This Mid-scale Research Infrastructure-1 Design proposal lays the groundwork for dramatically enhancing the power of upcoming imaging surveys through spectroscopic follow-up, with the design of the Fiber-Optic Broadband Optical Spectrograph (FOBOS) and the development of machine-learning methodologies that will use FOBOS to train imaging data.

1. Intellectual Merit

1.1. **Scientific Justification.** Led by NSF's Large Synoptic Survey Telescope[1] (LSST), astronomy is entering a new era of unprecedented deep-imaging data sets that will survey huge volumes of the Universe. From the emergence of the earliest galaxies from a "primordial soup" of gas and dust, to the peak of cosmic star formation and the current era of accelerated expansion, these surveys will provide unprecedented statistics at key epochs of cosmic history.

Even so, gaining physical insight from panoramic imaging surveys will require intensive spectroscopic follow-up. The power of combining photometry and dedicated spectroscopy is widely appreciated and perhaps best illustrated by the success of the Sloan Digital Sky Survey (SDSS) which used this combination to record the properties of 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 follow-up of LSST galaxies** at a level required to capitalize on the $1B the U.S. has invested 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!

Solving this problem requires not only more powerful spectroscopic facilities, but new ways to take better advantage of what will be necessarily more limited spectroscopy compared to the vast imaging surveys of the LSST era. The more promising path 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 Archaeology in the LSST era; 2) Preliminary design of FOBOS,[2] a state-of-the-art spectroscopic facility for WMKO,[3] optimized for providing critical training data using one of the world's largest telescopes; 3) Preliminary design of the coordinated FOBOS observations required as well as the systems needed to serve training data publicly. This **MSRI-1 Design** proposal lays out the path for maximizing panoramic imaging from LSST, WFIRST,[4] Euclid,[5] and other facilities with spectroscopic follow-up unparalleled in depth and sampling density. 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 data.

1.2. **Research Community Priority.** The need for spectroscopic follow-up 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):

[1]LSST will be begin science operations in 2023.

[2]FOBOS: Fiber-Optic Broadband Optical Spectrograph

[3]WMKO: W. M. Keck Observatory operates the two twin 10m Keck Telescopes on Maunakea, Hawaii.

[4]WFIRST is NASA's space-based Wide-Field Infrared Survey Telescope, expected to launch in the mid 2020's.

[5]Euclid is led by the European Space Agency with significant NASA involvement and will launch in 2021. Its primary mission is a 15,000 deg^2 imaging survey in optical and near-IR wavebands.

> 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.

Workshops organized by the National Optical Astronomy Observatory (NOAO) in 2013 and 2016, the latter at the NSF's request, reported specific spectroscopic needs for LSST follow-up in all science areas. In particular, the 2016 report notes that a critical resource in need of prompt development is to "Develop or obtain access to a highly multiplexed, wide-field optical multi-object spectroscopic capability on an 8m-class telescope." Based on these recommendations, we propose the 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, FOBOS can access more than 70% of the LSST footprint, more than adequate for building powerful spectroscopic training sets. Compared to the Prime Focus Spectrograph (PFS) on Japan's Subaru Telescope, FOBOS would be 1.7× faster, provide unique UV sensitivity (0.31–1 μm compared to 0.38–1.25 μm with PFS), and offer higher-density, more flexible target sampling over "deep-drilling" fields. Unlike PFS, FOBOS would be operated on a U.S. telescope with dedicated U.S. access and a commitment to supporting U.S.-led imaging facilities. FOBOS is also complementary to future telescopes and instruments that would be optimized to cover wider areas (several deg^2 per pointing) at shallower depths.

The need for deep spectroscopic follow-up is particularly acute for the major cosmological probes to be carried out by LSST, Euclid, and WFIRST, which all rely on "photometric redshifts:" measures of galaxy redshift, z — a direct proxy of distance and look-back time—based on imaging alone. Newman et al. (2015) summarize the case for a significant spectroscopic campaign to calibrate and train LSST photometric redshifts in order to improve cosmological constraints. They describe a redshift survey that, if carried out with FOBOS, would increase LSST's Dark Energy figure-of-merit by 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 "Big Data" era, 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 supported the Informatics and Statistics Science Collaboration and partnered with NSF on 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 complemented by FOBOS spectroscopy, we will develop astrostatistical techniques and applications over the proposal period that will refine the FOBOS instrument requirements, inform the emerging design and operational modes, and define required training data. Tackling these challenges requires a community-wide effort and will deliver widespread benefits. Our goal 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.

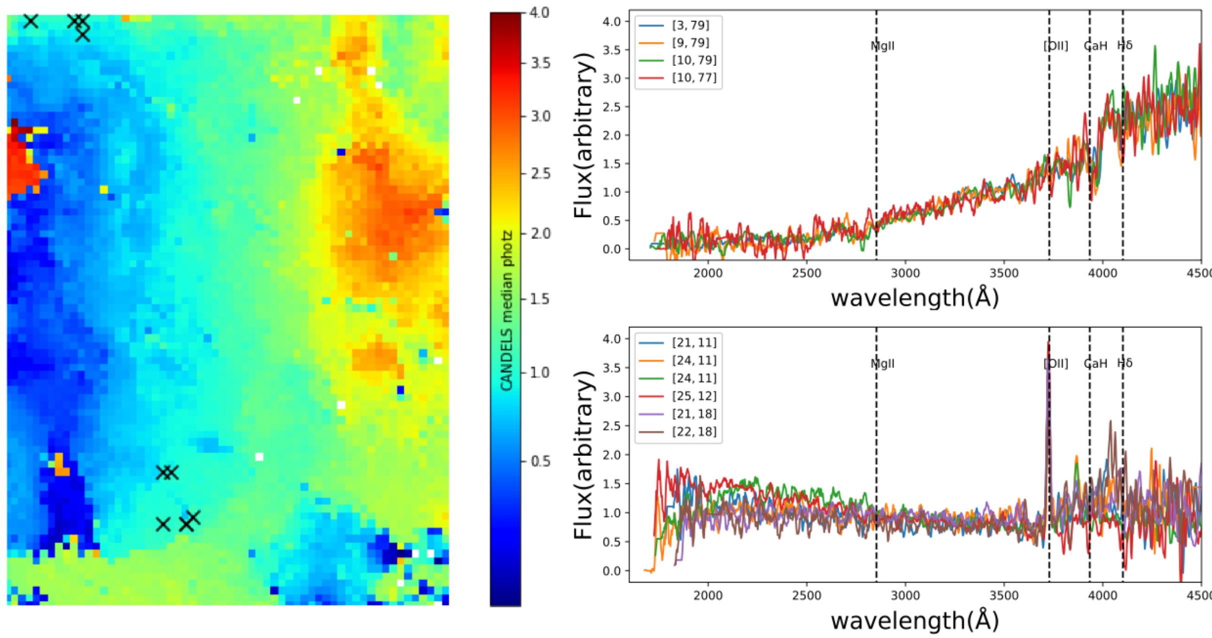

FIGURE 1. *Left*: A Self-Organizing Map (SOM; Kohonen, 1990) from Hemmati et al. (2018) encoding the relation between colors in an LSST+WFIRST-like color space and redshift, z. Position in the SOM is associated with a position in the multi-dimensional broad-band color space of galaxies. Galaxies observed in this space are assigned z values based on the median photo-z of galaxies from the CANDELS survey (color bar; Grogin et al., 2011). Such SOMs can be used to optimally define spectroscopic training samples for use with imaging surveys. *Right*: Galaxy spectra from VVDS (Le Fèvre et al., 2005); black crosses near the top and bottom of the SOM are plotted in the top and bottom panels, respectively. Note the similarity of the high-resolution spectra associated within the SOM, suggesting that a systematic spectroscopic exploration of the LSST color space would have far-reaching benefits to the science return of the mission beyond the photo-z application.

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 is accelerating due to the mysterious "dark energy," the origin of which remains unknown. Dark energy is one of the most fundamental, unsolved problems in both cosmology and particle physics. It has inspired enormous world-wide efforts — culminating in LSST, Euclid, and WFIRST — that seek highly precise measures of cosmic structure to constrain the evolving dark-energy equation-of-state.

These measures utilize angular correlations of galaxy positions, their gravitational lensing shear, and the cross-correlation between the two. Unfortunately, photometric distances (via photometric redshifts, or "photo-zs") are significantly less precise than spectroscopic redshifts (spec-zs), introducing significant biases. The spectroscopic validation of photo-zs we propose with FOBOS is therefore critical to the success of *all* imaging surveys in this respect. It would not only *increase the dark energy figure-of-merit in LSST by 40%* (Newman et al., 2015) but, importantly, provide vital confidence in cosmological results. FOBOS is particularly powerful in this application because it has no "redshift desert" thanks to its unique ability to measure spectroscopic redshifts above $z > 1.5$ via rest-frame UV features. This eliminates the need for expensive, space-based[6] near-IR spectroscopy.

Data-Science Challenge 1: Enable high-precision LSST photometric redshifts with strategically designed training spectroscopy: FOBOS is ideally suited to obtaining large ($>$ $10,000$ galaxies) and deep spectroscopic training sets required for LSST, WFIRST, and Euclid (see Newman et al., 2015). Applying machine-learning, template-based, and hybrid photo-z estimators to simulated data, we will build a methodology for designing FOBOS training sets that minimize observing time while maintaining adequate parameter-space sampling.[7] Self-Organizing Maps (SOM, Fig. 1) provide a state-of-the-art representation of a high-dimensional input space in projected 2D grid cells, allowing us to benchmark sampling of the photometric color space under various training set designs. We will use Bayesian Optimization techniques to evaluate the success of simulated training sets against the fidelity of full cosmological analyses that employ them. This will enable extremely rapid exploration of the optimal design space.

1.3.2. *A Comprehensive Picture of the Proto-galaxy Ecosystem.* Roughly three billion years after the Big Bang ($z{\sim}2$), the universe began a key epoch in which proto-galaxies transitioned from turbulent, gas-rich systems into the more ordered, star-dominated structures that populate the universe today. This period marks the peak of cosmic star formation and galaxy assembly. To understand it, we must study the entire galaxy "ecosystem," including not only the galaxies themselves but 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.

Although our first challenge emphasized photo-z training, here we will apply machine learning more broadly to infer physical parameters from multi-wavelength photometry as well as lower quality spectra trained with high-S/N data sets (cf., Fig. 2). Given the cost of deep spectroscopy, a central goal is to extract the maximum information from photometry and shallower spectroscopy so that more statistically powerful samples over larger cosmic volumes can be studied.

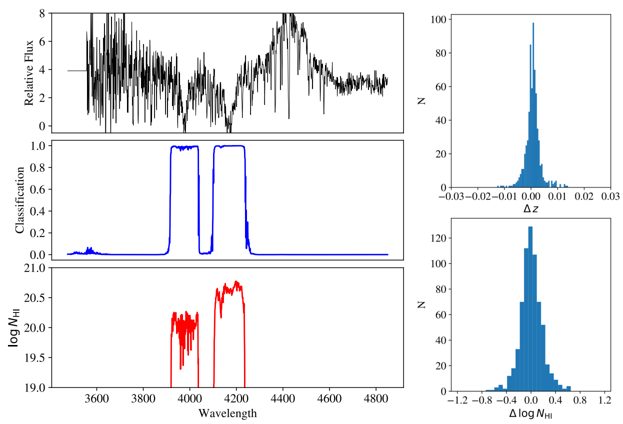

FIGURE 2. Application of machine learning to find and quantify the physical parameters of absorption by neutral hydrogen gas in spectra taken along quasar sight lines (adapted from Figs 7 and 14 from Parks et al., 2018). Two absorption systems in the spectrum (top-left) are identified (middle-left) and then "labeled" with an HI column density ($N_{\rm HI}$) (bottom-left) using a convolutional neural network (CNN). Redshift, z, (top-right) and $N_{\rm HI}$ (bottom-right) measurements obtained the CNN are in excellent agreement with derivations by experts. FOBOS will provide rich data sets for similar transfer of physical parameter labels to photometric and spectroscopic data.

Data-Science Challenge 2: Apply deep-learning algorithms to infer physical properties of galaxies at $z{\sim}2$ using using photometry. The range of observed spectral types is well-constrained by broad-band imaging (Figure 1),

[6]Ground-based near-IR spectroscopy is too contaminated by sky-line emission to provide spec-zs at the required level of completeness (Newman et al., 2015).

[7]Of course, the design of this sample also benefits the data-science challenges we detail in Section 1.3.2.

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. The challenge here is to identify the extent to which machine learning can deliver SDSS-like information — e.g., star-formation histories, stellar-population properties, dust content, inflow/outflow properties, and stellar masses — and determine design parameters for future training sets that will enable such inferences for millions of imaged galaxies at $z{\sim}2$.

Data-Science Challenge 3: Train short spectroscopic exposures in combination with 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. However, with improved photo-zs available from Challenge 1 and strong priors on spectral types (Challenge 2), the challenge here is to push machine-learning techniques to deliver *spectroscopic* redshifts (with 300 km s^{-1} accuracy) at the lowest signal-to-noise possible. Reductions by factors of 4–5 in exposure time would enable FOBOS to complete a 1M galaxy environment survey at z = 1–2 in just 20-30 nights.

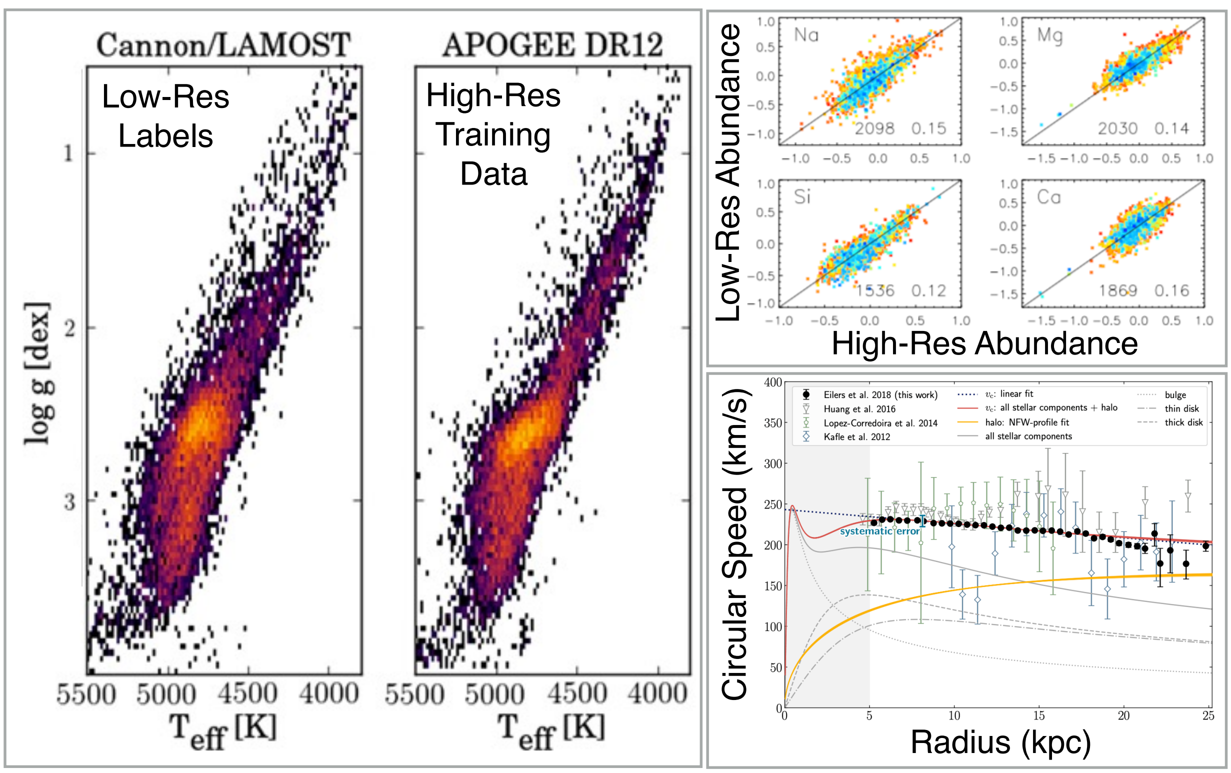

FIGURE 3. *Left*: Validation of *The Cannon* measurements of stellar effective temperature, $T_{\rm eff}$, and surface gravity, $\log g$, using low-resolution LAMOST spectra (left) compared to high-resolution APOGEE measurements (right; Ho et al., 2017). *Top-right*: Recovery of elemental abundances from low-resolution LAMOST spectra compared to high-resolution measurements from GALAH (Xiang et al., in prep). *Bottom-right*: The circular-speed curve of the Milky Way determined using a data-driven model that combines stellar parameters determined from APOGEE spectra with photometry from WISE, 2MASS, and Gaia, yielding the most precise measurements to date (Eilers et al., 2019).

1.3.3. *Unraveling the Formation History of our Local Group of Galaxies.* Our Local Group of galaxies—the Milky Way (MW), the Magellanic Clouds, Andromeda (M31) and Triangulum (M33) Galaxies, and a multitude of satellite galaxies—allows us to study one realization of the galaxy formation process in superb detail. In the next decade, LSST and WFIRST will increase the census of stellar streams and halo substructure in these galaxies by a hundredfold. Follow-up *stellar* spectroscopy will constrain stream orbits and the total mass they enclose (Sanderson et al., 2017) as well as the associated age and chemical composition (see below). As theoretical modeling also advances, these new data promise exciting insights on the formation history of the Local Group.

Radial velocity studies of stars in the MW halo or the M31 disk require observations of up to 10 hours on large telescopes (e.g., Cunningham et al., 2018). This again motivates machine-learning algorithms to extract physical quantities from both multi-band imaging and lower quality spectra (low resolution and S/N) using relatively small, yet high-S/N, training sets. For example, Ness et al. (2015) have developed *The Cannon*, a supervised learning approach that uses spectra with known stellar parameters to label spectra where those parameters are unknown (Fig. 3). Additionally, Ting et al. (2018) have developed *The Payne* which can infer 16 stellar-abundance labels from low-resolution spectra using a neural network and theoretical stellar spectra. Finally, Ting & Rix (2018) have combined Kepler-based astroseismology measurements with APOGEE spectra to determine stellar age to ∼25% precision using a neural network. Our proposed effort builds on new lines of inquiry based on these successes.

Data-Science Challenge 4: A nested network of stellar parameter training samples for resolved Milky Way and Local Group studies. As in our other challenges, the key driver here is the ability to extract maximum information from photometry, in this case stellar parameters. Our goal is to reach magnitudes significantly fainter than the detection limit of current and upcoming spectroscopic surveys of the Milky Way including Gaia, APOGEE,[8] the SDSS-V Milky Way Mapper, planned programs with 4MOST[9] and the Dark Energy Spectroscopic Instrument (DESI) Milky Way Survey, among others. Inferring stellar parameters beyond V∼18 will open up studies of the Milk Way's outer halo, the halo of M31, and stellar populations in local dwarf galaxies.

The immediate challenge is to design an optimized, nested set of training samples that connect data from the surveys above. This nested set will span high-S/N to low-S/N and high spectral resolution to low spectral resolution for sufficiently large, overlapping stellar samples. Subsets will have astroseismology from TESS[10] and PLATO.[11] Using simulated spectra with known input parameters, we will test methods for "label transfer" from information-rich spectra to information-poor spectra as we work down to fainter magnitudes, landing eventually at multi-band photometry alone. Within this nested set, low-resolution FOBOS data will fill in gaps at both high-S/N, where we will be training FOBOS data on higher resolution spectroscopy, as well as lower-S/N where we will be training photometry on FOBOS spectroscopy. The success of this multi-layered label transfer depends not only on the size of the training sets we can access or observe, but on how representative they are. Label transfer to WFIRST imaging of the M31 halo, or Local Group dwarfs in either hemisphere, is a particular concern. We will test it by evaluating label recovery on simulated stellar spectra with cosmologically-informed

[8]APOGEE, the Apache Point Observatory Galaxy Evolution Experiment has observed in both SDSS-III and SDSS-IV.

[9]4MOST: 4-meter Multi-object Spectroscopic Telescope.

[10]TESS is NASA's Transiting Exoplanet Survey Satellite.

[11]PLATO is ESA's PLAnetary Transits and Oscillations mission.

formation histories for M31 and dwarf galaxies, suitably differentiated from the Milky Way stars that anchor the training network.

2. Project Implementation

To meet the substantial spectroscopic needs of the U.S. astronomy community (Section 1.2), we propose three coordinated activities: 1) Organizing and evaluating the results of a community-wide effort to address a series of data-science challenges; 2) Completing preliminary design for the FOBOS instrumentation, informed in part by refining requirements as a result of (1); 3) Designing the operational modes, planning tools, data analysis software, and platforms necessary to deliver public training data that address these challenges. We focus our current request on the Preliminary Design Phase (PDP) of FOBOS development, following the the overall project plan and final deliverables outlined below. Anticipating progress in all three activities, we will request NSF MSRI-2 funding in 2021 to build and deploy FOBOS at WMKO, with observations and publicly served training data to follow. FOBOS would see first light in 2027 and carry a total cost of \$32M (without contingency in 2019 dollars).

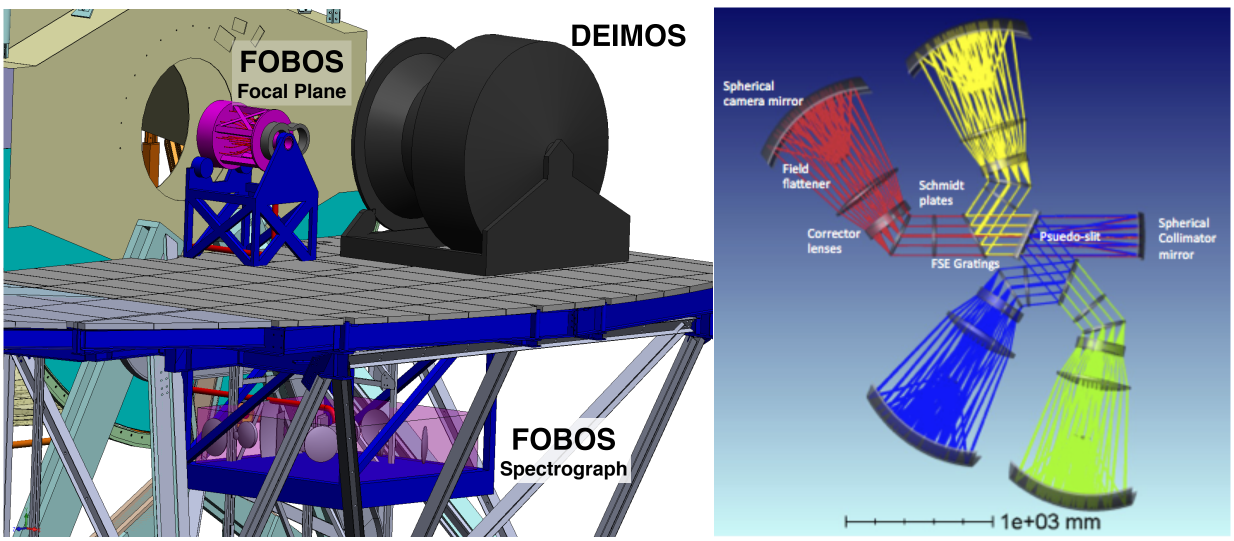

Figure 4. *Left*: Rendering of FOBOS instrument systems deployed at the Keck II Nasmyth port. By mounting the FOBOS spectrographs under the Nasmyth platform, other instruments like DEIMOS can maintain access to the telescope. *Right*: Rendering of one of the three four-armed FOBOS spectrographs.

2.1. **FOBOS Instrument Concept.** Mounted at the Nasmyth focus of Keck II Telescope at WMKO, FOBOS (Fig 4) will be one of the most powerful spectroscopic facilities deployed in the next decade. FOBOS includes a compensating lateral atmospheric dispersion corrector (CLADC, not pictured) to ensure 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 ($\sim$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. Fore-optics on the front end of each fiber demagnify and speed up the beam (from f/15 to f/5) for better coupling to the fiber numerical aperture and to minimize losses from focal ratio degradation. The focal plane plate rotates and translates to follow image positions as the telescope tracks across the sky. The fiber run is kept at less than 10m to maintain high throughput at UV wavelengths. 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 (Fig 4). Each spectrograph uses a series of dichroics to divide the 259 mm diameter collimated beam into four wavelength channels with combined, instantaneous coverage from 0.31–1 μm. Fused-silica etched (FSE) gratings provide mid-channel spectral resolutions of $R \sim 3500$ at high diffraction efficiency in each channel. The dispersed light is focused by an f/1.1 catadioptric camera[12] and recorded by an on-axis 4k×4k 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 peaks at 60% and 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. **FOBOS Instrument Design Effort.** FOBOS will complete its current conceptual design phase in fall 2019. Funding from this proposal will support the next phase of Preliminary Design. A schedule of milestones and additional information is 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-control design of the ADC will be completed.

Focal Plane System: 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 II Telescope and must comply with WMKO space envelopes, servicing needs, and other requirements. The focal plane system also includes the guide cameras.

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 use of Starbugs by 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.

Fiber System: We will complete the optical design and processing plan for affixing forward optics lenses to each fiber head. A micro-lens array solution will be developed for a central, fixed-position 4.5-arcsec diameter IFU for fast source acquisition. This work package 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.

[12]Based on the camera design for the Multi-Object Optical and Near-infrared Spectrograph (MOONS) on the Very Large Telescope (VLT).

2.3. **Addressing Data Science Challenges and Designing FOBOS Training Sets.** Our team includes leading experts on data science applications to astronomy and, specifically, LSST. 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 community workshops to motivate progress and discuss results. At the end of the proposal period, we will publish the results and developed software packages.

Our data-science challenges require work on simulated imaging+spectroscopic data sets where input physical properties (e.g., redshift) can be compared to output recovered values. Simulated imaging data (e.g., from LSST and WFIRST) are in-hand, while mock spectroscopy will be provided by a 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 desired FOBOS training sets, 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, quick-look 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, public release of FOBOS training data (e.g., Newman et al., 2013).

2.4. **MAISTRO: Target Allocation with Artificial Intelligence.** Powered by Starbugs fiber positioners, 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 MAISTRO,[13] 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 quick-look reduction package, science-driven performance metrics, *and real-time assessments of the 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 will incorporate updated target lists and priorities from the active observer and could easily be over-ridden at any time. Fractions of the full FOBOS multiplex might also be reserved "manual targeting" as required by the program PI.

2.5. **Publicly Available Automated Data Products.** While the FOBOS data simulator is required for our data-science challenges, it also forms the basis of a delivered data reduction pipeline (DRP) for this instrument. This software will provide both the quick reduction assessments needed for dynamic targeting, as well as full reductions for scientific analysis. In the proposal period, we will also develop a preliminary design for a data analysis pipeline (DAP).

[13]MAISTRO: Modular Artificial Intelligence System for Target Reallocation and Observing.

Unique among Keck instruments, the FOBOS DAP will take advantage of the fixed spectral format and common target classes to provide high-level data products, including Doppler shift, emission-line strengths, and template continuum fits (cf., Westfall et al.; SDSS-IV MaNGA DAP). The DAP will also produce results from relevant machine-learning applications (e.g., redshifts at low-S/N).

Raw data, reduced spectra, and high-level DAP science products will be publicly delivered via user-friendly platforms built on the Keck Observatory Archive. After associated proprietary periods, data will be served for *all* FOBOS observations, creating a rich legacy data set for the astronomical community. Both program PIs and the larger community will be encouraged to develop the DRP and DAP 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

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 Internship Program is aimed at advancing college students from Hawai'i into the STEM workforce. Almost 400 students have participated to date, of which 24% are Native Hawaiian and 38% are women. A longitudinal study of Akamai outcomes indicated that 87% were still in STEM, either in the workforce or continuing STEM studies (Barnes et al., 2018). ISEE and the Akamai program already have deep connections to WMKO; 45 interns have worked on projects related to instrument development and observatory operations over the past 15 years. Our funding request includes support for two Akamai interns.[14] One intern will develop aspects of the FOBOS instrument simulator and use this simulator to develop performance metrics for the Preliminary Design. The second will build machine-learning tools for Data-Science Challenge 3, specifically for obtaining spectroscopic redshifts at low S/N.

3.2. **Investing in future educators.** Also via the ISEE, we will support three graduate students to participate in the Professional Development Program (PDP) that build teaching skills through collaborative design of an inquiry activity. The PDP team conceives, develops, and tests the activity which culminates in a lab exercise run with undergraduates. The program emphasizes inclusive and equitable learning environments. Specifically, our team of graduate students will develop a lab unit related to FOBOS instrument development aimed at incoming community college transfer students enrolled in UCSC's highly-successful Lamat Program. In addition to enriching graduate student training, these efforts will positively impact 25 undergraduates from California community colleges, a large fraction from underrepresented minority groups.

3.3. **Student Training.** UCSC's **Astro 9** course introduces scientific research methods to 1st-year students by engaging small student teams on actual research projects supervised by graduate students, postdocs, and staff. The **Science Internship Program** (SIP) creates a similar environment for high-school students over a 10 week summer program. We will design projects for both programs focused on simulating data sets and introducing machine learning concepts used in our Data Science Challenges. Both PI Bundy and co-PI Westfall have served as research mentors in these programs.

[14]At UCO, an Akamai intern during Summer 2018 helped build a fiber test-bench at UCSC.

References

Barnes, A., Ball, T., Starr, C. R., et al. 2018, in 2018 ASEE Annual Conference & Exposition (Salt Lake City, Utah: ASEE Conferences), https://peer.asee.org/31030

Borne, K., Accomazzi, A., Bloom, J., et al. 2009, in astro2010: The Astronomy and Astrophysics Decadal Survey, Vol. 2010, P6

Council, N. R. 2015, Optimizing the U.S. Ground-Based Optical and Infrared Astronomy System (Washington, DC: The National Academies Press), doi:10.17226/21722

Cunningham, E. C., Deason, A. J., Rockosi, C. M., et al. 2018, arXiv e-prints, arXiv:1809.04082

Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, The Astrophysical Journal Supplement Series, 197, 35

Hemmati, S., Capak, P., Masters, D., et al. 2018, arXiv e-prints, arXiv:1808.10458

Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5

Kohonen, T. 1990, Nature, 346, 24

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193

Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, A&A, 439, 845

Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53

Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5

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

Parks, D., Prochaska, J. X., Dong, S., & Cai, Z. 2018, MNRAS, 476, 1151

Sanderson, R. E., Hartke, J., & Helmi, A. 2017, ApJ, 836, 234

Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2018, arXiv e-prints, arXiv:1804.01530

Ting, Y.-S., & Rix, H.-W. 2018, arXiv e-prints, arXiv:1808.03278