

**Mid-scale RI-1 (M1:DP) Keck-FOBOS:
Building 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. Newman, X. Prochaska, C. Rockosi, J. Rhodes, M.
Rich, M. Savage, A. Shapley, B. Siana, Y.-S. Ting

[[10-page limit, excluding references. This Mid-scale Research Infrastructure-1 “Design” proposal requests funds to complete the preliminary instrument design for the Fiber-Optic Broadband Optical Spectrograph (FOBOS) and build frameworks for enabling data-driven science goals via a FOBOS Public Survey.]]

1. INTELLECTUAL MERIT

1.1. **Scientific Justification.** ”including the unique research capabilities and lack of general availability of the requested infrastructure and its potential to significantly advance the Nations research infrastructure.”

[[3/4 page]]

Led by NSF’s Large Synoptic Survey Telescope (LSST) which deploys in 2023, 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. 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 be detected 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 “imaging” followed by dedicated spectroscopy yielded an unprecedented sample of over 1 million galaxies, mapping the present-day universe and making SDSS the most highly cited survey 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 obtaining the spectroscopic followup** at a level required to capitalize on our \$1B investment in LSST. In fact, SDSS-like spectroscopic followup of 1 million galaxies at LSST distances 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,” namely *Harnessing the Data Revolution* in order to maximize the information content of LSST via machine learning of optimally-designed spectroscopic training sets. This proposal presents a connected approach to the three critical components in this endeavor: 1) Development of statistical approaches to specific and ambitious data-science challenges that will address key questions about transitional epochs in cosmic history; 2) Design of FOBOS, a state-of-the-art spectroscopic facility on one of the world’s largest telescopes optimized for obtaining spectroscopic training sets; 3) Design and execution of a “Public Survey” with this facility as well as a data-serving platform to provide spectroscopy and training data to the U.S. community. This MSRI-1 design proposal lays out the path for maximizing the panoramic imaging of LSST with spectroscopic followup. Through a subsequent MSRI proposal we will deliver on our goals with an instrument deployment in 2026 and completed spectroscopic followup as LSST concludes in 2029.

We address data science challenges in four research areas in order to guide our instrument and survey design:

- (1) Significantly more powerful probes of Dark Energy and Cosmology
- (2) Comprehensive understanding of the proto-galaxy ecosystem at $z \sim 2$
- (3) Archaeological studies of our own Milky Way galaxy and its satellites to unravel their specific journeys through this transition.
- (4) Time domain spectroscopy...?

1.2. Research Community Priority. [[3/4 page]]

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 made the following recommendation:

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). Based on these recommendations, we propose the Keck-FOBOS instrument coupled with a suite of data-driven tools to address LSST’s spectroscopic requirements at a final cost 20 times less than a new Southern Hemisphere facility. Located in Hawaii, Keck-FOBOS would have access to greater than 70% of the LSST footprint, more than adequate for our primary goal of building powerful spectroscopic training sets. FOBOS would complement future ambitious facilities that could 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.” Newman et al. (2015) summarize the case 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 50% 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. Wide-spread community interest in advanced data science techniques continues to grow amidst calls for educational structures, 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. Data-Driven Science Challenges. We introduce data science challenges in three science areas that will be addressed in this proposal. Each enables significant advances in our

understanding of transitional epochs in cosmic history by leveraging FOBOS-derived spectroscopic training sets to maximize the information that can be extracted from LSST.

1.3.1. *Enhancing Dark Energy Probes via Precision Cosmic Distances.* [[1 page]]

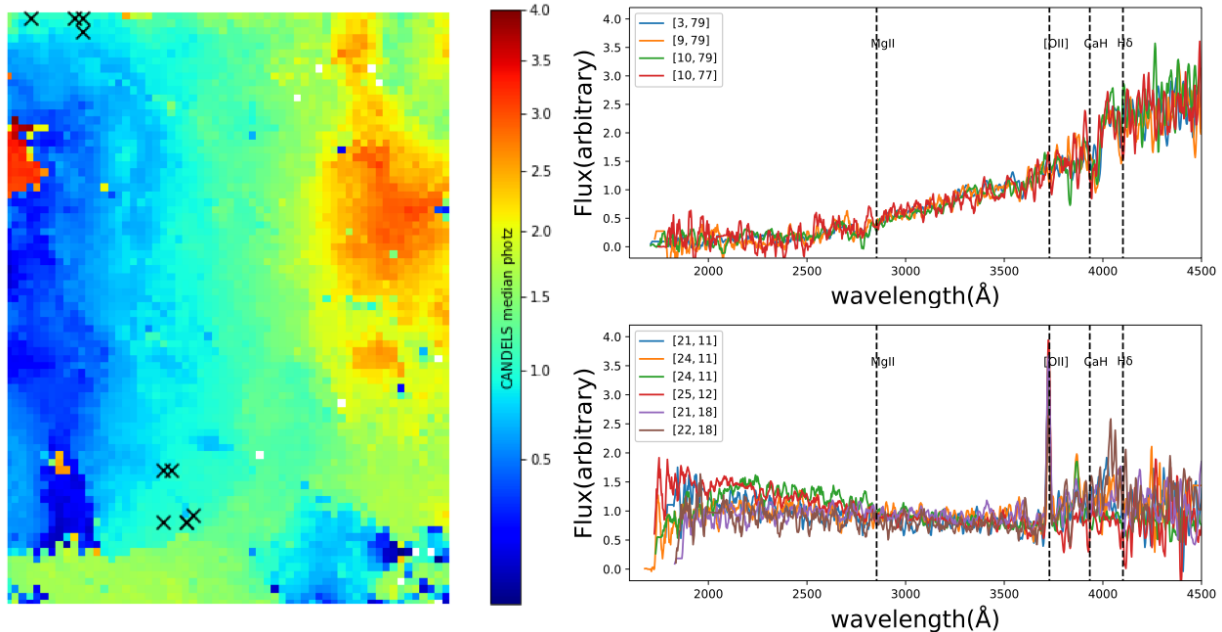


FIGURE 1. *Left:* A Self-Organizing Map (SOM) from Hemmati et al. (2018) visualizing the relationship between broadband galaxy brightness in different broadband filters (projected along the x and y axes) and observed spectroscopic redshift (indicated by the color map). SOMs guide the optimal construction of training samples by highlighting which galaxy classes require targeting. *Right:* Remarkably, the spectra associated with localized SOM regions are surprisingly similar, not just the associated redshifts.

The 2011 Nobel Prize in Physics was awarded for the discovery of an era of acceleration in the expansion rate of the universe that sets in when the universe was roughly half its current age. This accelerated expansion is often attributed to a mysterious “Dark Energy,” but the amplitude of this vacuum energy density is 120 orders of magnitude smaller than what naive models would predict.

Dark Energy is perhaps the single most important unsolved problem in cosmology and astrophysics. 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 expands as the universe expands, precise measures of cosmic structure allow us to reconstruct the cosmic expansion history, and at sufficient levels of detail, distinguish among various Dark Energy models.

Many of the most important cosmological probes require galaxy “redshifts,” a Doppler-like shift in the observed wavelengths of emitted light, as distant galaxies appear to recede from us as a result of cosmic expansion. Cosmological models, which provide access to Dark Energy parameters, are constrained in part via the conversion of redshift into distance. Accurate redshifts are conventionally derived via spectroscopy by fitting the observed wavelength locations of spectral features. Less precise redshifts, so-called “photometric redshifts” (or photo-*zs*) can be estimated with imaging photometry alone, a compelling option for large and faint data sets

where spectroscopic redshifts are infeasible. But, “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 50%* (Newman et al., 2015).

Data Science Challenge 1: Obtain precise LSST Photometric Redshifts ($\sigma_z/(1+z) \lesssim 0.005$ at $i(\text{AB}) < 25.3$) with Targeted Training Sample sizes of $<10\text{k}$. Our proposed FOBOS instrument is ideally suited to providing the spectroscopic training defined in Newman et al. (2015), but a complete program would require a 400-night investment in 10 m telescope time. This challenge demands a reduction in the required FOBOS training sample by a factor of ~ 4 via clever application of state-of-the-art machine learning techniques.

Neural network trained photo- z s have long been recognized for providing the best precision when sufficient training sets are available (e.g. Bundy et al., 2006), and significant effort is underway in optimizing their application to future cosmological imaging surveys. Hemmati et al. (2018) for example have exploited Self-Organizing Maps (SOMs, Fig 1) to sort multiband photometric data by observed redshift in order to select optimized training samples for spectroscopic followup.

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

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

1.3.4. *Time Domain.* **[[1/2 page]]**

2. PROJECT IMPLEMENTATION

2.1. FOBOS Instrument Concept. **[[1 page]]**

Mounted at the Keck II Telescope’s Nasmyth focus, 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 (Fig). A compensating lateral atmospheric dispersion corrector (CLADC) 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 that result from the telescope design. The final lens surface in the 3-lens CLADC also serves as the mounting plate for roaming Starbugs fiber positioners. Starbugs patrol a large on-sky area, enabling flexible targeting configurations that can be dynamically adjusted during observations.

A total of 1800 $150\ \mu\text{m}$ 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 $\sim 10\text{ m}$ 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 fiber output 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 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 to access the Nasmyth port. The end-to-end instrument throughput is greater than 30% at all wavelengths.

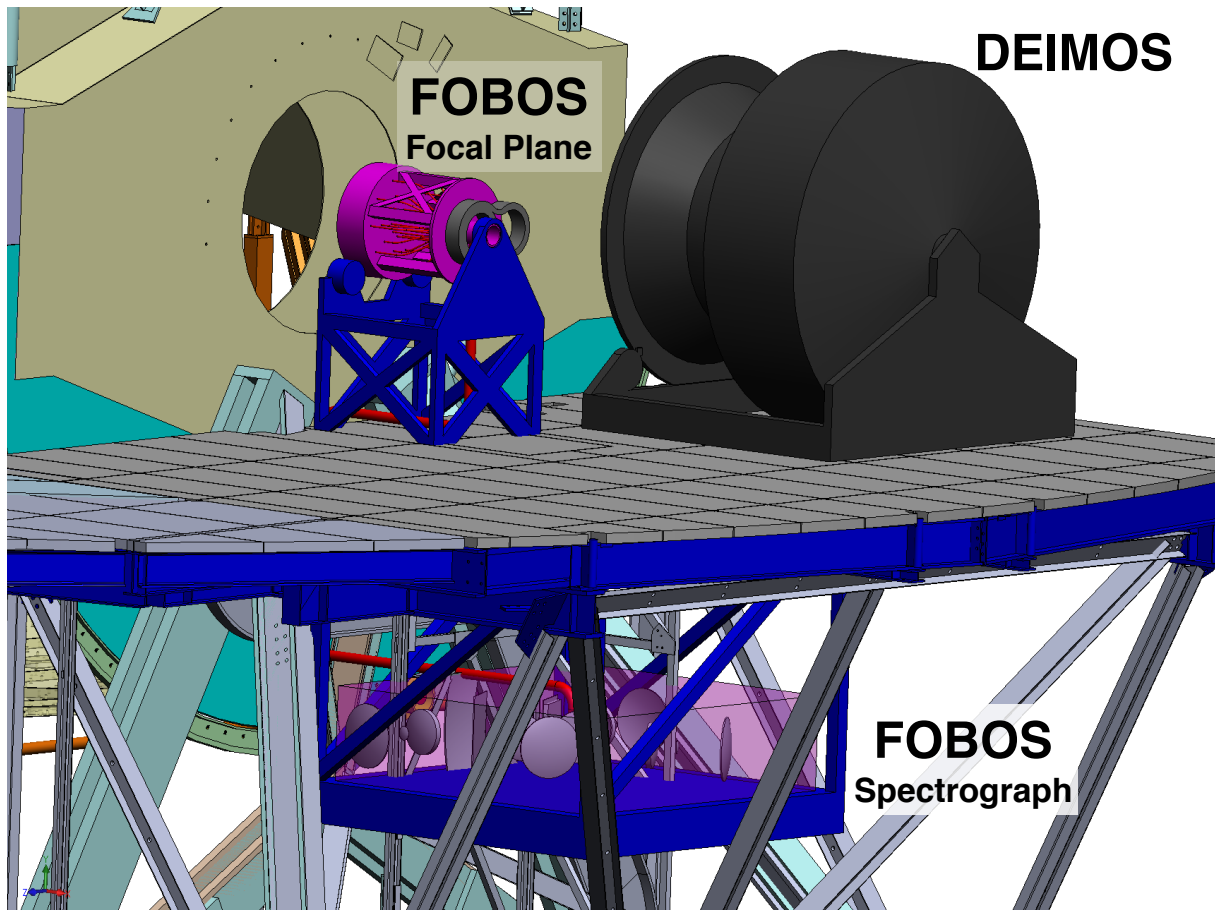


FIGURE 2. Rendering of FOBOS instrument systems deployed at the Keck II Nasmyth port. By mounting the FOBOS spectrographs under the Nasmyth platform, other instruments like DEIMOS can maintain access to the telescope.

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. Instrument Design Effort. [\[\[1 page\]\]](#)

2.3. Design of Public Survey and Training Sets. [\[\[1 page\]\]](#)

2.4. Target Allocation with Artificial Intelligence. [\[\[1/2 page\]\]](#)

2.5. Publicly Available Automated Data Products. [\[\[1/2 page\]\]](#)

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. **Student training.** **[[1/4 page]]**

3.2. **ISEE?.** **[[1/2 page]]**

3.3. **Data Science Major in Astrophysics.** **[[1/4 page]]**

”Preliminary proposals must include an outline of ongoing operations and maintenance plans, including an estimate of any needs for ongoing, NSF-supported operations and maintenance that may be requested outside of the Mid-scale RI program.”

”Results from Prior NSF Support should not be included. Also, links to URLs may not be used.”
no more than 2 pages for references

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>
- Bundy, K; Ellis, RS; Conselice, CJ; Taylor, JE; Cooper, MC; Willmer, CNA; Weiner, BJ; Coil, AL; Noeske, KG; Eisenhardt, PRM. “The Mass Assembly History of Field Galaxies: Detection of an Evolving Mass Limit for Star-Forming Galaxies,” *ApJ*, v. 651, 2006, p. 120–141. <http://adsabs.harvard.edu/abs/2006ApJ...651..120B>
- 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>

Budget and Budget Justification

”including budgets for any subawards. For preliminary proposals cost estimates may be preliminary estimates with the Basis of Estimates (BoE) included. Copies of vendor quotations should not be included in preliminary proposals. If the budget includes contingency, that contingency should cover the ”known unknowns” and be used to mitigate identified risks.”

Facilities, Equipment, and Other Resources:

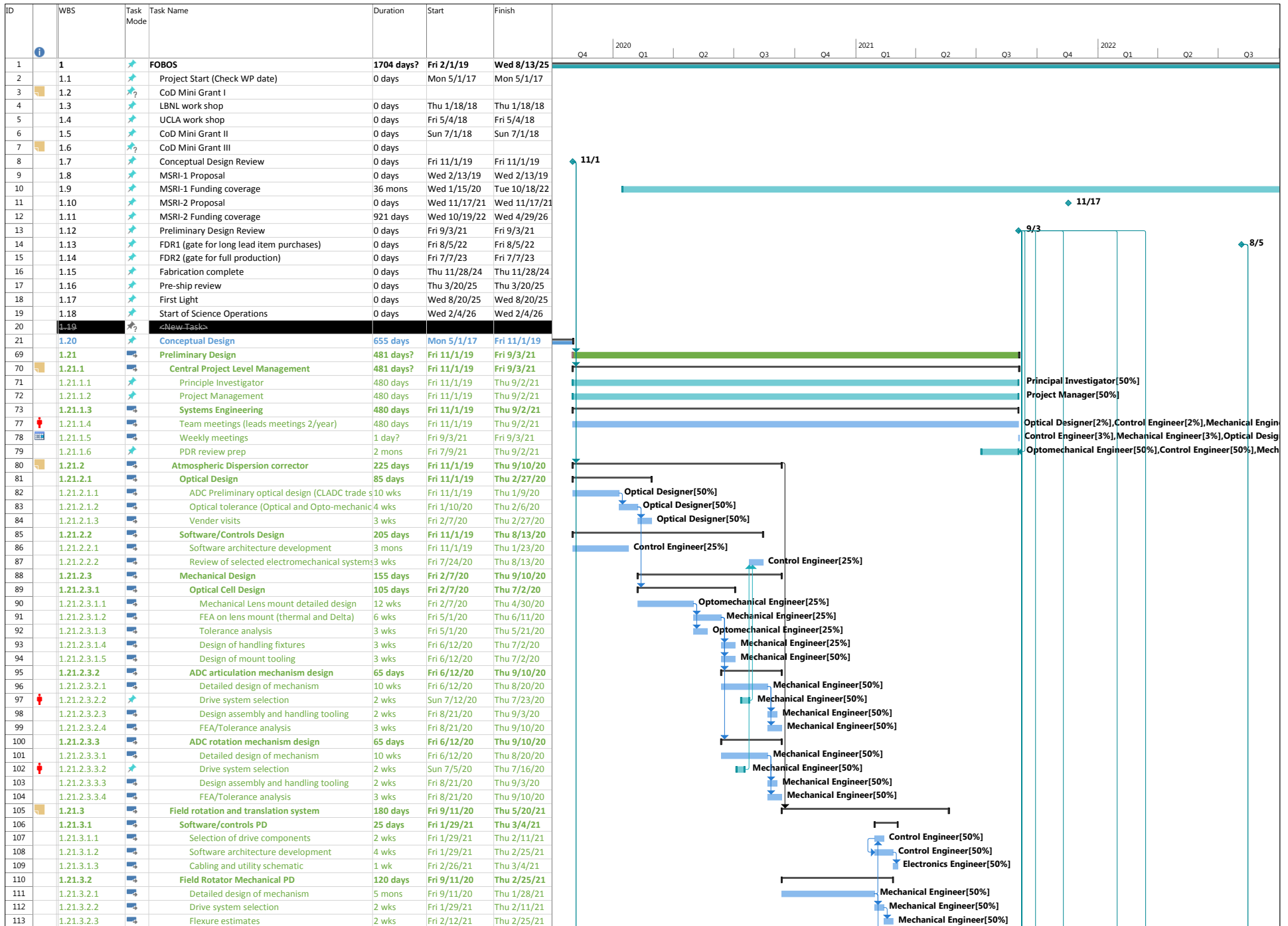
In order for NSF, and its reviewers, to assess the scope of a proposed project, all organizational resources necessary for, and available to a project, must be described in this section of the proposal. Proposers should describe only those resources that are directly applicable. The description should be narrative in nature and must not include any quantifiable financial information. Proposers should include a description of the internal and external resources (both physical and personnel) that are expected to be available to the project. Such information must be provided in this section, in lieu of other parts of the proposal (e.g., Budget Justification, Project Description).

Supplementary Documents:

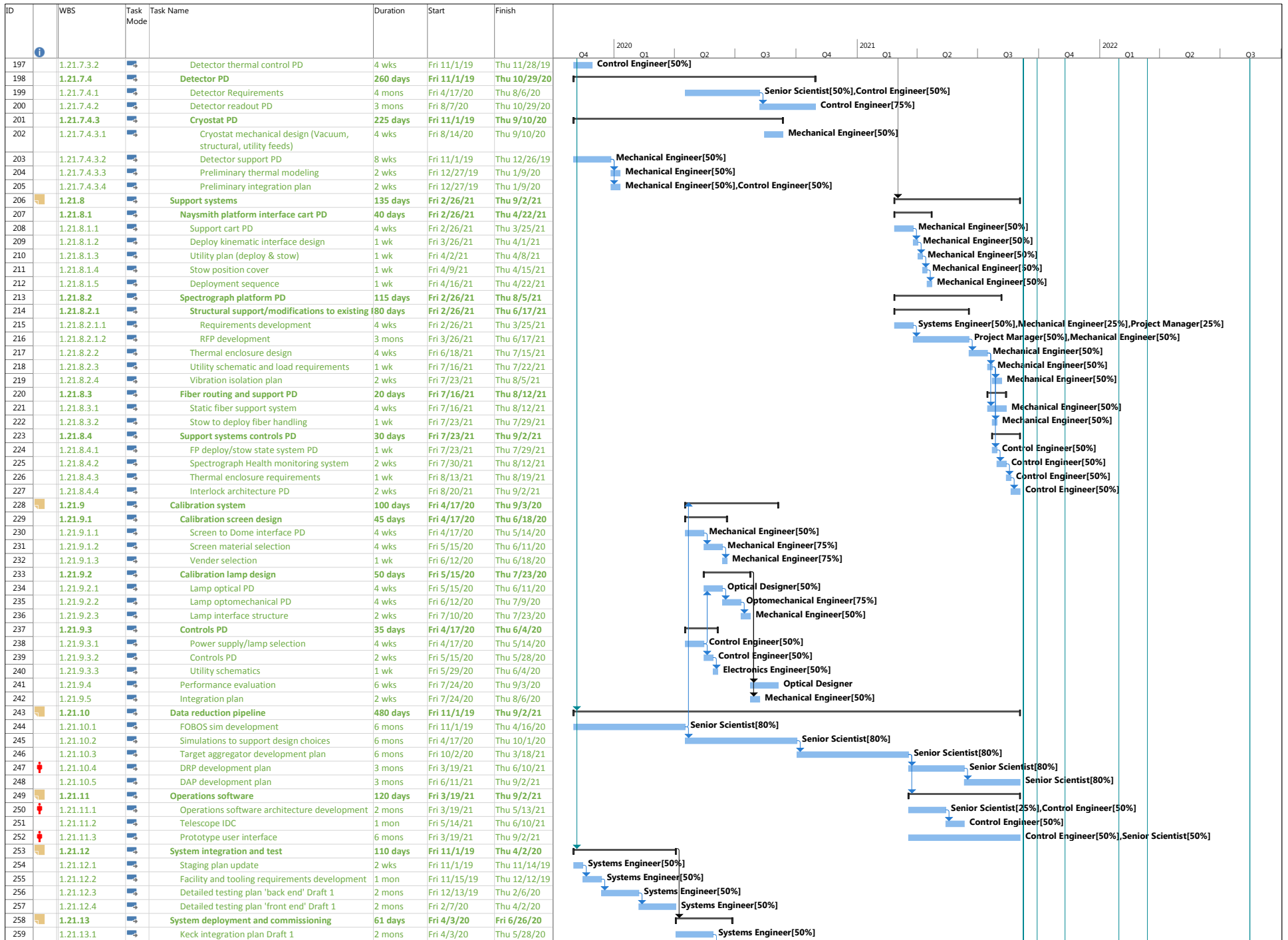
(to be entered in the Supplementary Documents section of FastLane)

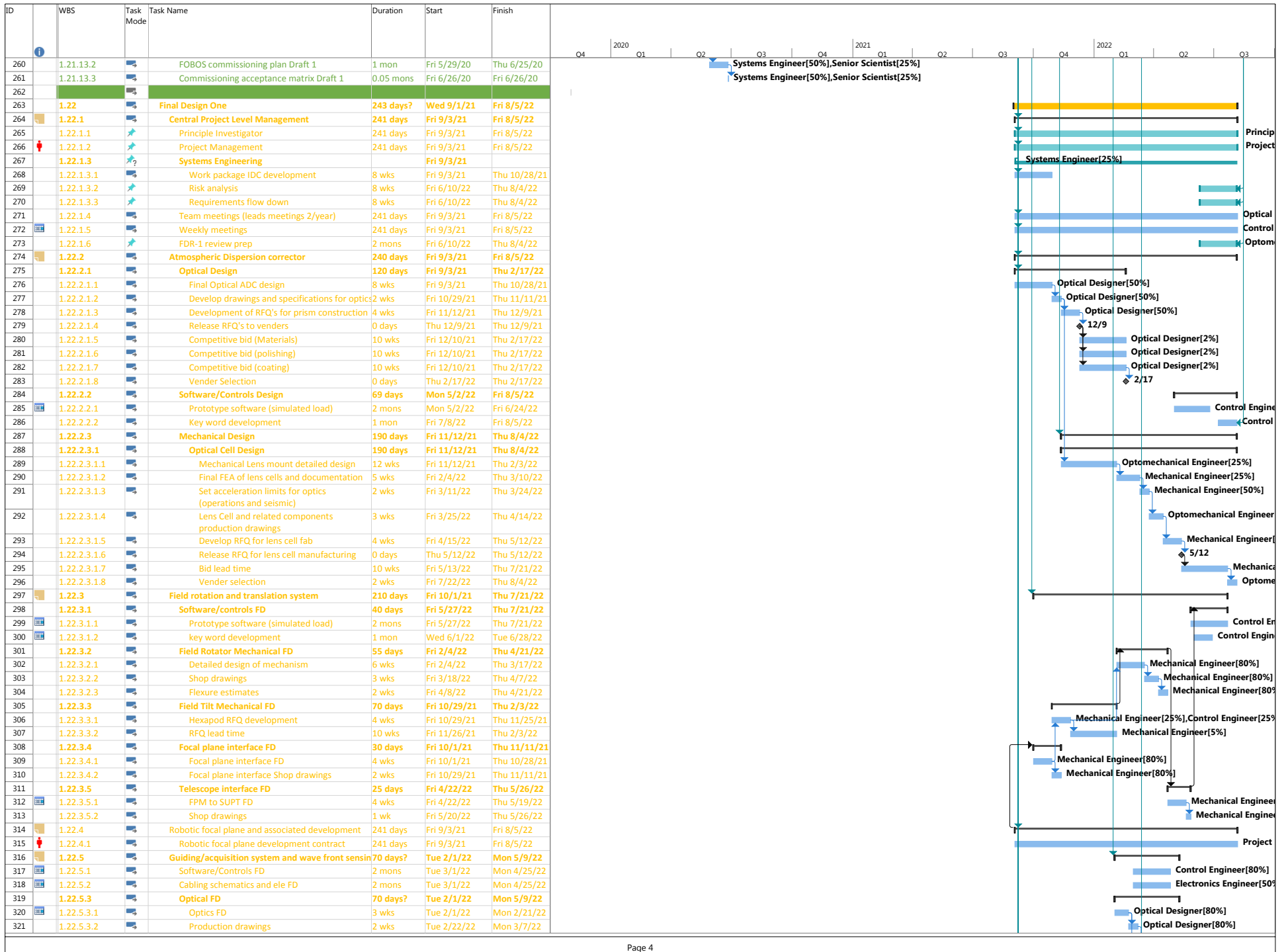
- (1) A list of the major team members, their affiliations, and their role in the project;
- (2) A list of Partner Organizations to be funded via subawards, and the role of each in the project;
- (3) An outline of the Project Execution Plan (PEP).

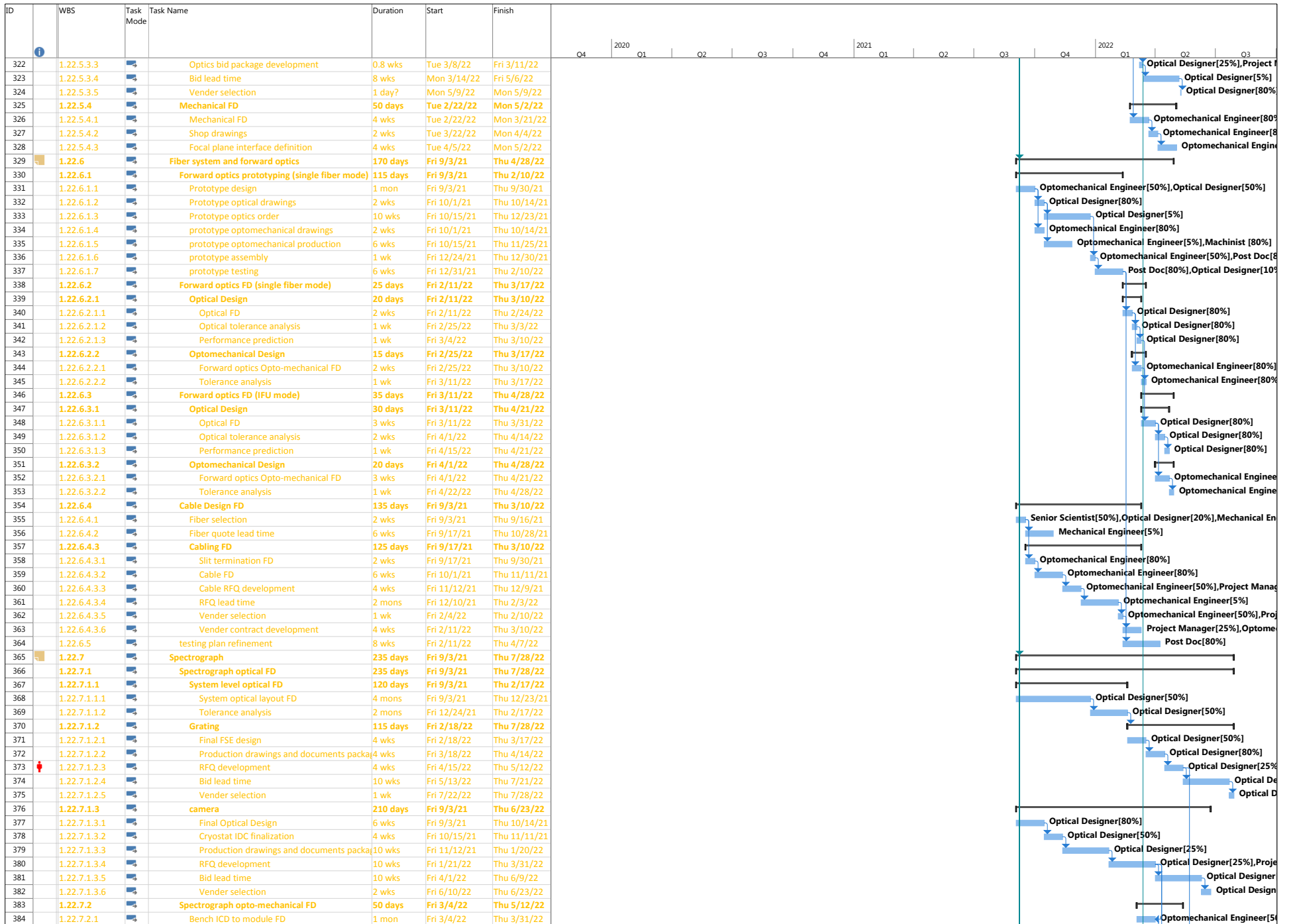
(See the LFM/MFG. Greater detail will be required in invited full proposals should that occur. See Full Proposal Preparation section for further information.)

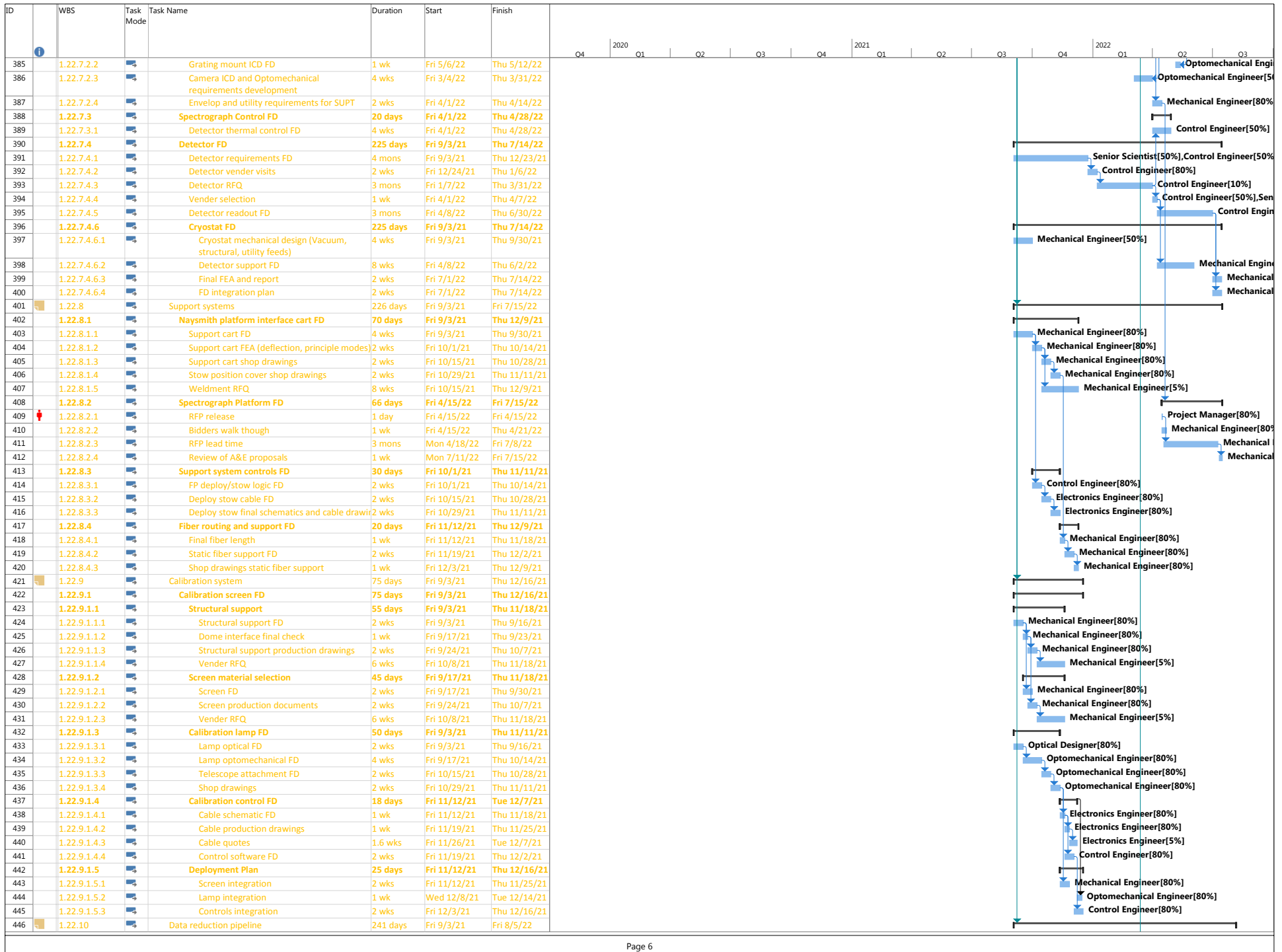


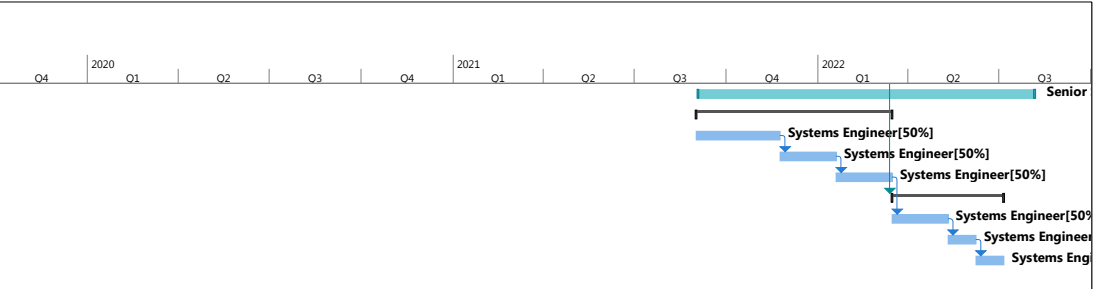










ID		WBS	Task Mode	Task Name	Duration	Start	Finish												
447		1.22.10.1		Data Lead (design choice support)	241 days	Fri 9/3/21	Fri 8/5/22												
448		1.22.11		System integration and test	140 days	Wed 9/1/21	Tue 3/15/22												
449		1.22.11.1		Facility upgrade requirements	3 mons	Wed 9/1/21	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																			