

Training for Big Astronomical Data with Keck-FOBOS: Comprehensive, Data-Driven Models of a Universe in Transition

K. Bundy, K. Westfall, N. MacDonald, P. Capak, A. Coil, C. Conroy, M. Cooper, R. Kupke, K.G. Lee, R. Mandelbaum, D. Masters, J. A. Newman, X. Prochaska, C. Rockosi, J. Rhodes, M. Rich, M. Savage, A. Shapley, B. Siana, Y.-S. Ting, G. Wilson

[[10-page limit, excluding references.]]

1. INTELLECTUAL MERIT

1.1. Scientific Justification. Led by NSF’s Large Synoptic Survey Telescope (LSST¹), astronomy is entering a new era of unprecedented deep-imaging data sets that will survey huge volumes of the universe when it was only one-half or one-third its current age ($z \sim 1-3$). These epochs mark important but poorly understood transitions in cosmic history. Early galaxies were emerging from a “primordial soup” of gas and dust, assembling now-fossilized structures that may persist even within our own Milky Way. Meanwhile, the rate of cosmic expansion was beginning to accelerate, as the Universe became increasingly dominated by “Dark Energy,” whose origin remains the single greatest mystery in astronomy and cosmology today.

Since Edwin Hubble’s observations over 100 years ago, major advances in our understanding of the universe have come from the two-step process of first taking images of the sky to locate sources of interest and then obtaining information-rich spectroscopy to reveal the nature of those sources. A modern example is the Sloan Digital Sky Survey (SDSS) whose combination of panoramic broad-band “imaging” followed by dedicated spectroscopy yielded unprecedented in-depth data on over 1 million galaxies, mapping the present-day universe and making SDSS one of the most highly cited surveys in the history of astronomy.

LSST’s all-sky images will be 1,000 times deeper and detect far more distant galaxies than SDSS, but **no current U.S. facility is capable of obtaining spectroscopic followup of LSST galaxies** at a level required to capitalize on the \$1B U.S. investment in that project. In fact, an SDSS-like spectroscopic study of 1 million galaxies at LSST depth would require 300 years of observing on the largest telescopes with current instrumentation!

The only way forward is encapsulated in one of NSF’s “10 Big Ideas,” *Harnessing the Data Revolution*: we can maximize the information content of LSST and other imaging facilities via machine learning from optimally-designed spectroscopic training sets. This proposal presents a coordinated framework with three critical components necessary for success in this endeavor: 1) Using simulated spectroscopic+imaging data to define the training sets required to address ambitious data-science challenges in Cosmology, Galaxy Formation, and Local Group Archeology in the LSST era; 2) Preliminary design of Keck-FOBOS², a state-of-the-art spectroscopic facility on one of the world’s largest telescopes optimized for providing the required training sets; 3) Preliminary design of the coordinated Keck-FOBOS observations required as well as the systems needed to publicly deliver training set data products. This MSRI-1 design proposal lays out the path for maximizing panoramic imaging from LSST, WFIRST³, Euclid⁴, and other facilities with unparalleled deep and high-sampling density spectroscopic followup. Through a subsequent MSRI proposal we will deliver on our goals with an instrument deployment in 2026, an array of spectroscopic programs, and associated public-ready training sets.

¹LSST will begin science operations in 2023.

²Keck-FOBOS: The Keck Observatory Fiber Optic Broadband Optical Spectrograph

³WFIRST is NASA’s space-based Wide-Field Infrared Survey Telescope, expected to launch in the mid 2020’s.

⁴Euclid is led by the European Space Agency with significant NASA involvement and will launch in 2021. Its primary mission is a 15,000 deg² imaging survey in optical and near-IR wavebands.

1.2. Research Community Priority. The need for spectroscopic followup in the LSST era was made clear in the National Research Council’s 2015 report, “Optimizing the U.S. Ground-Based Optical and Infrared Astronomy System” (Council, 2015) which recommended:

The National Science Foundation should support the development of a wide-field, highly multiplexed spectroscopic capability on a medium- or large-aperture telescope in the Southern Hemisphere to enable a wide variety of science, including follow-up spectroscopy of Large Synoptic Survey Telescope targets. Examples of enabled science are studies of cosmology, galaxy evolution, quasars, and the Milky Way.

In addition to this report, further details of spectroscopic needs for LSST in all science areas were disseminated after a 2013 workshop on this topic organized by the National Optical Astronomy Observatory (NOAO). **JAN: I think at least as relevant is the NSF-requested Kavli/NOAO/LSST report, <https://www.noao.edu/meetings/lsst-oir-study/>, which followed up on the Elmegreen report.** Based on these recommendations, we propose the Keck-FOBOS instrument coupled with a suite of data-driven tools to address the spectroscopic requirements of LSST and other photometric surveys at a final cost 20 times less than a new Southern Hemisphere facility. Located in Hawaii, Keck-FOBOS would have access to more than 70% of the LSST footprint, more than adequate for our primary goal of building powerful spectroscopic training sets. Compared to Prime Focus Spectrograph (PFS) on Japan’s Subaru Telescope, Keck-FOBOS would be $1.7\times$ faster, provide UV sensitivity with a wavelength range of 310–1000 nm (PFS covers 380–1250 nm), and offer high-density and more flexible target sampling over “deep-drilling” fields. Keck-FOBOS would be operated on a U.S. telescope with dedicated U.S. access and a commitment to supporting U.S.-led photometric surveys. FOBOS is also complementary to future ambitious facilities that would be optimized to cover wider areas (several deg^2 per pointing) at shallower depths.

The need for deep spectroscopic followup is particularly acute for LSST’s major cosmological probes which rely on “photometric redshifts:” measures of the redshifts of objects – which indicate how far back in time and space we are looking – based on imaging alone. Newman et al. (2015) summarize the case for this and describe a redshift survey which, if carried out with Keck-FOBOS, would increase LSST’s Dark Energy Figure-of-Merit by a factor of 40% at a cost of less than 5% of the LSST budget. The urgent case for spectroscopic redshift training has been the subject of numerous publications (e.g., Laureijs et al., 2011; Masters et al., 2015; Hemmati et al., 2018).

Meanwhile, the astronomy community recognizes that the coming era of “Big Data” astronomy culminating in LSST necessitates “harnessing the data revolution.” Widespread community interest in advanced data science techniques continues to grow amidst calls for educational programs, conference series, and research funding to support the growth of a new field, “Astroinformatics,” which exploits the interface between astrophysics and statistics (Borne et al., 2009). Astronomy’s largest organizations, including the American Astronomical Society and the International Astronomical Union, have supported active working groups on astroinformatics and astrostatistics since 2015. LSST itself has built the Informatics and Statistics Science Collaboration and partnered with NSF to fund the Data Science Fellowship Program to train astronomy graduate students in data science techniques. Our proposal builds on and contributes to these ongoing efforts.

1.3. Science Goals and Data Science Challenges. We identify ambitious “data science challenges” for the LSST era that would address major goals within each of three core topics. By simulating future wide-field imaging data as well as Keck-FOBOS spectroscopy, we will develop

astrostatistics techniques and applications over the proposal period that will refine the Keck-FOBOS instrument requirements, inform the emerging design and operational modes, and define required training sets. Tackling these challenges requires a community-wide effort and will deliver wide-spread benefits. Our specific purpose with this proposal is to establish community priorities and success metrics and to coordinate the various groups working in this area—many represented among our Senior Personnel.

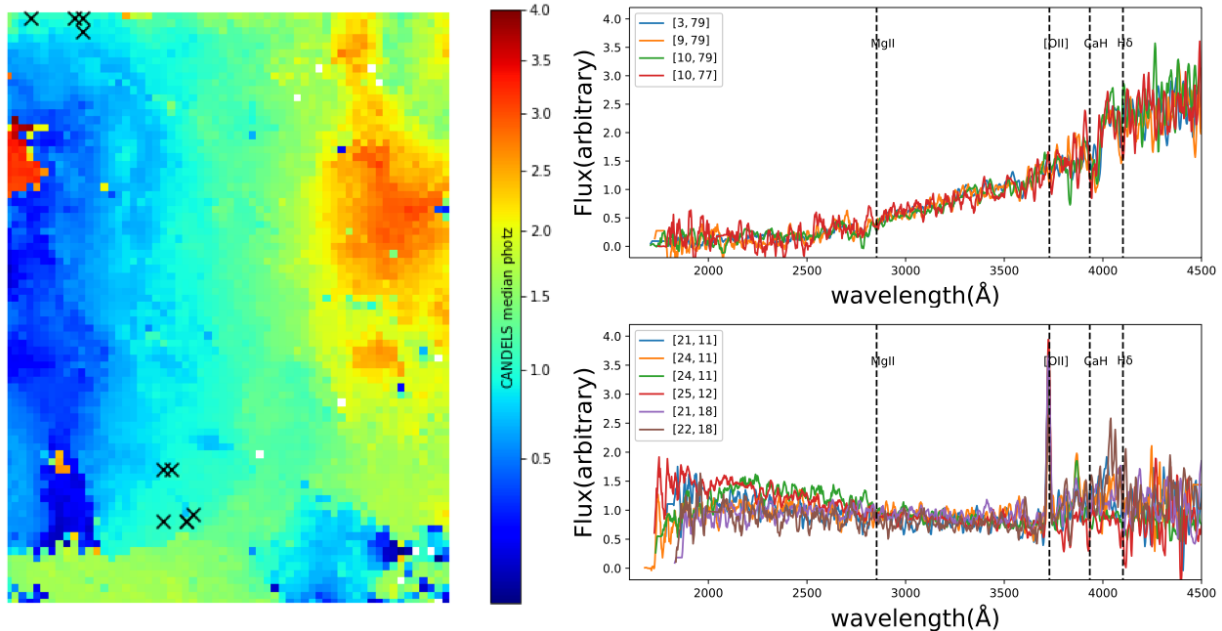


FIGURE 1. *Left:* A Self-Organizing Map (SOM) from Hemmati et al. (2018) visualizing the relationship between galaxy brightness in different broadband filters (projected into a two-dimensional space) and observed spectroscopic redshift (indicated by the color map). SOMs guide the optimal construction of training samples by highlighting which galaxy classes require targeting. *Right:* The spectra associated with localized SOM regions have similar spectra, as well as similar redshifts. **JAN: Contra the previous text, I would say that the similarity of the spectra is neither remarkable nor surprising, since to be assigned to the same cell the galaxies have to have extremely similar SEDs, and hence spectra. A bigger open question is can you get similar apparent SEDs from multiple redshifts (e.g., are there degeneracies with other parameters); the SOM is not necessarily one-to-one with redshift, but should be deterministic of spectral shape in any event.**

1.3.1. *Enhancing Dark Energy Probes via Precision Cosmic Distances.* The 2011 Nobel Prize in Physics was awarded for the discovery that the expansion of the universe has been accelerating instead of slowing down due to gravity as previously expected, starting when it was roughly half its current age. This accelerated expansion is often attributed to a mysterious “Dark Energy,” the origin of which remains unknown.

Dark Energy is perhaps the single most important unsolved problem in both cosmology and particle physics. As such, it has inspired enormous world-wide effort and the construction of dedicated ground and space-based facilities. These include LSST, Euclid, and WFIRST. The goal of these experiments is the precision mapping of cosmic structure. Because this structure grows as the universe does, these measurements allow us to reconstruct the cosmic expansion history and distinguish amongst Dark Energy models.

Most cosmological probes require measurements of galaxy “redshifts,” the Doppler-like shift in the observed wavelengths of emitted light; like structure on the largest scales, the wavelengths of light also grow with the cosmic expansion as it travels to us. By comparing the relationship between redshifts and other observable quantities (such as angles or brightnesses) to expectations from cosmological models (including dark energy parameters), we can constrain the nature of our Universe. Accurate redshifts are conventionally derived via spectroscopy by comparing the observed wavelengths of spectral features to the characteristic wavelengths of those features where light was emitted. Less precise redshifts, called “photometric redshifts” (or photo- z s), can be obtained from imaging (a.k.a. “photometric”) data alone, a compelling option for large and faint data sets where spectroscopic redshifts are infeasible. However, “to infer cosmological parameters not limited by systematic errors, accurate redshift measurements are needed” (Hemmati et al., 2018).

Spectroscopic redshifts are critical for both the training of photometric redshift algorithms and for calibrating results in order to correct for biases. Complete photo- z training samples can *increase the Dark Energy Figure of Merit in LSST by 40%* (Newman et al., 2015) and Keck-FOBOS is particularly powerful in this respect because it has no redshift desert. Keck-FOBOS can measure spec- z s above $z > 1.5$ via rest-frame UV features, eliminating the need for expensive, space-based near-IR spectroscopy.

Data-Science Challenge 1: Enable High-precision LSST Photometric Redshifts ($\sigma_z/(1+z) \lesssim 0.02$ at $i(\text{AB}) < 25.3$) with Targeted Training Spectroscopy. Delivering optimal photometric redshifts with minimal errors per object will require sets of $> 10,000$ spectra for training purely machine learning-based algorithms or optimizing our knowledge of galaxy spectra and calibration errors for template-based and hybrid algorithms. Our proposed FOBOS instrument is ideally suited to providing this training set, considering the requirements in Newman et al. (2015). Our goal is to design and deliver an optimized set of spectroscopic redshifts which will enable photometric redshifts to be accurately ‘painted on’ to LSST imaging-only objects, and thereby improve both dark energy and galaxy evolution science from LSST.

1.3.2. *A Comprehensive Picture of the Proto-galaxy Ecosystem.* [[1 page]]

Roughly 4 billion years after the Big Bang ($z \sim 2$), the universe entered a key epoch in which proto-galaxies transitioned from interacting, gas-rich systems into the more ordered, star-dominated structures that populate the universe today. This period marks the peak of global star formation rate and galaxy assembly history. To understand it, we must not only study the galaxy population at this epoch but the entire galaxy “ecosystem” which includes their gas-filled environments. The goal is to build a comprehensive picture of the physical processes that fuel proto-galaxy growth, shape their internal structure, and influence their environment.

LSST’s panoramic imaging will detect huge numbers of galaxies at this epoch. Targeted followup with Keck-FOBOS will allow us to ascribe detailed galaxy and environmental information from deep spectroscopic training samples to the much larger cosmic volumes surveyed with broad-band imaging.

Data-Science Challenge 2: Apply Deep Learning to infer star formation rates and formation histories, dust content, wind properties, and stellar masses from $z \sim 2$ photometry. The range of observed spectral types is remarkably constrained by broad-band imaging (Figure 1, right panel), suggesting a far greater potential for imaging data to reveal physical properties with sufficient training than conventional modeling of spectral energy distributions (SEDs) would suggest. Applying machine learning, our challenge is to deliver SDSS-like information for millions of imaged galaxies at $z \sim 2$. With simulated data sets, we will

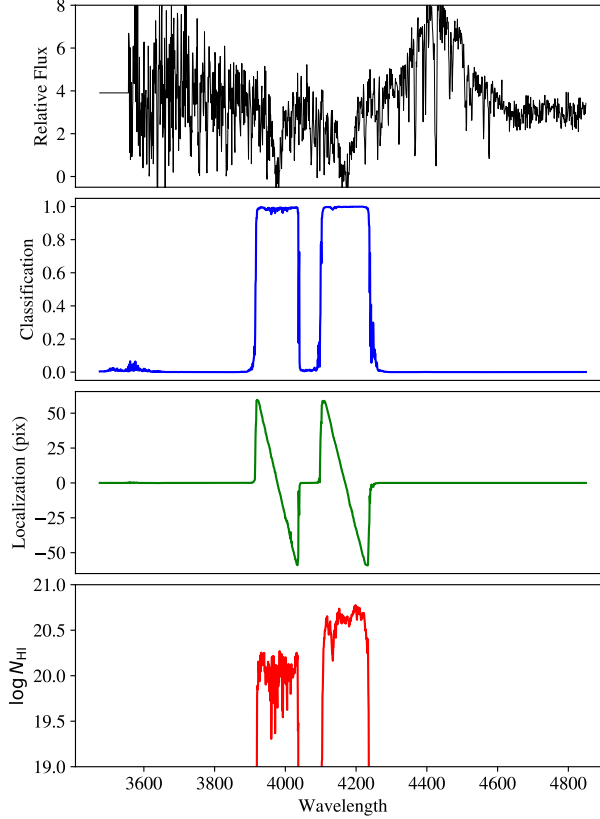


FIGURE 2. Example of machine learning applied to absorption features in rest-frame UV spectroscopy to detect two intervening gas clouds along the line-of-sight (from Parks et al., 2018). A spectrum is shown at top with lower panels indicating associated labels conveying physical information. Keck-FOBOS will provide a rich set of similar features and the opportunity to transfer labels to imaging and higher redshift data sets.

investigate derived uncertainties and biases and explore benefits from incorporating additional imaging information like morphology, structure, and size from a wide range of wave-bands (e.g., LSST plus Euclid plus WFIRST). The exercise will define requirements for Keck-FOBOS instrument performance and the FOBOS Public Survey design.

Data-Science Challenge 3: Enable label transfer from rest-frame optical to UV stellar and ISM indicators. There are many powerful gas and stellar spectral features just redward of the Lyman- α line at 1216 Å. By combining Keck-FOBOS UV and near-IR spectroscopy (e.g., from PFS) at $z \sim 2$, we can transfer “labels” best modeled in the rest-frame optical to spectra at UV wavelengths, at least for certain types of galaxies. This “label transfer” will dramatically enhance interpretation of JWST discoveries of the first galaxies ($z \sim 10$) for which rest-frame UV imaging and spectroscopy will be most accessible. A similar application can ascribe the escape fraction of Lyman continuum radiation observed in the Keck-FOBOS Public Survey to constrain the sources responsible for “reionization” at $z \sim 6$. With simulated spectral observations, we will determine the extent of label transfer that is possible and set requirements on training samples.

Data-Science Challenge 4: Train short spectroscopic exposures in combination with LSST photometry to provide environmental diagnostics for 1M galaxies at $z = 1-2$. Photometric redshifts, while acceptable in large cosmological analyses, wash out information about the local position of galaxies with respect to one another. To characterize a galaxy’s

local environment and identify its neighbors requires (observationally expensive) spectroscopic redshifts (spec-zs). However, with improved photometric redshifts available from Challenge 1 and strong priors on spectral types (Challenge 2), machine learning techniques can yield *spectroscopic* redshifts at much lower signal-to-noise than conventional redshift measurements. Specifically, our challenge is to develop a methodology that can measure 300 km s^{-1} accuracy spec-zs on spectra obtained in just 10 minutes with Keck-FOBOS. This would enable an SDSS-like environmental study of 1M galaxies at $z = 1\text{--}2$ in just 20 nights of 10 m telescope time, making it a compelling sub-component of the Keck-FOBOS Public Survey.

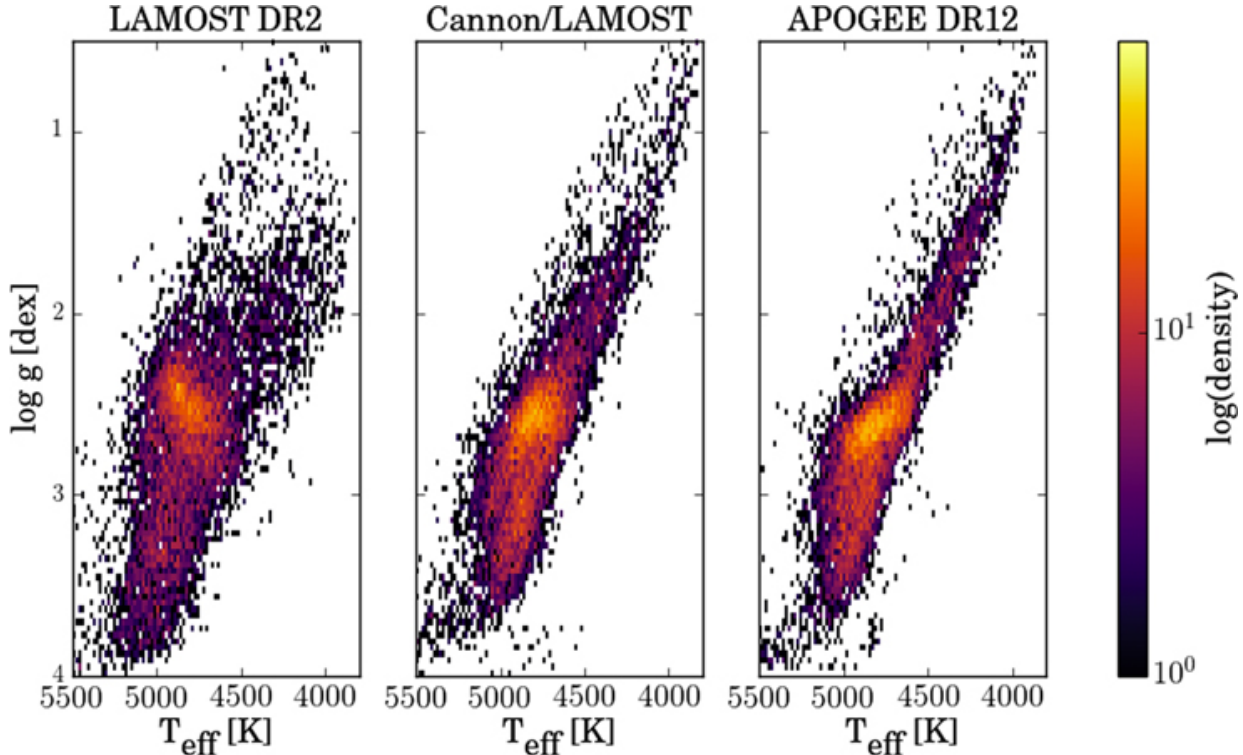


FIGURE 3. Verification of application of *The Cannon* to LAMOST spectra based on stellar parameters determined by the high-resolution APOGEE data from Ho et al. (2017). Each panel shows the derived effective temperature, T_{eff} , and surface gravity, $\log g$, for each star, with the color representing the density of stars at each position. The left panel shows the results for the LAMOST spectra using a direct fitting approach, the right panel shows the results derived from the high-resolution APOGEE data, and the middle panel shows the results of using *The Cannon* to determine the stellar parameters using the low-resolution LAMOST spectra trained by the APOGEE-derived parameters. Results from *The Cannon* are more accurate and astrophysically plausible.

1.3.3. Unraveling the Formation History of our Local Group of Galaxies. [[1 page]]

Our Local Group of galaxies — composed of the Milky Way (MW) Galaxy, the Magellanic Clouds, the nearby Andromeda (M31) and Triangulum (M33) Galaxies, and a multitude of satellite galaxies — is just one realization of the galaxy-formation process, but it is the one that we can study in the greatest detail. Large-scale imaging surveys executed over the past 25 years have revolutionized our census of the Local Group. In particular, SDSS and Pan-STARRS have unveiled numerous stellar streams and other halo substructures in both the MW and M31, including a stellar bridge stretching between M31 and M33. We expect a hundredfold growth

in the census of halo substructures in the MW via the upcoming LSST and WFIRST surveys. Follow-up spectroscopy of Local Group member stars allow us to, e.g., constrain the orbits of stellar streams and the present-day enclosed mass of the galaxies they orbit (refs), as well as the age and chemical composition of their stellar populations (refs). At the same time, cosmological simulations [refs], like IllustrisTNG [others], can now simulate the full chemo-dynamical evolution of Local-Group-like overdensities in the Universe to which data can be meaningfully compared. Finally, the Gaia satellite (ref) is currently revolutionizing our understanding of the MW by providing distances and on-sky motions for more than a billion stars spanning the full extent of its disk. This simultaneous maturation of both the theoretical and observational data will allow us to form physically motivated models for the formation history of the Local Group and its constituents.

As we obtain deeper images and more varied data sets, follow-up spectroscopy of Local Group objects of interest becomes ever more difficult. As it is, e.g., Keck-DEIMOS programs to measure the radial velocities of stars in the MW halo or the M31 disk require observations of up to 10 hours, depending on the population being probed [Tollerud, Dorman, Cunningham]. Given such long integration times, one approach is to maximize the number of targets observed in a single pointing. However, one can also appeal to machine-learning algorithms to infer the relevant physical quantities statistically from both multi-band imaging and lower quality spectra (low resolution and S/N) using a relatively small, yet high-S/N, training set. There has been a significant push over the past 5 years toward building such machine-learning applications.

For example, [Ness+] have developed *The Cannon*, a supervised learning algorithm that uses spectra with known stellar parameters to label spectra where those parameters are unknown. In one application, they determined three fundamental parameters for 55000 APOGEE spectra using a 1% training sample. Additionally, [Ting+] have developed *The Payne*, which uses a neural network and theoretical stellar spectra to determine 25 stellar-abundance labels providing the detailed chemical make-up of each observed star. Example applications of these techniques are shown in Figure X.

Our proposed effort builds on new lines of inquiry based on these successes, both in terms of application of these machine-learning techniques to new data sets and development of new techniques as we discuss below.

Data-Science Challenge 5: The chemical evolution and assembly history of the MW stellar halo. LSST and WFIRST will reveal a trove of substructure in both the MW and M31 halo. We will design an observational program that would employ Keck-FOBOS to observe main-sequence turn-off and red-giant stars in these substructures within the MW. These and additional data (APOGEE, H3) will be used as training sets to build data-driven models for the stellar parameters (temperature, surface gravity, metallicity) for all halo stars with LSST+2MASS+WISE+WFIRST multi-band photometry. These will be combined with dynamical data and compared with cosmological simulations to build a generative model for the assembly history of the MW stellar halo.

Data-Science Challenge 6: The differential chemical evolution of M31 and MW. A natural extension of Data-Science Challenge 6 is to perform the same analysis for the halo of M31. However, we cannot expect to obtain high-quality spectra of individual main-sequence stars at the distance of M31 with Keck-FOBOS. Moreover, training a chemical evolution model based on training sets composed of stars in the Milky Way could lead to systematic errors: The Milky Way and Andromeda have distinct chemical evolution histories (ref), despite being relatively similar in many other respects. We will therefore design an observational program that will obtain deep observations of giant stars in the M31 halo. These data will drive a machine-learning algorithm

that combines a model of the MW halo with results from cosmological hydrodynamical simulations to constrain the differential history of the MW and M31 stellar halos.

Data-Science Challenge 7: Stellar parameter determinations for a billion stellar spectra. While providing on-sky motions and photometry for 1.7 billion stars in the MW, fewer than 10% of stars will have a full complement of three-dimensional space motions, fewer than 0.3% will have basic stellar parameters, and only 0.1% will have measured chemical abundances. In addition, Gaia distance measurements have errors that increase quadratically with distance. To realize the full potential of the Gaia astrometric catalog, one needs 3D position and 3D velocity vectors and chemical abundances for each star. We will therefore design training sets to observe with Keck-FOBOS that, when combined with existing high-resolution datasets (e.g., APOGEE, WEAVE) will allow us to build data-driven models of distance, temperature, surface-gravity, and stellar abundance for *all* stars in the Gaia dataset. These data will allow us to isolate coeval populations in the Galactic disk that can be combined with very high-resolution simulations of the Milky Way to provide a detailed evolutionary history of our Galactic home.

2. PROJECT IMPLEMENTATION

This proposal involves three coordinated activities: 1) Organizing and evaluating the results of a community-wide effort to address simulated Data Science Challenges; 2) Completing Preliminary Design for the Keck-FOBOS instrumentation, informed in part by refining requirements as a result of (1); 3) Designing the operational modes, planning tools, data analysis software, and serving platforms necessary for delivery of public training sets. Anticipating significant progress in all three activities, we will request NSF MSRI-2 funding in 2021 to build and deploy Keck-FOBOS at the telescope, carry out required observations, and publicly serve the data products. FOBOS would see first light in 2027 and carry a total cost of \$32M (without contingency in 2019 dollars). While we focus the current request on work required for the Preliminary Design Phase, we outline the overall project plan and final deliverables in order to motivate this work.

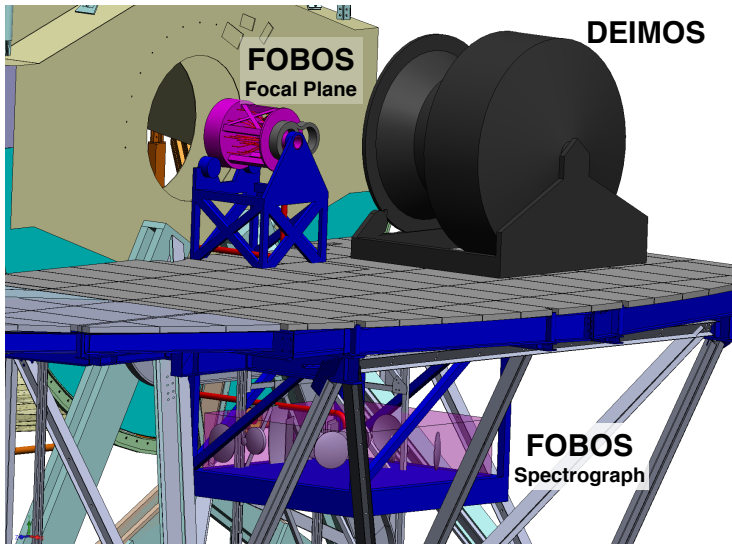


Figure ??: Rendering of FOBOS instrument systems deployed at the Keck II Nasmyth port. By mounting the FOBOS spectrographs under the Nasmyth platform, other instruments like DEIMOS can maintain access to the telescope.

2.1. Keck-FOBOS Instrument Concept. Mounted at the Nasmyth focus of Keck II Telescope at WMKO⁵, the Fiber Optic Broadband Optical Spectrograph (FOBOS) will be one of the most powerful spectroscopic facilities in the next decade. FOBOS consists of several key components

⁵WMKO: William M. Keck Observatory operates the two twin 10 m Keck Telescopes on Mauna Kea, Hawaii.

(Fig ??). A compensating lateral atmospheric dispersion corrector (CLADC, not pictured) ensures that target light from all wavelengths falls on allocated fibers while also correcting image aberrations at the edges of the 20 arcmin diameter Keck field. Each of the CLADC lenses is 946 mm in diameter, the first two closely spaced with lateral relative motions enabled by three barrel-mounted actuators. The final CLADC lens surface serves as the vertical mounting plate for roaming Starbugs fiber positioners. It translates to track focal plane tilt. Starbugs patrol a large on-sky area (~ 1 arcmin), enabling flexible and dynamic targeting configurations with adjacent fibers as close as 10 arcsec.

A total of 1800 150 μm core diameter fibers are deployed at the curved focal plane, which rotates and translates to maintain image positions as the telescope tracks across the sky. The fiber run is kept at less than 10 m to maintain high throughput at UV wavelengths, and special care is given to stress-relief cabling to minimize variable focal ratio degradation over the fiber run.

Sets of 600 fibers feed each of three identical spectrographs. Each spectrograph uses a series of dichroics to divide the input light into four wavelength channels with combined coverage from 310 to 1000 nm and mid-channel spectral resolutions of $R \sim 3500$. The dispersed light in each channel is focused by an f/1.1 catadioptric camera and recorded by an on-axis $4\text{k} \times 4\text{k}$ CCD mounted at the center of the first camera lens element. Spectrographs are mounted in a temperature controlled housing installed under the Nasmyth Deck to allow space for other Keck instruments above. The end-to-end instrument throughput is greater than 30% at all wavelengths.

FOBOS includes observatory level systems for precise instrument calibration using dome-interior screen illumination, a metrology system for accurate fiber positioning, and guide cameras for field acquisition and guiding. The instrument design envisions future upgrades including alternate collecting modes that deploy multiple fiber bundles, feeds to other fiber-based spectrographs at different wavelengths or spectral resolutions, and the ability to support and benefit from image corrections with Ground-Layer Adaptive Optics.

2.2. Keck-FOBOS Instrument Design Effort. Keck-FOBOS will complete its current conceptual design phase in October 2019. Funding from this proposal will support preliminary design beginning in November 2019. A schedule of milestones is attached and more information provided in the Project Execution Plan (PEP). Major components of the preliminary design effort are described below.

Atmospheric Dispersion Compensator (ADC). The opto-mechanical design, tolerancing, lens cell design, motion systems, and software controls design of the ADC will be completed.

Focal Plane System. The final ADC lens element serves as the focal plane mounting plate for the fiber positioners. This focal plane system must rotate and translate to track the field and refraction angles from the ADC. Mechanical design, including flexure analysis and the selection of drive mechanisms and potential vendors will be completed. This system also defines one of the interfaces to the Keck Telescope and must comply with Keck Observatory space envelopes, servicing needs, and other requirements. The focal plane system also interfaces with guide cameras for field acquisition and guiding.

Starbugs fiber positioners. Starbugs are a positioning technology developed and deployed by the Australian Astronomical Observatory (AAO) which has partnered with our team to generate a conceptual design for Starbugs in the context of FOBOS. Design requirements for Starbugs in FOBOS are more relaxed than the currently on-sky TAIPAN instrument thanks to the larger physical plate scale at Keck. AAO will serve as a vendor during preliminary design but is interested in exploring a partnership and in-kind contribution model in the construction phase. In addition to the Starbugs themselves, a fiber metrology system (for accurate closed-loop positioning) will also be developed.

Fiber System. We will complete the optical design and processing plan for affixing forward optics lenses to each fiber’s head (these demagnify and speed up the beam for proper fiber coupling). A micro-lens array solution will be developed for a central, fixed-position 4.5-arcsec diameter IFU for fast source acquisition. This workpackage also includes the stress-relief cable system and fiber termination hardware and processing.

Spectrographs. The optical systems and components (slit, collimator, dichroics, gratings, and camera), an analysis of acceptable tolerances and performance, their mechanical supports, software controls, and the overall enclosure will all be advanced through preliminary design. Detectors, cryostats, read-out electronics and systems for thermal management will be designed.

2.3. Addressing Data Science Challenges and Designing FOBOS Training Sets. Our team includes leading experts on data science applications to astronomy and LSST specifically. We will also use our established connections to LSST’s Informatics and Statistics Science Collaboration (ISSC) to advertise, recruit, and coordinate efforts to tackle the Data Science Challenges described in Section 1.3. Our proposal request includes two open workshops to motivate progress and discuss results. At the end of the proposal period, we will publish the results and developed software packages.

The Data Science Challenges require work on simulated imaging+spectroscopic data sets where input physical properties (e.g., redshift) can be imposed and the output, recovered values compared against the input. Simulated imaging data (e.g., from LSST and WFIRST) are in-hand, while mock spectroscopy will be provided by a Keck-FOBOS instrument simulator, an initial version of which has already been developed. Further advances to be supported by this proposal include improved error modeling and simulating systematic effects from detector artifacts, image quality aberrations informed by the emerging detailed optical design, and variable observing conditions.

The resulting success in addressing each Data Science Challenge will define a level of readiness and set requirements on the associated Keck-FOBOS training sets required, including number of sources, pointings, magnitude limits, signal-to-noise thresholds, and observing conditions. Preliminary observing design and a description of required operational modes to efficiently observe these training sets will begin with this proposal. Operational modes will set requirements on target aggregation and prioritization systems, field acquisition speed, field rotation range, zenith avoidance zone, reconfiguration time, calibrations, read-out time, quicklook reduction software and processing rates. We will develop integrated program concepts that efficiently combine required observations. Detailed survey and execution plans will be completed in the next phase of this project (MSRI-2). Roughly 20% of Keck observing time is open to the public, and as in previous federally-funded projects, we fully expect that Senior Personnel at Keck institutions will be successful in collaborative efforts to secure significant amounts of additional telescope observing time to enable rapid, publicly release of training data with any proprietary period waived (e.g., Newman et al., 2013).

2.4. MAISTRO: Target Allocation with Artificial Intelligence. Powered by Starbugs fiber positioners, Keck-FOBOS will enable fast, dynamic reallocation of fibers. To efficiently determine the best options given a wide range of possible targets and desired observing outcomes, we will develop a preliminary design for MAISTRO⁶ an “artificial intelligence” (AI) targeting system that will learn optimization strategies for assigning targets from a database of overlapping observing programs with pre-defined priorities. The AI package will aggregate data quality using a quicklook reduction package, science-driven performance metrics, *and real-time assessments of the*

⁶MAISTRO: Modular Artificial Intelligence System for Target Reallocation and Observing.

observing conditions to make dynamic targeting recommendations. For example, if conditions are slightly less than optimal, MAISTRO would reconfigure Starbugs to brighter objects in a field or implement a different program prioritization. MAISTRO would incorporate updated target lists and priorities from the active observer and could easily be over-ridden at any time. Fractions of the full Keck-FOBOS multiplex might also be reserved “manual targeting” as required by the P.I.

2.5. Publicly Available Automated Data Products. The typical proprietary period for raw data acquired at Keck is 18 months. However, typical of other public surveys (e.g., SDSS), this period will be shortened to one year for the FOBOS Public Survey.

Both as part of our design effort and for long-term use, we will develop a data-reduction pipeline, building on work already done for other fiber-based observations, like SDSS and DESI. This software will provide both the quick reduction assessments needed for our dynamic targeting system and the more detailed reduction to produce the data for scientific analysis. Reduced data will be delivered to the community (e.g., via the Keck Observatory Archive) after the proprietary periods are finished for *both* PI-led and public survey observations.

Finally, we will also provide a data-analysis pipeline that provides high-level data products. There are two aspects to analysis of the data. First, we will provide software to perform the traditional measurements of properties like Doppler shift, emission-line strengths, and internal kinematics that are measured from FOBOS-observed spectra. This software will build on existing software we have built for the SDSS-IV MaNGA survey (Westfall et al.), and it will be executed for *any* data taken with FOBOS and released along with the reduced spectra. This is a substantial effort and unheard of for observations taken outside of a large-scale survey effort. Second, we will provide the results of our various machine-learning applications in the FOBOS Public Survey (e.g., the LSST-source redshifts as determined by the FOBOS observed training set).

Important to the success of both the data-reduction and data-analysis software will be PI and community involvement in their refinement to meet the needs of specific science applications. These software packages will be open source and publicly served (e.g., using GitHub).

3. BROADER IMPACTS

”include a discussion of student training, increased participation of underrepresented groups and a description of tangible benefits to the wider U.S. research community (access, data products, technology, etc.).”

3.1. Akamai: Training the next generation of Hawaiian STEM professionals. Led by the Institute for Scientist and Engineer Educators (ISEE) at UCSC, the Akamai program provides resources for STEM training of Hawaiian college students through internships and professional development courses. Akamai particularly aims to serve the Native Hawaiian population and other under-represented groups; approximately a quarter of all interns are Native Hawaiian and 38% are women. Traditionally, most Akamai interns are pursuing engineering or computer science in their undergraduate education.

ISEE and the Akamai program already have deep connections to the W. M. Keck Observatory, involving many interns in projects related to instrument and observatory development over the past decade. We aim to involve Akamai interns in the development of the FOBOS instrument design, a software-based simulator that will inform the instrument design and eventually result in a sensitivity calculator and the data-reduction pipeline, and the many machine-learning applications germane to our public survey. In fact, we have already had an Akamai intern work with us to setup a fiber test-bench at UCSC during Summer 2018. We are excited by the opportunity to

include funding for these interns as part of our proposal and continue to seek new opportunities to generate connections with the current and future Hawaiian workforce.

3.2. Investing in future educators. Also via the ISEE, we will lead Professional Development Programs (PDPs) that train graduate students to develop an inquiry based short-course on instrument development, data-reduction procedures, and machine-learning methods. The ISEE PDP began in 2001 via a grant from the NSF to the Center for Adaptive Optics (CfAO) at UCSC for training the CfAO graduate students and postdoctoral researchers, but has now expanded to include many more disciplines, departments, and international partners. The program trains graduate students and postdocs to collaboratively design a well-focused inquiry activity within a small team. The activity is conceived, developed, and tested within the team, and the program culminates in the team implementing the activity with a group of undergraduates. The program emphasizes inclusive and equitable learning environments.

3.3. Undergraduate Student Training. Multiple avenues exist within the current curriculum to include UC undergraduate students in the development of FOBOS and its public survey. At UCSC in particular, we will guide freshman and first year transfer students through two quarters in Astro 9, a recently started course that aims to introduce scientific research method early in students' tenure via timely research projects developed and led by UCSC graduate students, postdocs, and staff. Both PI Bundy and co-PI Westfall have been involved in projects over the past two years, including a project to measure the rotation curves of galaxies observed by the SDSS-IV MaNGA survey. Introducing undergraduates to astronomical instrumentation would be a unique contribution to this course.

REFERENCES

- Borne, K; Accomazzi, A; Bloom, J; Brunner, R; Burke, D; Butler, N; Chernoff, DF; Connolly, B; Connolly, A; Connors, A; Cutler, C; Desai, S; Djorgovski, G; Feigelson, E; Finn, LS; Freeman, P; Graham, M; Gray, N; Graziani, C; Guinan, EF; Hakkila, J; Jacoby, S; Jefferys, W; Kashyap; Kelly, B; Knuth, K; Lamb, DQ; Lee, H; Lored, T; Mahabal, A; Mateo, M; McCollum, B; Muench, A; Pesenson, M; Petrosian, V; Primini, F; Protopapas, P; Ptak, A; Quashnock, J; Raddick, MJ; Rocha, G; Ross, N; Rottler, L; Scargle, J; Siemiginowska, A; Song, I; Szalay, A; Tyson, JA; Vestrand, T; Wallin, J; Wandelt, B; Wasserman, IM; Way, M; Weinberg, M; Zezas, A; Anderes, E; Babu, J; Becla, J; Berger, J; Bickel, PJ; Clyde, M; Davidson, I; van Dyk, D; Eastman, T; Efron, B; Genovese, C; Gray, A; Jang, W; Kolaczyk, ED; Kubica, J; Loh, JM; Meng, XL; Moore, A; Morris, R; Park, T; Pike, R; Rice, J; Richards, J; Ruppert, D; Saito, N; Schafer, C; Stark, PB; Stein, M; Sun, J; Wang, D; Wang, Z; Wasserman, L; Wegman, EJ; Willett, R; Wolpert, R; Woodroffe, M. “Astroinformatics: A 21st Century Approach to Astronomy,” In *astro2010: The Astronomy and Astrophysics Decadal Survey*, v. 2010, 2009, p. P6. <https://ui.adsabs.harvard.edu/\#abs/2009astro2010P...6B>
- Council, NR. *Optimizing the U.S. Ground-Based Optical and Infrared Astronomy System*, The National Academies Press, Washington, DC, ISBN 978-0-309-37186-5, 2015
- Hemmati, S; Capak, P; Masters, D; Davidzon, I; Dore, O; Mobasher, B; Rhodes, J; Scolnic, D; Stern, D. “Photometric redshift calibration requirements for WFIRST Weak Lensing Cosmology: Predictions from CANDELS,” *arXiv e-prints*, 2018. <http://adsabs.harvard.edu/abs/2018arXiv180810458H>
- Laureijs, R; Amiaux, J; Arduini, S; Auguères, J; Brinchmann, J; Cole, R; Cropper, M; Dabin, C; Duvet, L; Ealet, A; et al. “Euclid Definition Study Report,” *arXiv e-prints*, 2011. <http://adsabs.harvard.edu/abs/2011arXiv1110.3193L>
- Masters, D; Capak, P; Stern, D; Ilbert, O; Salvato, M; Schmidt, S; Longo, G; Rhodes, J; Paltani, S; Mobasher, B; Hoekstra, H; Hildebrandt, H; Coupon, J; Steinhardt, C; Speagle, J; Faisst, A; Kalinich, A; Brodwin, M; Brescia, M; Cavuoti, S. “Mapping the Galaxy Color-Redshift Relation: Optimal Photometric Redshift Calibration Strategies for Cosmology Surveys,” *ApJ*, v. 813, 2015, p. 53. <http://adsabs.harvard.edu/abs/2015ApJ...813...53M>
- Newman, JA; Abate, A; Abdalla, FB; Allam, S; Allen, SW; Ansari, R; Bailey, S; Barkhouse, WA; Beers, TC; Blanton, MR; Brodwin, M; Brownstein, JR; Brunner, RJ; Carrasco Kind, M; Cervantes-Cota, JL; Cheu, E; Chisari, NE; Colless, M; Comparat, J; Coupon, J; Cunha, CE; de la Macorra, A; Dell’Antonio, IP; Frye, BL; Gawiser, EJ; Gehrels, N; Grady, K; Hagen, A; Hall, PB; Hearin, AP; Hildebrandt, H; Hirata, CM; Ho, S; Honscheid, K; Huterer, D; Ivezić, Ž; Kneib, JP; Kruk, JW; Lahav, O; Mandelbaum, R; Marshall, JL; Matthews, DJ; Ménard, B; Miquel, R; Moniez, M; Moos, HW; Moustakas, J; Myers, AD; Papovich, C; Peacock, JA; Park, C; Rahman, M; Rhodes, J; Ricol, JS; Sadeh, I; Slozar, A; Schmidt, SJ; Stern, DK; Anthony Tyson, J; von der Linden, A; Wechsler, RH; Wood-Vasey, WM; Zentner, AR. “Spectroscopic needs for imaging dark energy experiments,” *Astroparticle Physics*, v. 63, 2015, p. 81–100. <http://adsabs.harvard.edu/abs/2015APh....63...81N>
- Newman, JA; Cooper, MC; Davis, M; Faber, SM; Coil, AL; Guhathakurta, P; Koo, DC; Phillips, AC; Conroy, C; Dutton, AA; Finkbeiner, DP; Gerke, BF; Rosario, DJ; Weiner, BJ; Willmer, CNA; Yan, R; Harker, JJ; Kassin, SA; Konidaris, NP; Lai, K; Madgwick, DS; Noeske, KG; Wirth, GD; Connolly, AJ; Kaiser, N; Kirby, EN; Lemaux, BC; Lin, L; Lotz, JM; Luppino, GA; Marinoni, C; Matthews, DJ; Metevier, A; Schiavon, RP. “The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts,” *ApJS*, v. 208, 2013, p. 5. <http://adsabs.harvard.edu/abs/2013ApJS..208....5N>

Parks, D; Prochaska, JX; Dong, S; Cai, Z. “Deep learning of quasar spectra to discover and characterize damped Ly α systems,” MNRAS, v. 476, 2018, p. 1151–1168. <http://adsabs.harvard.edu/abs/2018MNRAS.476.1151P>

4. FACILITIES, EQUIPMENT, AND OTHER RESOURCES

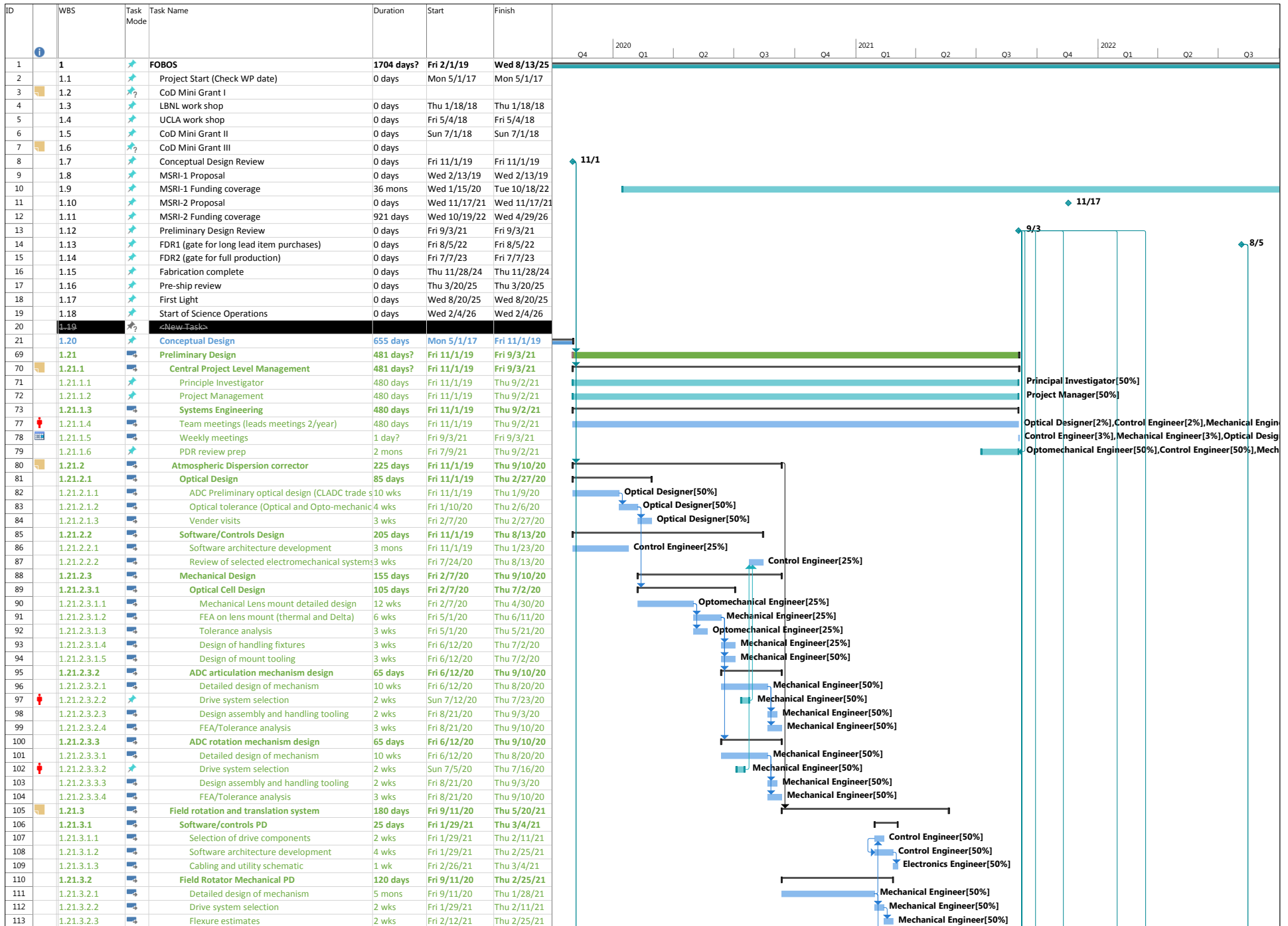
UC Observatories (UCO) manages a world-renown facility, also on the UCSC campus, for the design, construction, and testing of astronomical instrumentation. UCO a long heritage of producing state-of-the-art instrumentation, including many spectrometers, for the Lick and Keck Observatories. This long track record is thanks to the talented staff of optical designers, engineers, and instrument scientists that we seek to leverage in this project. Specific equipment and expertise of relevance includes optical fiber manipulation tools, precision machining equipment and opto-mechanical design, optics and detector design expertise.

This proposal effort also relies on the generous support of Keck Observatory which has funded the initial stages of FOBOS development and provided engineering and technical guidance on the instrument design and its interface to the observatory.

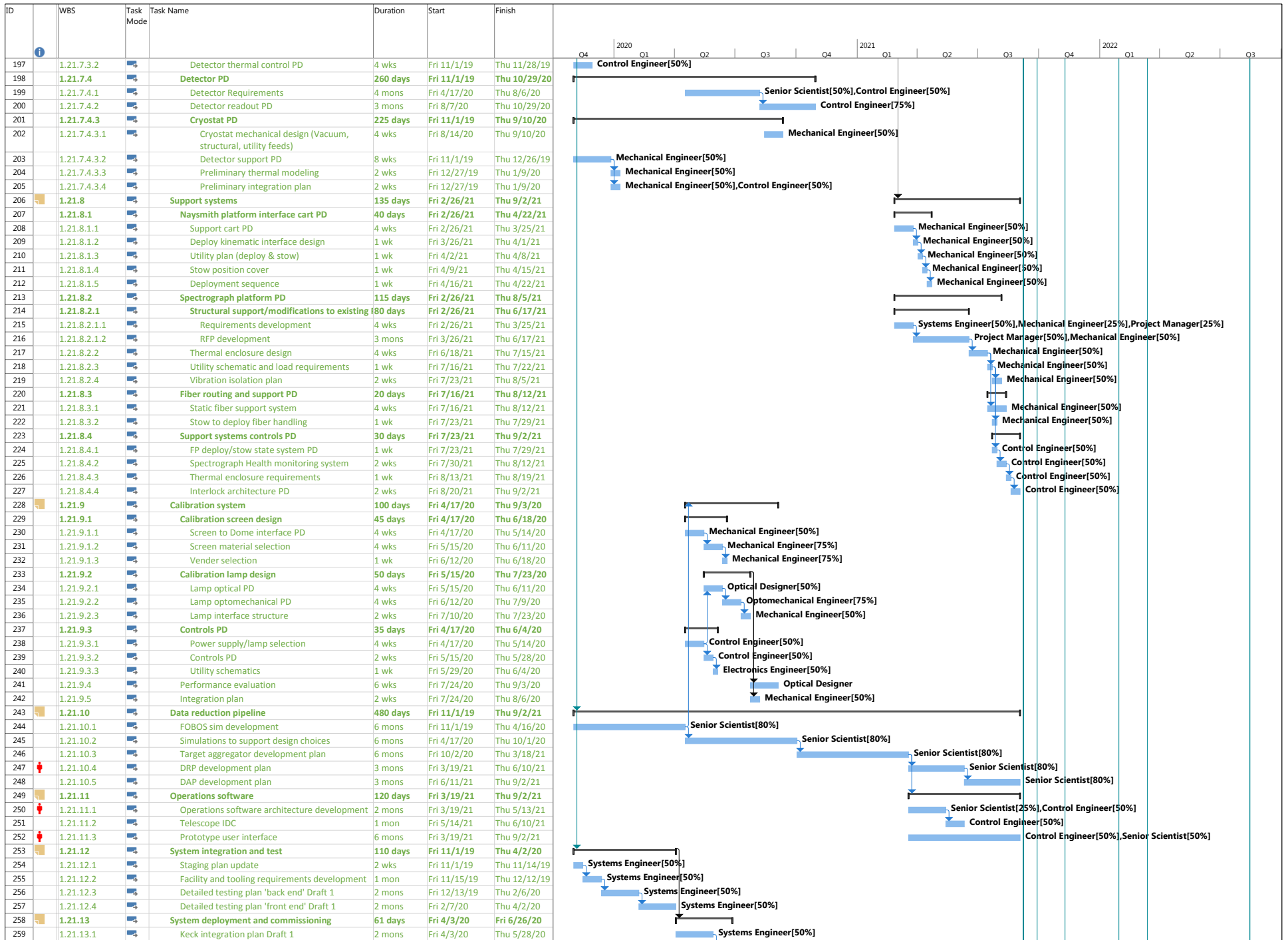
Supplementary Documents:

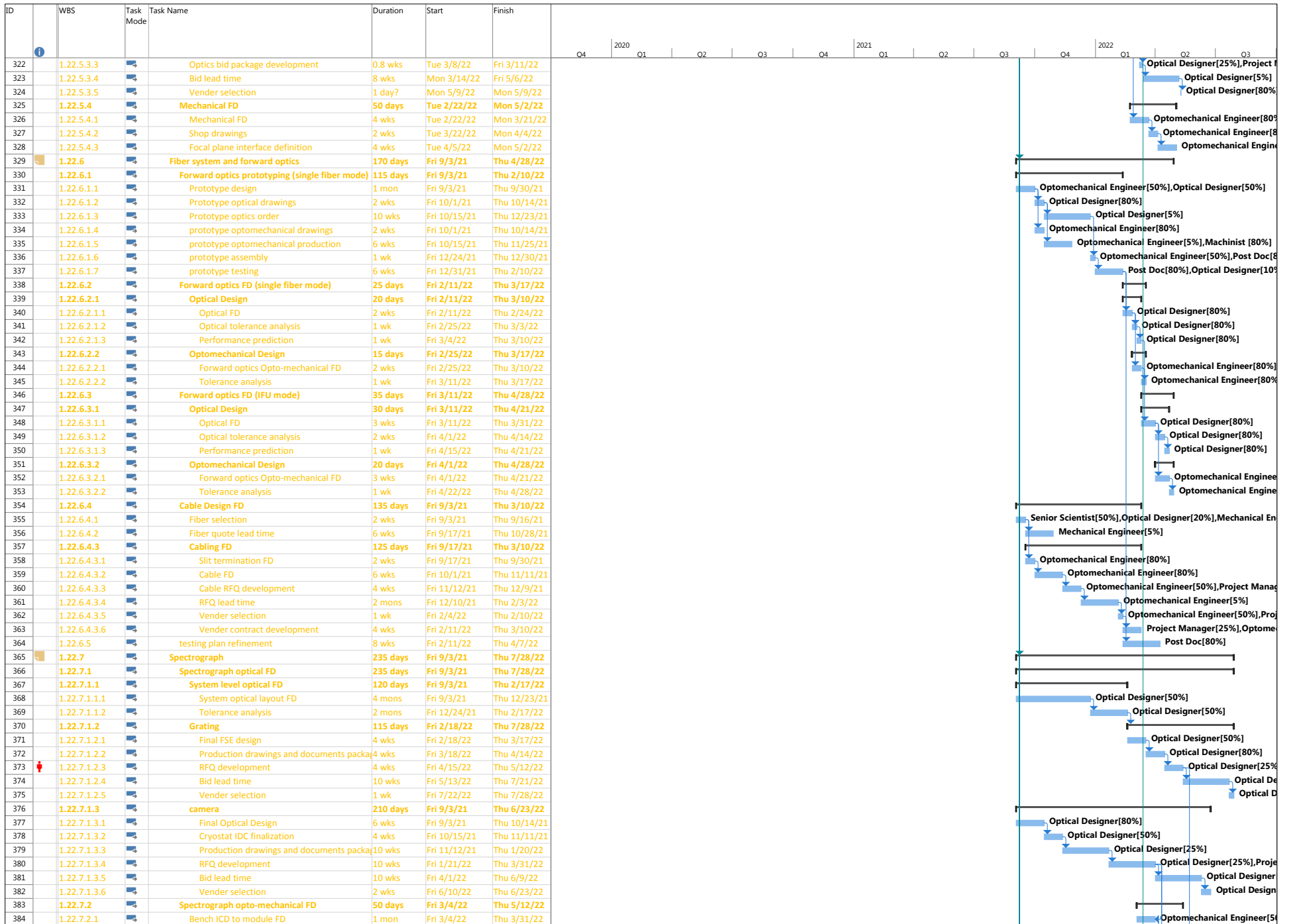
(to be entered in the Supplementary Documents section of FastLane)

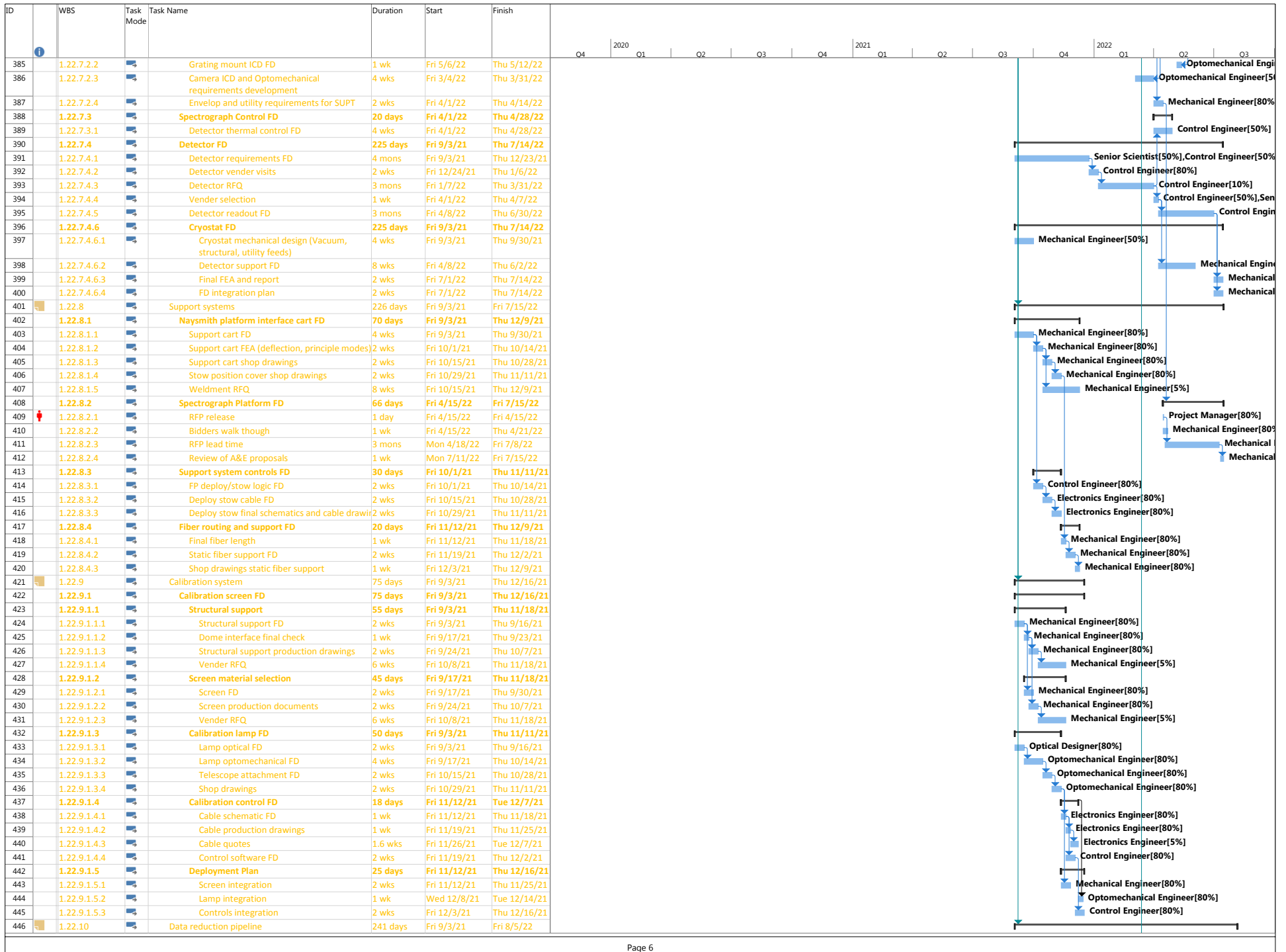
- (1) A list of the major team members, their affiliations, and their role in the project;
- (2) A list of Partner Organizations to be funded via subawards, and the role of each in the project;
- (3) An outline of the Project Execution Plan (PEP).

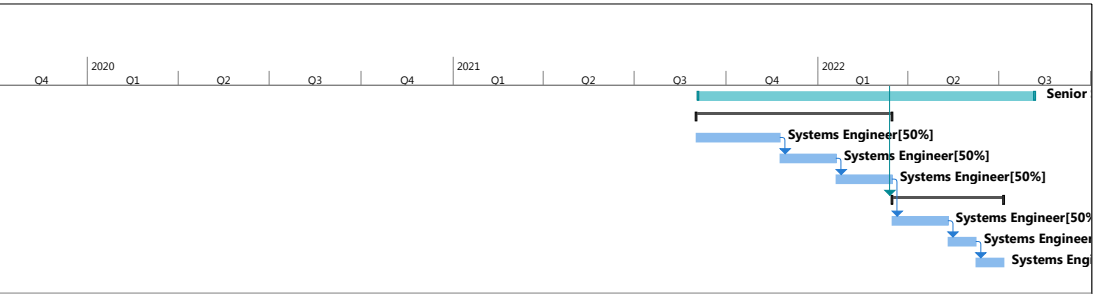










ID		WBS	Task Mode	Task Name	Duration	Start	Finish												
447		1.22.10.1		Data Lead (design choice support)	241 days	Fri 9/3/21	Fri 8/5/22												
448		1.22.11		System integration and test	140 days	Wed 9/1/21	Tue 3/15/22												
449		1.22.11.1		Facility upgrade requirements	3 mons	Wed 9/1/21	Tue 11/23/21												
450		1.22.11.2		Back end final integration plan	2 mons	Wed 11/24/21	Tue 1/18/22												
451		1.22.11.3		Front end final integration plan	2 mons	Wed 1/19/22	Tue 3/15/22												
452		1.22.12		System deployment and commissioning	80 days	Wed 3/16/22	Tue 7/5/22												
453		1.22.12.1		Keck integration plan FD	2 mons	Wed 3/16/22	Tue 5/10/22												
454		1.22.12.2		FOBOS Commissioning plan FD	1 mon	Wed 5/11/22	Tue 6/7/22												
455		1.22.12.3		Commissioning acceptance matrix FD	1 mon	Wed 6/8/22	Tue 7/5/22												
456																			