our needs, even if it turns out to be entirely within the Rubin/Roman deep drilling fields (which would also be necessary for us). As a result of all this, some objects with redshifts from SuMIRe may be useful for our purposes, and we would avoid retargeting such objects, but it is very far from covering our requirements.

One can imagine (or hope) that more spectroscopy will be obtained in the future in support of Rubin Observatory LSST photometric redshift training and characterization. Required spectroscopic sample sizes are similar to those for Roman (20,000–30,000 objects in total), but for Rubin weak lensing analyses, they should be limited at $i < 25.3$ instead of $H \lesssim 24.5$. As a result, much of—but not all—the color space required for Rubin will also be relevant for Roman. In an optimistic scenario, significant DESI2 time could be devoted to Roman DDFs, targeting supernova hosts for transient cosmology and faint galaxies to enhance photo-z's; test spectroscopy for samples extending down to $i = 24.5$ with exposure times of up to 8 hours has been obtained with DESI. Additionally, there will be some Subaru nights provided to the US community that could also be used for photo-z training/characterization spectroscopy. DESI programs would cover large enough areas for sample/cosmic variance to be a non-issue so long as at least two distinct fields are covered; Subaru/PFS programs may or may not, depending on the strategy pursued.

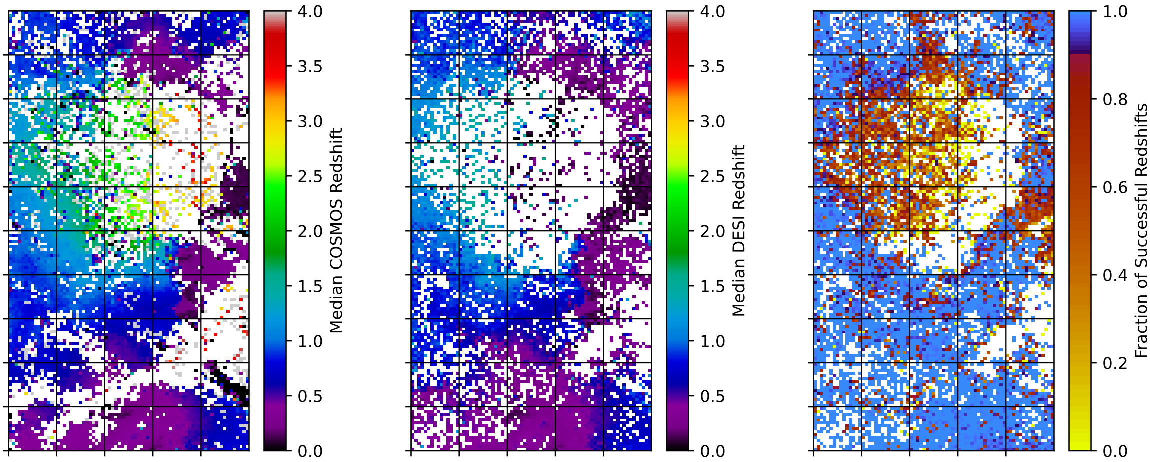

Figure 7: A SOM of $ugriZYJHKs$ color space from Masters et al. (2015) color-coded by median redshift from non-DESI COSMOS spec-z's (left), the DESI DC3R2 survey spec-z's (middle), and the fraction of successful redshift measurements from DC3R2 as a function of color space (right). In all panels, cells shown as white have incomplete spectroscopic coverage: they lack objects with spec-z's, but there are photometric objects in these cells. **If we cannot measure spec-z's for these cells, then the inferred redshift distributions of wide-field galaxies will be biased, impacting the 3×2 point analysis.** Much of color space at $z = 1.5$–2.5 lacks spectroscopic coverage (i.e., white cells near teal/green/yellow-green cells in the left two panels). Though DC3R2 is relatively shallow compared to the proposed depth of SuPR, we can see that DESI is highly complementary to PFS because DESI can obtain spec-z's for objects at $z < 1.5$, allowing our campaign to ideally focus on higher redshift objects whose [OII]λ3727 lines or 4000 Å break fall outside of DESI's wavelength coverage. Adapted from McCullough et al. (2024) Figs. 2 and 3.

We therefore consider it possible that some—perhaps 10k–20k in optimistic scenarios—of the 20,000–30,000 spectra needed for Roman photo-z work could come from either DESI or PFS programs in support of Rubin. These programs will still not be sufficient for our needs, both because they would be i-limited (when we need H-limited samples) and because DESI in particular has no infrared spectroscopic capabilities, limiting its utility at $z = 1.5 - 3$. In contrast, Subaru can provide redshifts via the [O II] 3727 Angstrom doublet to $z \sim 2.35$, and via UV absorption lines at higher redshifts thanks to the long exposure times we would use (with Lyman alpha potentially contributing as well). If other broad spectroscopic training sets optimized for Rubin are available, we would design PFS samples to be complementary to those efforts and in that way reduce the total sample of objects for which Roman–Subaru spectroscopy is needed. In that case, we would opt to target fewer total PFS pointings (sparsely covering the deep drilling fields to still limit sample/cosmic variance) but use longer exposure times to raise redshift success rates.

4.8 Complementarity to other Roman–Subaru programs

We are aware of two other proposals that will focus on deep PFS spectroscopy, one focused on obtaining redshifts for Type Ia Supernovae and the other on spectroscopy of high-redshift galaxies, and have been in communication with the relevant teams. We summarize the synergies and antisynergies with those proposals below.

Supernova follow-up: Synergies with the supernova-focused proposal are limited due to orthogonal field constraints; the Roman High Latitude Time Domain survey fields are located at high ecliptic latitude, and are either unobservable with LSST (which plays a key role in HLWAS photo-z's) or extremely expensive to observe from Subaru (in the case of Euclid Deep Field South). Our team believes that it would be infeasible to expect, let alone guarantee, that we can obtain the required level of relative photometric calibration accuracy if we must use photometry from a instrument other than LSSTCam, all the more so if the fields are not contiguous with the HLWAS. Furthermore, for fields without LSST DDF imaging, obtaining the deep LSST DDF-depth optical imaging needed for modern photo-z redshift distribution calibration methods would be inordinately expensive in observing time, as described above.

However, some synergies remain. First of all, there will likely be supernovae detectable in the HLWAS deep drilling fields that could augment samples from the Time Domain Survey fields, and such objects (or their hosts) could likely be targeted amongst our own targets with minimal impact. Furthermore, the spectroscopy we will obtain would still be useful for improving photometric redshift and stellar mass estimates with the non-LSST photometry that would be obtained in the time domain fields, which can then be used both to identify high-z SN hosts and to map foreground lensing magnification (with dark-matter-to-stellar mass relations calibrated via galaxy-galaxy lensing) for high-z SN fields. Both of these applications will be much less sensitive to photo-z

systematics than 3x2pt analyses, so photometric calibration uncertainties should not be a significant issue (e.g., characterizing the foreground magnification to 10% of its value would still render the effect far subdominant to intrinsic luminosity variations in supernovae).

High-redshift galaxies: We anticipate that another proposal for long exposures with PFS will be submitted by a team focused on characterizing high-redshift galaxies discovered by Roman. In this case, our field requirements are likely synergistic rather than orthogonal, as the same deep Rubin + Roman photometry that we require in our photometric redshift training/characterization fields is what is needed to enable the identification of high-z galaxies. It is likely that these fields would be more optimal than the Time Domain Survey fields for this work, given that the Rubin imaging in Euclid Deep Field-South will be shallower than for other DDFs, and no comparably deep imaging will be available in the northern time domain field.

As a result, it is likely that objects for both SuPR-Deep and high-z galaxy surveys could be targeted simultaneously with different fibers; the principal challenge would be how to optimally share PFS time with targets from both teams. One could imagine, for instance, targeting the faintest/hardest objects from each sample for the maximum possible exposure time, and trading off fibers for brighter objects between surveys once sufficient S/N is obtained on them. We are very open to exploring these synergies further.

References

Bernstein, G. & Huterer, D. 2010, *M.N.R.A.S.*, 401, 1399

Blanton, M. R. 2006, *Ap. J.*, 648, 268

Boruah, S. S., Eifler, T., Miranda, V., et al. 2024, arXiv e-prints, arXiv:2403.11797

Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *Ap. J. S.*, 200, 13

Buchs, R., Davis, C., Gruen, D., et al. 2019, *M.N.R.A.S.*, 489, 820

Campos, A., Yin, B., Dodelson, S., et al. 2024, arXiv e-prints, arXiv:2408.00922

CCS Definition Committees. 2025, Interim Report: The Community Definition of Roman's Core Community Surveys, `https://asd.gsfc.nasa.gov/roman/comm_forum/forum_17/Core_Community_Survey_Reports-rev03-compressed.pdf`, accessed: February 25, 2025

Driver, S. P., Andrews, S. K., Davies, L. J., et al. 2016, *Ap. J.*, 827, 108

Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, *A. J.*, 155, 1

Greene, J., Bezanson, R., Ouchi, M., Silverman, J., & the PFS Galaxy Evolution Working Group. 2022, arXiv e-prints, arXiv:2206.14908

Gruen, D. & Brimioulle, F. 2017, *M.N.R.A.S.*, 468, 769

Gruen, D., McCullough, J., Amon, A., et al. 2023, The Messenger, 190, 28

Hearin, A. P., Zentner, A. R., Ma, Z., & Huterer, D. 2010, *Ap. J.*, 720, 1351

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, *M.N.R.A.S.*, 465, 1454

Khostovan, A. A., Kartaltepe, J. S., Salvato, M., et al. 2025, arXiv e-prints, arXiv:2503.00120

Ma, Z. & Bernstein, G. 2008, *Ap. J.*, 682, 39

Masters, D., Capak, P., Stern, D., et al. 2015, *Ap. J.*, 813, 53

Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, *Ap. J.*, 841, 111

Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, *Ap. J.*, 877, 81

McCullough, J., Gruen, D., Amon, A., et al. 2024, *M.N.R.A.S.*, 531, 2582

Myles, J., Alarcon, A., Amon, A., et al. 2021, *M.N.R.A.S.*, 505, 4249

Newman, J., Blazek, J., Chisari, N. E., et al. 2019, *B.A.A.S.*, 51, 358

Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015a, Astroparticle Physics, 63, 81

Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015b, Astroparticle Physics, 63, 81

Newman, J. A. & Gruen, D. 2022, *A.R.A.A.*, 60, 363

Stanford, S. A., Masters, D., Darvish, B., et al. 2021, *Ap. J. S.*, 256, 9

The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv e-prints, arXiv:1809.01669

To, C.-H., DeRose, J., Wechsler, R. H., et al. 2024, *Ap. J.*, 961, 59

Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *Ap. J. S.*, 258, 11