University of Arizona

30 August 2016

Final Report

1 May 2014 to 31 March 2016

University of Arizona

High Energy Physics Program

at the Cosmic Frontier

DOE/HEP Grant - Final Report **Submitted 30 August 2016**

Project Title:	University of Arizona High Energy Physics Program
DOE/HEP Contact:	Dr. Kathy Turner
DOE Award Number:	DE-SC0011625
Lead Principal Investigator:	Ken Johns, johns@physics.arizona.edu Physics Department 1118 E 4th St University of Arizona Tucson, AZ 85721 (520) 621-6791
Recipient Organization:	Arizona Board of Regents PO Box 3308 Vice President for Research Dr. Kimberly Espy Tucson, AZ 85722-3308 (520) 626-6000
Period of this grant:	1 May 2014 to 31 March 2016
Report covers the period:	1 May 2014 to 31 March 2016
Co-Principal Investigators:	Elliott Cheu, elliott@physics.arizona.edu

Table of Contents

Introduction

The High Energy Physics Group at the University of Arizona has conducted forefront research in elementary particle physics. This report covers our work on dark energy physics using the Large Synoptic Survey Telescope (LSST) in the Cosmic Frontier sector. During the grant period, May 1, 2014 to March 31, 2016, we provided engineering design support for power supplies for the LSST camera and developed tools for understanding systematic effects pertaining to dark energy measurements through weak lensing.

This Final Report marks the end of this grant.

Cosmic Frontier Report

The group of Elliott Cheu, Alexandra Abate and graduate student, Lidens Cheng, continued its work on the camera and dark energy science with the Large Synoptic Survey Telescope (LSST). We have been heavily involved with the development of the power supplies for the LSST camera. We also contributed significantly to the systematic studies necessary for achieving the requisite sensitivity in dark energy measurements through weak lensing.

Camera Power Supplies

Until 2015, the UA LSST group was responsible for the Power Distribution Cards (PDC) for the LSST camera and the associated Camera Control Software (CCS). Given the stringent noise, power and space requirements of the LSST camera, we helped design custom power supplies, as no commercial supplies would meet the demanding specifications. The requirements for isolation, monitoring and sequencing also eliminated the possibility of using a commercial solution.

In 2014 we completed the design and construction of a prototype card for the PDC system. This board was designed with a very stringent set of requirements. Unfortunately, the space constraints for the PDC crates were tightened significantly, leading to a redesign of the whole PDC system. In 2014 and 2015 we worked closely with SLAC on a new design for the PDC cards. However, due to resource issues at UA, we eventually transferred the project to SLAC.

We also were involved in the development of the command and control system (CCS) for the PDC system. In 2014 we had successfully demonstrated a prototype system using a BeagleBone Black small form factor computer. We utilized this system to continue development of the power control system to drive our prototype PDC cards. When we transferred leadership of the PDC system to our SLAC collaborators, we stopped further development of on the CCS system. As of 2016 we no longer have any involvement in the camera hardware or software.

Photo-z Calibration

One of the limiting systematics in weak lensing analyses will be the ability to accurately determine a galaxy's redshift. In LSST redshifts are determined using the photometric technique, utilizing a series of optical band pass filters. Current studies indicate that codes to reconstruct redshifts using this method are approaching their statistical limit. So, calibration of these photometric methods is necessary in order to achieve the requisite precision needed to determine precise dark energy constraints using LSST. As such, a sample of galaxies with known redshifts is required. We can perform this calibration directly, by obtaining a sub-sample of around 10^5 galaxies with accurate redshifts measured via spectroscopy. However, this sample must be statistically complete to >99.9% of the whole cosmological analysis sample [1] and existing deep surveys have only managed to secure accurate redshifts for up to 60% of targets [2]. The best alternative option is to utilize cross-calibration statistics [3], but to date the methods have not been proved to work for surveys as deep as LSST.

We undertook an in-depth investigation of how well LSST can measure dark energy in the event we cannot rely on cross-calibration methods to calibrate the redshifts of very faint galaxies. For various levels of statistical completeness we studied the size of the bias on the dark energy parameters, and from our simulation identified which sectors of the galaxy distribution contribute most to mis-characterization of photo-z errors, and hence bias on dark energy estimates. We expected the results of these studies to produce recommendations for targeting specific galaxies for spectroscopic follow-up.

In our study we generated a simulation of the LSST cosmological data set from a realization of another small simulation. We used this to estimate the *actual* precision of LSST photometric redshifts, along with the precision that would be constrained from a series of spectroscopic sub-samples, with different magnitude limits. The photo-z error distributions were parameterized as a mixture of three Gaussians, and an expectation-maximization algorithm was developed to perform the estimation. The actual versus constrained photo-z errors were implemented in a Fisher Matrix analysis of cosmic shear measurements to reveal the resulting bias on the dark energy. We also checked the size of calibration sample required to beat down sample variance coming from the variation of galaxy types selected in any random draw of a target calibration sample.

For LSST we found that for direct calibration methods around 10^6 galaxies are required to mitigate sample variance coming from galaxy type variation for samples with magnitude limits >=23.5. A spectroscopic calibration sample with a magnitude limit $m_i<24.5$ (nearly one magnitude below the LSST cosmological sample limit of $m_i<25.3$) could bias dark energy parameter estimates by about 150%, and with $m_i<23.5$ by at least 500%. A calibration sample with magnitude limit $m_i<25$ showed negligible bias on dark energy parameters. A positive finding was that nearly all of the bias on dark energy came from missing faint spiral galaxies in the redshift range $1<z<1.5$. (See *Figure 1*) If these galaxies can be especially targeted perhaps the requirements on statistical completeness can be relaxed and direct calibration of photo-z errors could be possible. However, we note that this conclusion depends strongly on the features of our simulation.

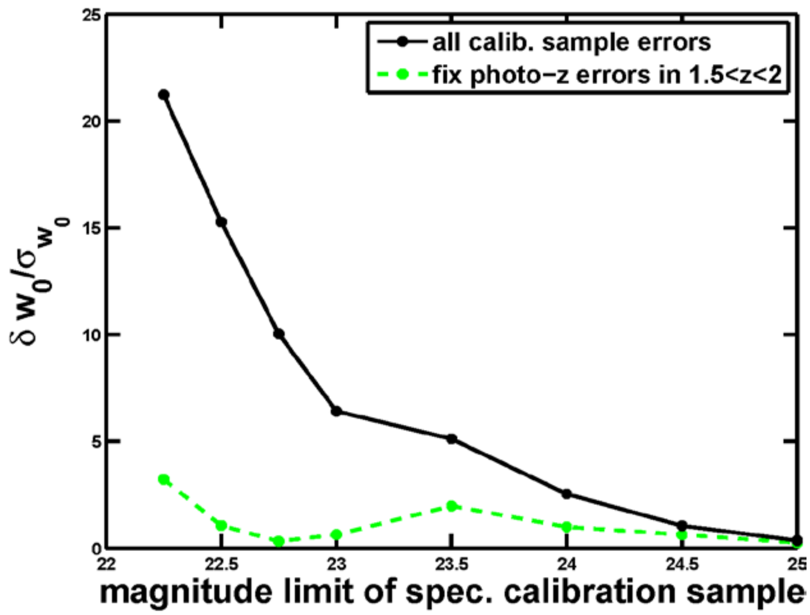

Figure 1 The resulting relative bias on the dark energy equation of state parameter w_0 as a function of the magnitude limit of the photo-z calibration sample (solid black line). The dashed green line shows the bias when the photo-z error distribution is fixed to the true one in just the redshift range 1.5<z<2.

From here we pivoted to work towards a more general conclusion and turned the question we were asking around to be: what is the optimum redshift selection for a magnitude limited calibration sample? Here optimum translates to minimizes bias on the dark energy parameters. We employed a Markov Chain Monte Carlo (MCMC) technique to select different proportions of calibration galaxies from redshift bins and find the selection that minimizes the total dark energy bias (sqrt[$dw_0^2 + dw_a^2$]). However the MCMC failed to converge properly and investigation into the different chain steps found that there was no discernable difference in the calibration sample redshift distribution between steps that had large dark energy bias and low dark energy bias. Therefore we decided to hold off on this investigation until more sophisticated simulations are produced as part of the Dark Energy Science Collaboration's Data Challenge 2 or 3 data sets that will be ready in 2017-2018.

<u>Study of the Effect of a Mixed CCD Focal Plane on LSST Photo-z</u>

As part of a study to evaluate the effect of a mixed focal plane made up of CCD chips manufactured by two different vendors, Abate ran a machine learning photo-z estimation algorithm on simulated data (created by S. Schmidt, UC Davis) based upon such a scenario. The results found that correcting all chips to a "standard" u-band did not degrade photo-z performance, when using either a template-based photo-z algorithm (as employed by S.

Schmidt) or a machine learning photo-z algorithm (as employed by Abate). A mixture of both sensors in the focal plane so each object has half of its observations occur within each sensor type, and where the data are not corrected to a “standard” u-band, proved to yield slightly higher precision than either alone (see *Figure 2*).

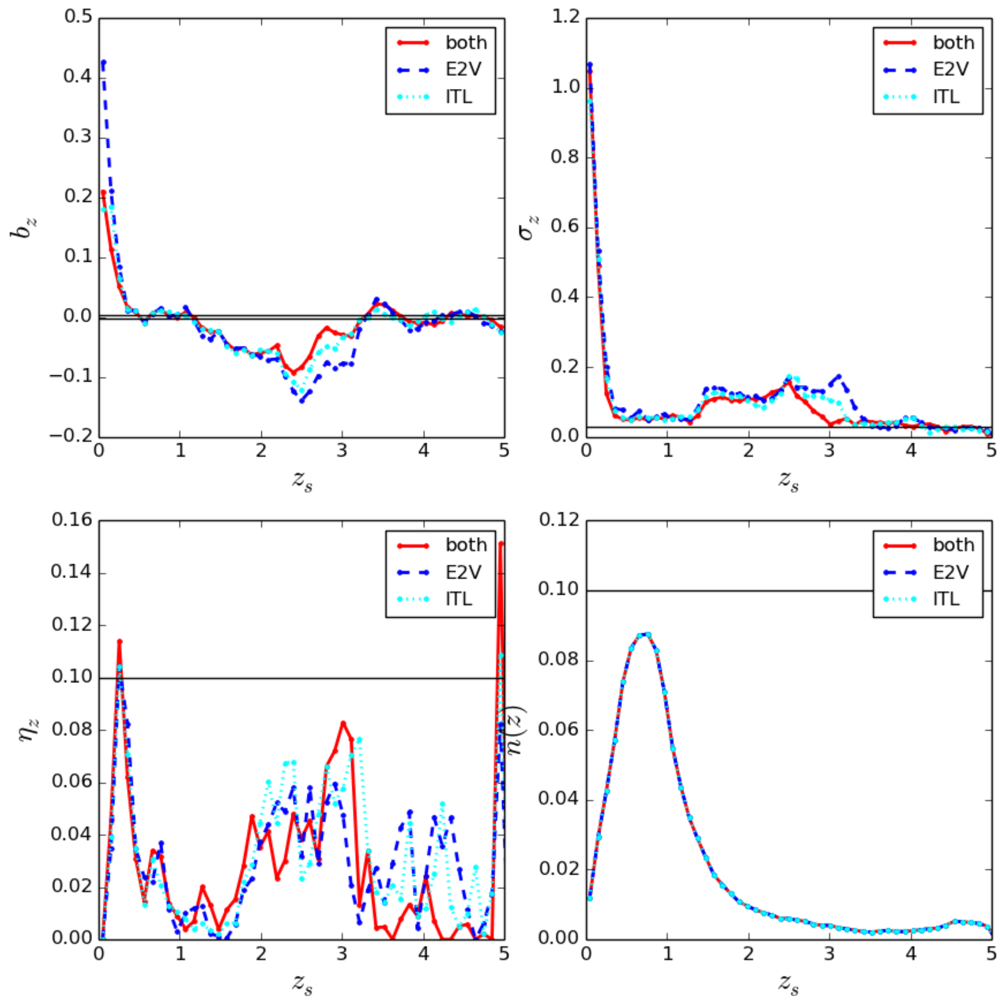

Figure 2 Photo-z precision comparing: (i) only E2V focal plane sensors (blue dashed), (ii) only ITL focal plan sensors (cyan dot-dashed), (iii) a mixture of both sensors, half observations are E2V and half ITL (red solid).

<u>Photo-z Simulations Pipeline</u>

In preparation for running large scale simulations of LSST galaxy photometry Abate developed a code base that performs a set of generic calculations: redshift galaxy SEDs, integrating (quickly and accurately) the amount of flux within a filter transmission function, basic cosmological calculations etc. The code base can be found here:

https://github.com/DarkEnergyScienceCollaboration/PhotoZDC1

A number of tasks were completed towards the photo-z Dark Energy Science Collaboration (DESC) Task Force (run by Abate) that is charged with producing a simulated data set for DESC's Data Challenge 1. The goal of Data Challenge 1 is to test the ability of state of the art photo-z codes to recover a robust estimation of the probability a galaxy is at some redshift. A secondary goal is to propagate the resulting photo-z precision as a function of redshift and galaxy type into other analysis codes (e.g. weak lensing, large scale structure).

The main tasks are summarized below.

- development of emission line model via training a model (following [4]) to assign emission line properties given the principal components of a galaxy spectrum continuum (in collaboration with Z. Liu, C. Schafer, J. Newman at CMU, *Figure 3*)

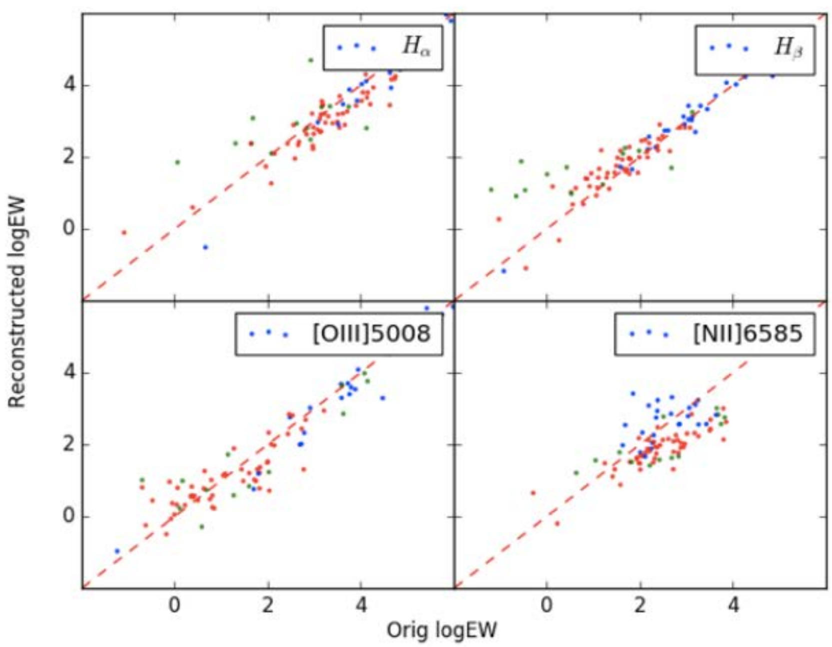

Figure 3 A model trained using SDSS data, where principal components from the continuum spectra are used to predict the measured log equivalent width of a set of emission lines, was applied to a second set of empirical galaxy continuum spectra, the Brown et al spectra [5]. The actual equivalent widths of the "Brown" galaxies (Orig logEW) were recovered robustly.

- development of a simple k-nearest neighbor method to map a simulated galaxy spectrum to a spectrum based upon empirically measured galaxies (Abate).
- development (still underway) of a sophisticated method to achieve the above using diffusion maps and landmarks (in collaboration with J. Shin, A. Lee, P. Freeman at CMU, *Figure 4*)

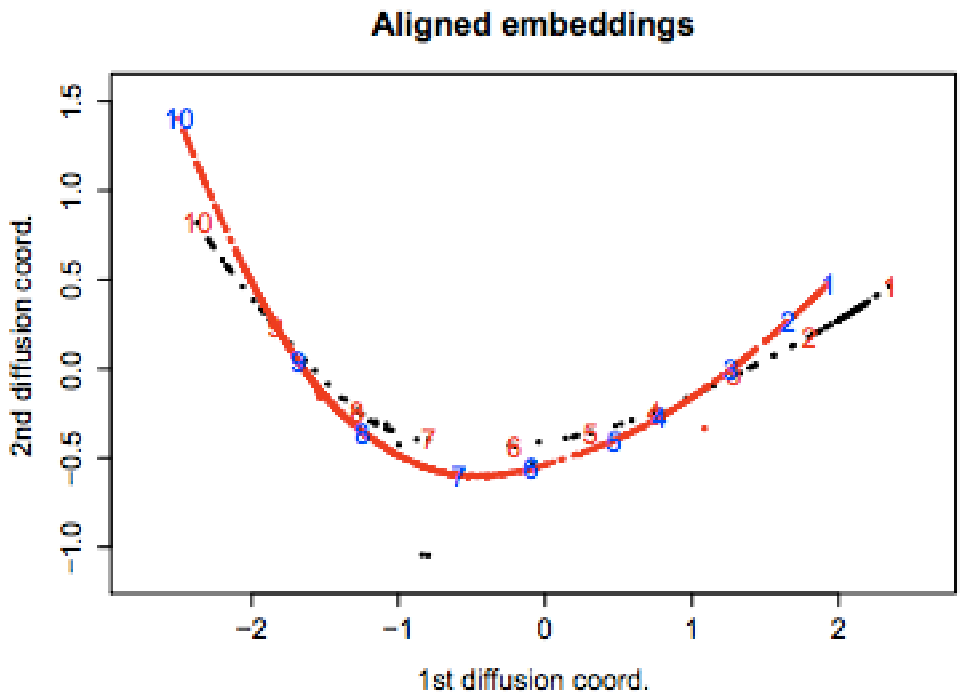

Figure 4 *Aligned embeddings of "Brown" spectra (black points) and simulated spectra (red points) with K = 10 landmarks for each set (shown as blue and red numbers, respectively). This algorithm does not require a training set but the pseudo-reference pairs are found by directly matching K original spectra using the l2-norm in observables space. These alignments will be used to assign a simulated spectrum a spectrum drawn from locally similar empirical Brown spectra.*

- development of simulation pipeline: input cosmological simulation ->output galaxy photometry (Abate, *Figure 5*)
- creation of codes to validate the galaxy colors in cosmological simulations (Abate)
- creation of codes to auto-validate photo-z estimation submission results on the data challenge data set, including testing the robustness of the p(z) from each code (Abate).

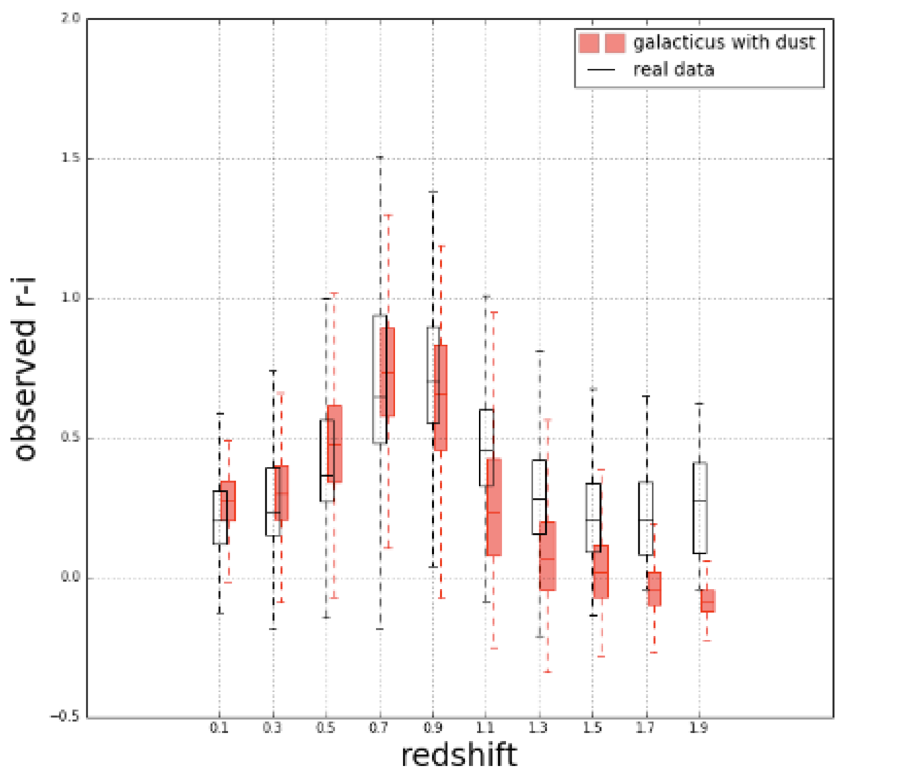

Figure 5 Simulated r-i color distribution from the Galacticus cosmological simulation [6] after mapping empirical galaxy spectra (the "Brown" spectra) onto the simulated galaxies using a simple k-nearest neighbor method (red boxes), CFHT-Subaru test bed data set (black outlined boxes). Both data sets are selected so i<23.5. The color distributions agree fairly well up to z=1 (CFHT-Subaru is incomplete above z=1) even though no tuning has yet been performed on the Galacticus simulation parameters.

Investigation of Atmospheric Transmission on LSST y-band

Abate ran a simplified, yet extreme simulation of the effect of the atmospheric water vapor on the transmission by the LSST y-band: half the observations with extreme water vapor absorption (see *Figure 6*, solid line), half with little absorption (see *Figure 6*, dashed line). It is clear from these two realizations of the y-band that they will both probe slightly different parts of a galaxy spectrum, therefore they contain potentially different *information*.

If there is different information contained within the “extreme” and “normal” y-bands, then data originating from the “extreme” filter should be separable from data originating from the “normal” filter. The first test was checking if a Random Forest Classifier could learn the difference between the data from the two different y-bands (when it was also given the ugriz band data, needed to help break degeneracies).

Four data sets (1 to 4) were simulated that were an increasing challenge for the classifier: different galaxies were distributed over a range of redshifts, and then the observed LSST ugrizy was generated (where the y observation was half the time made in either the extreme version or the normal version).

- Data set 1: no photometric errors, galaxies at ten unique redshifts,
- Data set 2: no photometric errors, galaxies at continuous redshifts,
- Data set 3: photometric errors, galaxies at ten unique redshifts,
- Data set 4: photometric errors, galaxies at continuous redshifts

Figure 7 shows the results in the form of the precision and recall values. For data set 1 (the easy one) the classifier does a perfect job of separating the y-bands (precision and recall = 1; no false positives or false negatives). For data set 4 (the realistic one) however, the recall has dropped to only at around 0.75, meaning only about 75% of the y-bands were distinguishable, and the precision has dropped to 0.65, so only 65% of the y-bands were classed correctly, not a lot better than guessing.

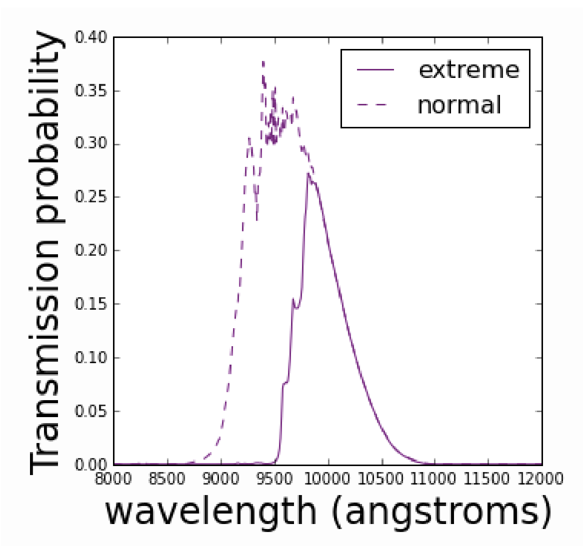

Figure 6 The two y-bands corresponding to extreme water vapor absorption (solid line) and normal water vapor absorption (dashed line).

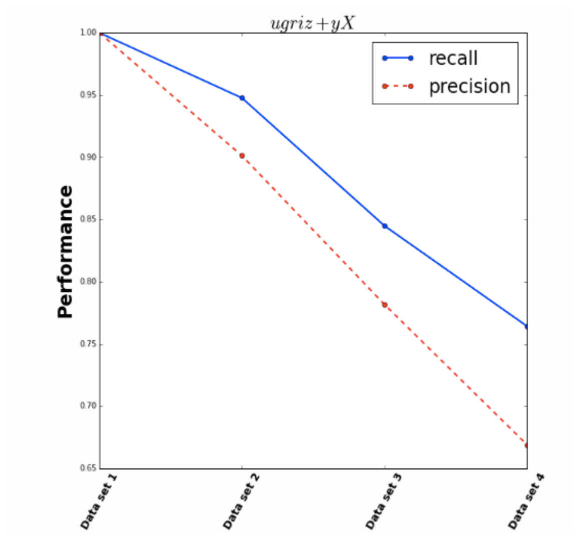

Figure 7 Precision and recall from classing one of the y-bands, as a function of the complexity of the data set. Data set 4 represents a fairly realistic scenario

Though it appears to be small, to confirm if we can use this extra information from the separate y-bands to improve the estimation of the galaxies' redshifts, a Random Forest Regressor was then trained on data set 4. The trained regressor constrained a mapping between the flux observed in each of the ugrizy bands and the redshift. *Figure 8* shows the result after comparing the percentage of galaxies in each redshift bin that were better estimated using either a single y-band (to full depth) or two y-bands (each to half depth). There is no improvement after treating the different effective y-bands separately. These results were presented at the "Foreground Physical Effects on LSST Weak Lensing Science: A Workshop on the Impact of the Last Kiloparsec" at UC Davis workshop in December 2015.

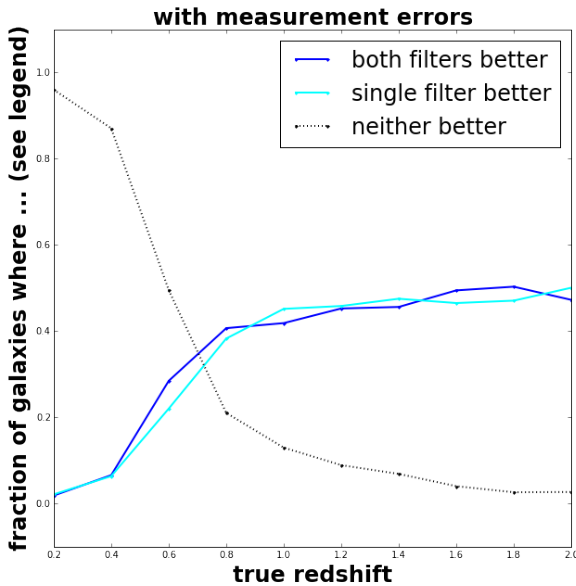

Figure 8 fraction of galaxies where (i) two y-bands to half depth (solid blue) yields better photo-z (ii) one standard y-band to full depth (solid cyan) yields better photo-z, (iii) neither case yields better photo-z (black dotted)

<u>IGM Studies</u>

Although, on average, analysis has been done to study the impact of the intergalactic medium (IGM) on galaxy redshifts, no group has studied the variance associated with the IGM and its effect on the photo-z of very distant galaxies. In the rest frame neutral hydrogen in the IGM may absorb photons with wavelengths 912A to 1216A (the Lyman series to Lyman limit), and at redshifts around z=1.6, this range begins to shift into the LSST filter set. Photo-z at such high redshift have not been widely used beyond the ``drop out" technique to detect very high redshift galaxies (this also relies on Lyman limit absorption from the IGM). As a consequence, the absorption by the IGM is modelled in most photo-z codes only via the Madau law [7], which gives a relation for the mean transmission by the IGM. However the absorption along the IGM varies significantly with the line of sight, e.g. by the presence of a rare but highly absorbing Damped Lyman Alpha system, and this will add extra statistical error into the photo-z estimates. In addition the presence (and distribution) of absorbers will be correlated with the amount of matter along the line of sight. This may lead to a correlation of photo-z error with cosmological probes; an especially sensitive probe may be cosmic magnification which requires a precise knowledge of galaxy number counts at the detection limit of the survey.

We worked to address this issue by generating a series of line-of-sight distributions using the empirically observed distributions of IGM absorbers. More recently, we also included the effects of damped Lyman alpha attenuation. We then calculated the transmission through each line of sight for a source galaxy situated at a range of redshifts. This is shown in *Figure 9*. We then developed a fitting function to return the mean and variance of the IGM transmission given a galaxy redshift and observed wavelength. We also simulated the expected LSST magnitudes of a galaxy at various redshifts through each line of sight transmission, through the mean transmission, and with zero IGM absorption. The observed magnitudes were run through the BPZ photometric redshift estimation software [8] and photometric redshifts were estimated. Using these results we have been able to determine the impact of IGM on photo-z reconstruction.

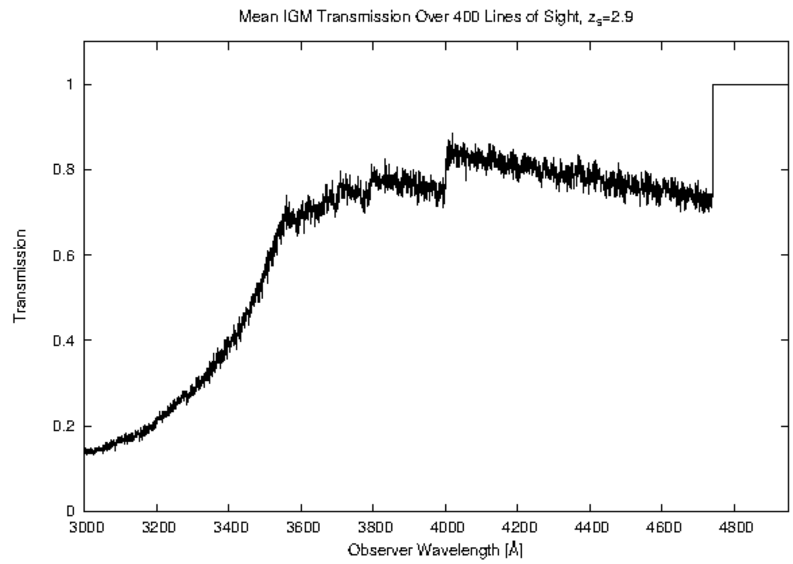

Figure 9 Mean transmission curve for a single galaxy averaged over 400 lines of sight. The galaxy redshift is 2.9.

We have implemented Inoue and Iwata's 2008 [9] Monte Carlo simulation that models IGM attenuation. We simulated 9000 lines of sight and their respective observed transmission. Since IGM absorption depends on the line of sight, we are investigated the stochastic effect of the IGM.

We have also updated the model to their 2014 version [10]. We ran simulations of different galaxies set at a series of redshifts and estimated the contribution of the variation of the IGM to the photo-z errors, in particular the photo-z bias. See *Figure 10*.

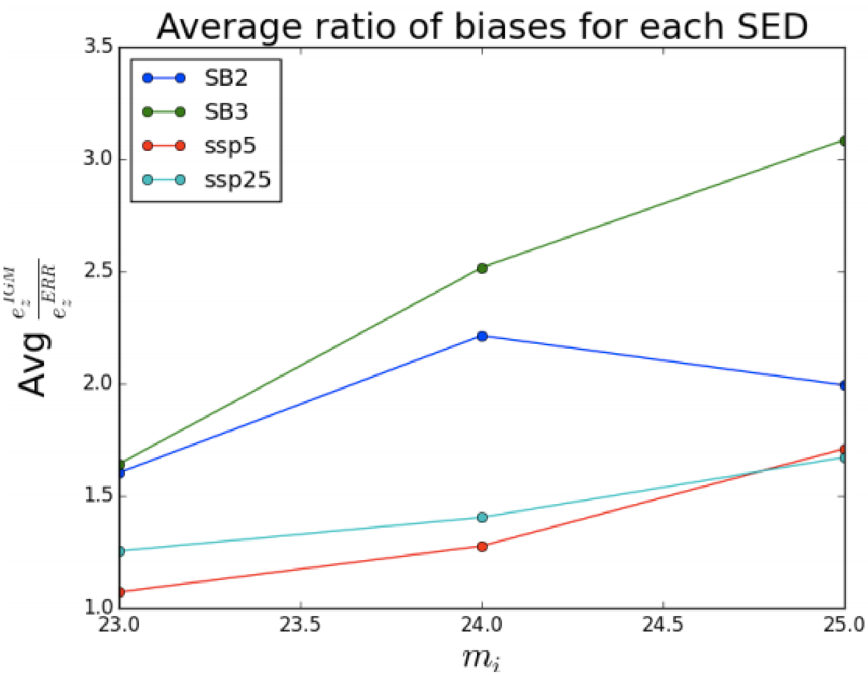

Figure 10 Relative size of photo-z bias due to variation of IGM compared to the fiducial bias expected from a constant IGM with LSST full depth photometric errors. Each line corresponds to a different starburst galaxy spectrum.

Appendix 1: Graduate Student PhDs

During the period of the grant, May 1, 2014 to March 31, 2016, no PhDs were awarded for our Cosmic Frontier work.

Appendix 2: References cited in the report

[1] Cunha, C. E., Huterer, D., Lin, H., Busha, M. T., & Wechsler, R. H. 2014, MNRAS, 444, 129

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

[3] Newman, J. A. 2008, ApJ, 684, 88-101

[4] Beck, R., Dobos, L., Yip, C.-W., Szalay, A. S., & Csabai, I. 2016, MNRAS, 457, 362

[5] Brown, M. J. I., Moustakas, J., Smith, J.-D. T., et al. 2014, ApJs, 212, 18

[6] Benson, A. J. 2012, New Astronomy, 17, 175

[7] Madau, P. 1995, ApJ, 441, 18

[8] Benitez, N. 2000, ApJ, 536, 571

[9] Inoue, A. K., & Iwata, I. 2008, MNRAS, 387, 1681

[10] Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805

Appendix 3: Publications and Reports

Large Synoptic Survey Telescope: Dark Energy Science Collaboration
A. Abate et al.
arXiv:1211.0310

Spectroscopic needs for imaging dark energy experiments.
J. A. Newman et al.
Astropart. Phys., Vol 63, p. 81, 2014

LSST: From Science Drivers to Reference Design and Anticipated Data Products
Z. Ivezic et al.
arXiv:0805.2366

Spectroscopic Needs for Calibration of LSST Photometric Redshifts
S. J. Schmidt, J. A. Newman, A. Abate, & the Spectroscopic Needs White Paper Team 2014
arXiv:1410.4506

Spectroscopic Needs for Training of LSST Photometric Redshifts
A. Abate, J. A. Newman, S. J. Schmidt & the Spectroscopic Needs White Paper Team 2014
arXiv:1410.4498