### 

**Simulating Views of our Universe: A Sustainable Framework for Astronomical Survey Telescopes**

John Peterson (Purdue University), Andrew Connolly (University of Washington), and Garrett Jernigan (Eureka Scientific)

### Introduction

Modern astrophysics has entered a new era of high precision and data volume. With the significant progression in the size of modern wide-field large telescopes as well as the ability to build larger and larger cameras, the astronomical data production rate has exploded. The field is undergoing a dramatic transition in the approach towards analysis. In the past modest sets of astronomical data were collected and carefully studied by individual scientists using specialized data specific algorithms. Current and future data volumes are so vast that all analysis has to be fully automated. At the same time, the scientific goals of modern astrophysics increasingly require higher accuracy. The quantitative investigation of the mysterious dark matter and dark energy, the complete mapping of our Milky Way and Solar System, and the detailed study of other galaxies all the way to the edge of our observable horizon all require more and more precision. In particular, the accurate understanding of the imaging quality (characterized by a point-spread-function (PSF)), color-dependent transmission of light through the Universe, Earth’s atmosphere, and a telescope (photometry), and the angular positions of objects on the sky relative to one another (astrometry) are all critical to the success of modern astronomical scientific measurements. In this environment, where both data volumes are large and the precision and accuracy of measurements is critically important, we have been developing a efficient and high fidelity simulation framework to simulate astronomical images. Simulated images have an obvious advantage over the real images from modern telescopes in that we know the “truth” as defined by a user selected catalog of sky objects. and therefore we can evaluate the accuracy and performance of candidate algorithms. This knowledge canbe used to validate the the accuracy of the analysis of both real and simulated images. The billions of dollars invested in modern astrophysical telescopes would broadly benefit from a generalized well-tested astrophysical simulation framework.

To achieve these goals, we have constructed tools for building large catalogs and databases for state-of-the-art parameterized models of stars, galaxies, and other astronomical objects to generate a synthetic Universe. Then, to produce high fidelity images, we have a novel photon Monte Carlo approach where we numerically follow photons from an astronomical source, through the atmosphere (for ground-based telescopes), through the mirrors and lenses, and into the Silicon of the detector in a series of sophisticated simulation codes. Therefore, we construct a view of the sky by building up the simulated images one photon at a time which ensures high fidelity, The simultaneous requirements of high fidelity and practical numerical efficiency strongly favors the simulation and each separate photon.This novel approach is fast enough with carefully designed algorithms and the use of modern parallel processing. Despite the numerical challenge we have already generated over 15 Terabytes of highly realistic images, which is comparable to sizes of many existing telescope surveys. We use a variety of modern numerical techniques as well as by employing large-scale grid-based computing with thousands of processors. The prototypes of these codes have been written for the Large Synoptic Survey Telescope (LSST), which will be the world’s largest survey telescope when it begins operating at the end of the decade.

With this proposal we plan to extend the framework to encompass a variety of telescopes (some of which are operating now) and user-specified synthetic astronomical catalogs. In the next ten years, eleven separate survey telescopes (VLT Survey Telescope, LSST, UKIRT Infrared Deep Sky Survey, Visible and Infrared Survey Telescope for Astronomy, PanSTARRS 1, PanSTARRS 4, Dark Energy Survey, EUCLID, Wide-Field Infrared Survey Telescope, and Sky-Mapper) will be in operation of construction. Developing a sustainable simulation framework will enable the astronomical community to simulate UV, optical and near infrared views of the sky using a shared software environment. This will enable the community to scale current astronomical research analysis to data sets a 1000-fold larger than current surveys, to understand the characteristics and limitations of current and planned surveys, and to evolve the computational and statistical tools needed to extract the science from these resources as they come on line Maintaining high fidelity is required in this new regime where uncertainties are dominated by systematics and not statistical noise.

### Team Qualifications and Previous Work

The PIs are experts in instrument and simulation in astrophysics, as well as applying those simulations to real data. Peterson performed most of his previous work in X-ray astronomy. The main focus of his X-ray work was the X-ray spectroscopy of clusters of galaxies where his work uncovered the cooling flow problem in cluster of galaxies (Peterson & Fabian 2005). Peterson and Jernigan have collaborated extensively in applying novel photons Monte Carlo methods to astrophysics (Peterson et al. 2004, 2007). Most recently this was in the development of a photon Monte Carlo for the XMM-Newton X-ray satellite. Jernigan has also worked previously in developing detectors and simulations for past and future NASA mission (for example HETE2, RXTE, EXIST, and AXTAR). Jointly Peterson and Jernigan working on the construction of a large cluster catalog using X-ray (Chandra and XMM) and optical data for use in dark energy measurements (Peterson et al. 2009). Connolly is part of an active collaboration between computer scientists, statisticians and astrophysicists at Carnegie Mellon University and the University of Washington. This collaboration, dating to 2001, has demonstrated sustained success and was cited by the President of the American Statistical Association (ASA) as an exemplary interdisciplinary research team. All PIs are members of the LSST consortium. Peterson, Jernigan, Connolly have been collaborating for three years in building the image simulation framework for the LSST project.

Connolly is supported by several NSF grants. AST-0709394, MSPA-AST: Image Coaddition, Subtraction and Source Detection in the Era of Terabyte Data Streams ($427,933, 9/1/2007-8/31/2011) is the most closely related of these grants. Outcomes from this work include: the development of non-parametric techniques for the detection of sources within sequences of astronomical images through the use of image coaddition and subtraction , and algorithms for measuring the clustering of galaxies using n-point correlations functions that scale to high-performance parallel architectures . Connolly is currently the simulation scientist for the Large Synoptic Survey Telescope. He is also exploring the use of Hadoop as a model for the processing of astronomical data .

Beyond the computational and astrophysical aspects of this research the PIs have demonstrated a commitment to developing and releasing applications for data intensive cosmology and for integrating research and education (e.g. Connolly was the technical lead for the development of Sky in Google Earth, Google Sky http://earth.google.com}), which enables the seamless exploration of astronomical images of the sky).

This proposed project is possible because of prior NSF-supported work to develop a high fidelity simulator of the astronomical sky, the atmosphere, the LSST telescope and camera for LSST. The image simulation team is composed of ~15 scientist, students and staff at three locations. The principals for this proposed project lead these three teams: Connolly of University of Washington, Peterson of Purdue University and Jernigan of UC Berkeley and SLAC. The LSST project continues to support the improved fidelity and efficiency of the image simulator. The funding for this proposed new project would be entirely devoted to generalization of the image simulator to include telescopes and cameras other than LSST, and improve the ability of the user to easily extend the catalogs of spectral and spatial models of astronomical sources . This extended scope of the simulator would provide this resource to a broader range of optical astronomers. The long-term goal is to develop an image simulation tool that the full NSF astronomical community can use and improve as an open source project similar to IRAF, sets of IDL analysis packages and the tools that comprise the Virtual Observatory. This simulation framework is designed to interoperate with all the existing analysis platforms.

### A High Fidelity Simulation Software Framework

To achieve the goal of simulating accurate astronomical images, we have adopted a complete end-to-end approach. The data quality of real images is known to be a complex mixture of effects from the atmosphere, optics, and detector and it is important that all are included in a self-consistent manner. For example, one of the primary science drivers for the LSST that mandates high fidelity is the measurement of cosmic shear due to gravitational lensing of galaxies. Figure XXX (also figure 14.3 from the LSST Science Book) shows the lensing power spectra as a function of angular scale for two different example cosmological models, The precision of ellipticity measurements of each galaxy in an individual 15 second LSST exposure must reach a range near 1 part in 104 to match the errors show in the figure. Accuracy that matches this precision requires a science pipeline that includes a complex calibration procedure that estimates both statistical and non-statistical errors. Simulated images of stars (point-spread-function, PSF, calibrators) and galaxies (sources of weak lensing signal) must, therefore, be high fidelity realizations of the actual PSF (which is non-Gaussian and complex) at a level of accuracy that far exceeds the requirements on the ellipticity measurements. A low fidelity simulator that represents the atmospheric effects of the PSF as a Gaussian may provide a useful estimate of the precision and quantity of data needed to reach a measurement goal. However, a high fidelity photon-by-photon derived PSF is the only way to prove that the needed “accuracy” can be achieved. We are not claiming that all predictions of science results require high fidelity but that the primary science goals, that push the limits of the survey, do need to reach this level of detail. Expensive, modern survey projects require an improved ability to predict the performance of the design prior to construction.

Below we outline our prototype codes and describe how each part needs to be generalized to build a tool for the community.



Figure XXX: The lensing power spectra constructed from galaxies split into three broad redshift bins: z < 0:7; 0:7 <

z < 1:2, and 1:2 < z < 3. The solid curves are predictions for the\_fiducial\_CDM model and include nonlinear

evolution. The boxes show the expected measurement error due to the sample variance and intrinsic ellipticity errors The thin curves are the predictions for a dark energy model with w = -0:9. Ten years of LSST data are needed to reach the required accuracy for shear measurements of galaxies that would match these measured errors.

In Figure 1 we show the overall flow of information through our simulation framework; broken into three separate components or subsystems. In the first subsystem (referred to as the sky database) we store astrophysical catalogs as SQL databases. These data are derived from cosmological N-body simulations, models for Galactic structure, and simulations of Solar System sources. Variability (including transients sources), and proper motion are incorporated within these models together with the spectral, photometric, astrometric and morphological properties of each source. In total the astrophysical databases contain approximately 5 TB of observational and simulated data.

The second subsystem (referred to as the instance catalog generator) incorporates the observing strategy and observing conditions associated with a telescope. This could be as simple as a single pointing on the sky or a sequence of observations with a specified cadence and survey geometry. Each simulated pointing provides a position and time of the observation, together with the appropriate sky conditions (e.g. moon phase and angle, and sky brightness). Querying the sky database we can derive, positions (propagated to the time of observation), morphologies, and magnitudes (for a given filter and after applying variability) for any source in the field. The result of this is a parameterized view of the sky that can be expressed in the form of catalogs with the appropriate physical and statistical properties (including light curves) or through the generation of images using the third component of the simulation framework.

Imag Sim Flow Chart Nov 2008.pdf

**Figure 1** Data flow for the simulation framework. The base catalog contains the underlying astronomical catalogs that are stored in SQL databases. These catalogs are queried based on the pointing of the telescope to generate instance catalogs that are either formatted for output for users or used as input to the image simulator. Images are simulated using fast ray-trace algorithms.

In the final subsystem, image generation, photons are drawn from the spectral energy distributions associated with each source and these photons are ray-traced through the optical system before being converted into electrons by simulating the camera physics. The resulting images are “readout” using a model of the electronics of the camera and output as individual FITS images (the astronomical standard for binary images).

The framework, as designed, is extensible and scalable (i.e. it is capable of being run on a single processor or across many-thousand core compute clusters). Computationally intensive routines are written in C/C++ with the overall framework and database interactions using Python. Access to the sky catalogs is through SQL queries. The purpose of this design is to enable ease of access for users as well as the generation of a wide range of data sets for use in evaluating the design of a telescope or survey; from all-sky catalogs used to test photometric calibration, to time domain catalogs to study the impact of survey cadence on our ability to recover variable or transient sources, to images used in testing weak lensing shear measurements.

### *Parameterizing the Universe*

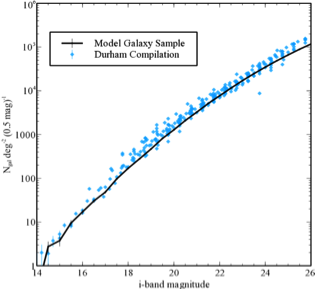


Figure 2 Comparison of the input sky database for the simulation framework with observations from deep imaging and spectroscopic surveys. The left panel shows the distribution of sources for a 2x2 degree region of the sky. The clustering is derived from the N-body simulations of de Lucia et al (2006). The central panel compares the galaxy number counts from the catalog with a compilation of observations wide field and deep imaging surveys. The right panel shows the redshift distributions for the galaxies. The form of the redshift distribution matches the DEEP2 redshift survey.

At the core of the simulations are the distributions of extragalactic, Galactic, and Solar System sources. Our existing simulation framework incorporates galaxies derived from the Millennium simulations of de Lucia et al (2006). These models extend dark matter N-body simulations to include gas cooling, star formation, supernovae and AGN and are designed to reproduce the observed colors, luminosities, and clustering of galaxies as a function of redshift. The underlying catalogs extend to r=28 and cover a redshift interval 0<z<6. Spectral synthesis models (Bruzual and Charlot 1993) are used to generate spectral energy distributions for all sources within a catalog (derived independently for the bulge and disk components and including the effects of internal reddening as a function of inclination). Morphologies are modeled using linear combinations of Sersic profiles together with a point source component that accounts for AGN activity. On ingestion into the database the simulated catalogs are compared against the observed properties of galaxies (as a function of magnitude, color and redshift) and the source densities adjusted to match existing deep imaging and spectroscopic surveys. Figure 2 shows the resulting spatial distribution of the resulting galaxy catalogs, the number-magnitude relation and the redshift distribution for a representative sample of galaxies.

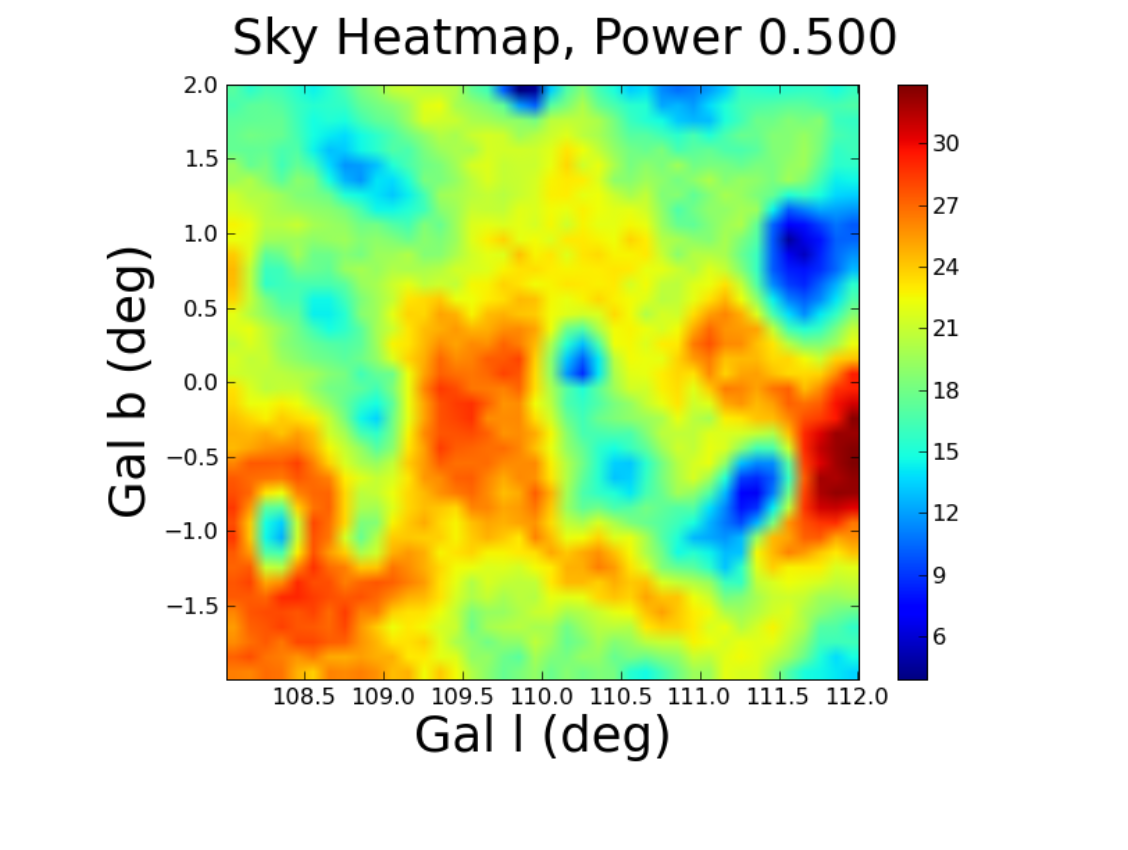
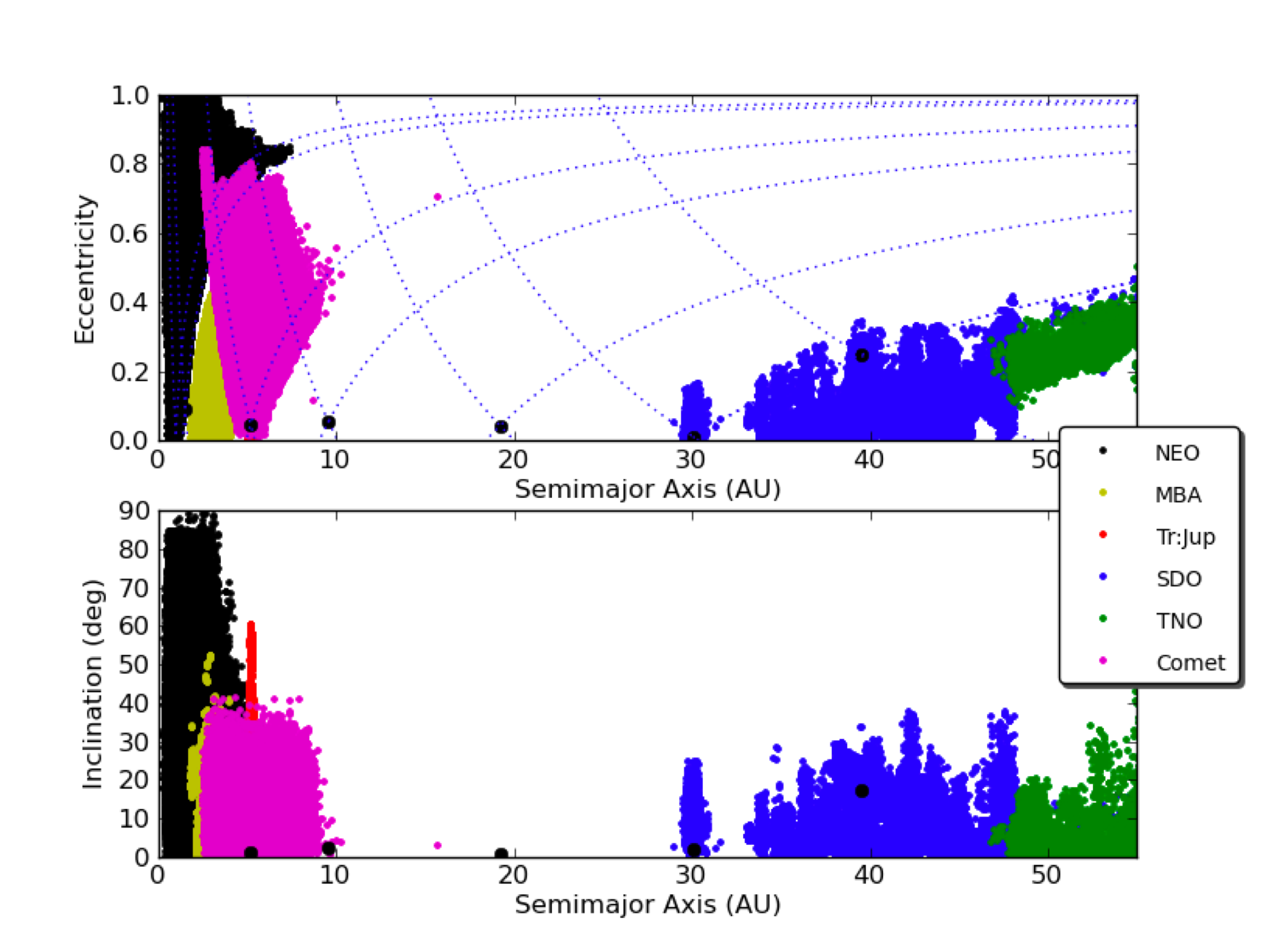
 

Figure 3 The density of stars for a region in the Galactic plane. The variation in density reflects the 3D distribution of dust in the model.

Galactic structure within the simulation framework is modeled using the work of Juric et al (2008). These simulations are designed to match the observed distributions of stellar colors to depths exceeding r>24. Each star in the simulation is matched to template spectral energy distributions (SEDs) using both spectral synthesis models (Kurucz 1993) and empirical spectral models. Populations for main-sequence F, G, and K stars, RGB stars, blue horizontal branch stars, RR Lyrae, and Cepheids are incorporated within the current database. White dwarf stars are taken from Bergeron et al. (1995) and SEDs for M, L, and T dwarfs are generated from a combination of spectral models and empirical spectra from the SDSS. Proper motions for each star are based on the kinematic survey of Bond et al. (2010) and light curves are assigned to variable stars based on their stellar colors. Galactic reddening is applied using the three-dimensional Galactic model of Amores & Lepine (2005). To provide consistency with extragalactic observations the dust model in the Milky Way is renormalized to match the Schlegel et al. (1998) dust maps at a fiducial distance of 100 kpc (Figure 3 shows the impact of the reddening on the density of stars within a region close to the galactic plane). With this level of accuracy in stellar density we can characterize not just the PSF of the optical system but also how well you can use these stars to constrain the PSF (including spatial variations across a focal plane) as a function of stellar spectral type, position on the sky, and survey geometry.

Asteroids are simulated using the Solar System models of Grav et al. (2007). They include: Near Earth Objects (NEOs), Main Belt Asteroids, the Trojans of Mars, Jupiter, Saturn, Uranus, and Neptune, Trans Neptunian Objects, and Centaurs. Spectral energy distributions are assigned using the C and S type asteroids of DeMeo et al (2009). 11 million asteroids are stored within the database catalog (with positions sampled once per night for a ten year duration). Positions are returned given a query specifying rgion of the sky and time of observation using a fast polynomial interpolation scheme. With typically 8000 asteroids per ten-degree region of the sky accurate source positions can be returned in few milliseconds. The right panel of Figure 3 shows the distribution of eccentricities as a function of semi-major axis for the simulated populations.

The simulated data sets incorporated within the current design are broad in scope (from our own solar system to the distant universe). They represent, however, only one view of the universe. A primary goal of this proposal is, therefore, to provide a documented and scalable API that will enable users to incorporate their own datasets into this simulation framework. In this way, the community can ingest different cosmological models (e.g. varying the properties of Dark Energy), add new populations of sources (e.g. new types of astrophysical transients or variable sources), or change the dynamical models that underlie the Galactic structure (e.g. to find local dwarf galaxies or streams of stars) and automatically determine how detectable these effects would be for range of telescope and survey designs. In detail we propose to:

* extend the underlying spectral energy distributions of the galaxy and stellar populations to the near-infrared to enable the simulation of optical and infrared telescopes.
* implement models for gravitational shear that will be incorporated at the catalog and image level (for studies of dark energy and dark matter). The range of shear models developed will include: constant shear for simplified tests (e.g. GREAT10 2010), shear maps that have been integrated over redshift for a given cosmology and power spectrum (i.e. a single shear screen), and a series of evolving shear screens for different cosmological models that are applied as a function of redshift and are consistent with the expected power spectrum and growth of structure.
* enable external users to upload a table of sources (e.g. a population of stars or a new cosmological model) into the database framework. Table uploads will be provided by a simple API that will automatically index the resulting catalogs using a Hierarchical Triangular Mesh (Budavari et al 2010) to maximize query speeds.
* enable users to upload high resolution images into the simulation framework to determine the impact of different telescope and atmospheric models on the resulting image resolution (e.g. how well can a ground based telescope separate pairs of lensed quasars).

### *3.4 Photon-based Image Simulations*

The framework described above provides a parametric view of the sky that can be used to generate catalogs of stars, galaxies and Solar System sources across the sky. From these catalogs we generate images by drawing photons from the spectral energy distribution of

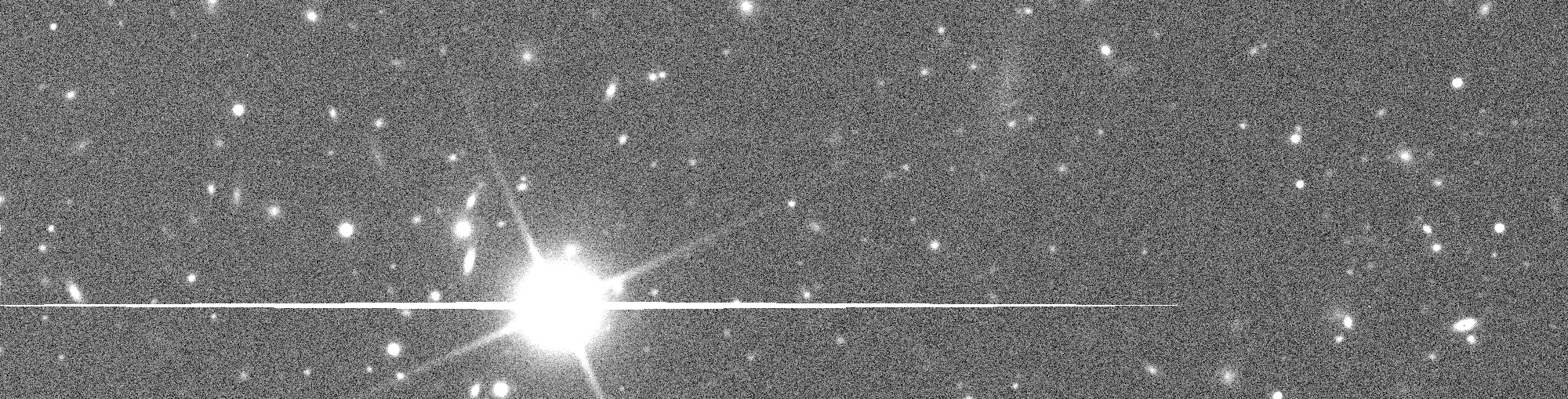


Figure 4 The top panel shows a three-color simulated image (g, r and i) of a full chip of stars and galaxies that incorporates all of the effects of the stellar and galaxy source densities, and full raytracing of every photon through the telescope and camera optics. The bottom panel shows a region of this chip that highlights the diffraction spikes and saturation trails from a bright star where the size is much smaller to show more detail.

each source (scaled to the appropriate flux density based on apparent brightness and accounting for the spatial distribution of light for extended sources). Each photon is then ray-traced through the telescope and camera to generate a CCD image. Examples of these images are shown in Figure 4. In the following sections we describe the details of the image simulation, how it will be modified to model other telescopes.

To simulate images with high-fidelity, photons are reflected and refracted by the optical surfaces within the telescope and camera (see Figure 5). The mirrors and lenses are simulated using geometric optics techniques. Fast techniques for finding intercepts on aspheric surfaces and altering the trajectory of a photon by reflection or wavelength-dependent refraction have been implemented to optimize the efficiency of the simulated images. Each optical element can move according to its six degrees of freedom and all surfaces incorporate uncertainties about their shapes through a spectrum of perturbations (enabling the modeling of mirror misalignments, tracking errors, and pointing uncertainties). Wavelength and angle-dependent transmission functions are incorporated within each of these techniques including the simulation of the telescope spider. Photons are destroyed, in a Monte-Carlo sense, when they pass through the filter in accordance with the wavelength and angle-dependent transmission functions. Ray tracing of the photons continues into the silicon of the detector with conversion probabilities, refraction (as a function of wavelength and temperature) and charge diffusion within the silicon modeled. After conversion photons are pixelated and, in the readout process, blooming, charge

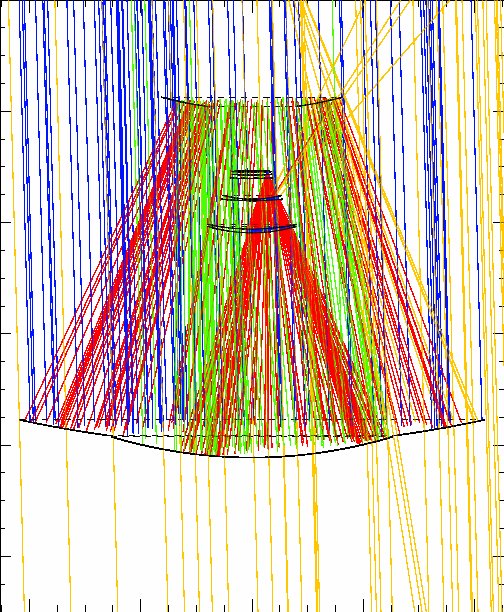
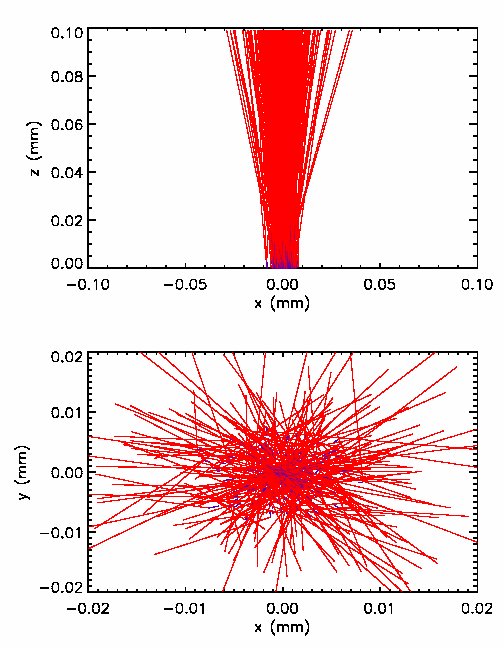
 

Figure 5 The left panel shows a ray trace diagram of photons reflected and refracted through mirrors and lensing using the LSST configuration. The blue, red, and green rays are the successful bounces and the yellow lines represent stray light. The right pane shows the detector simulation. The top right panel shows the photons as they enter the CCD silicon, the purple lines are the photon trajectory as it is refracted in the silicon, and the red line shows the path of the photo-electron after the photon is converted. An electric field drifts the electron to the top. The bottom right panel shows the top view of the same set of interactions.

saturation, charge transfer inefficiency, gain, hot pixels and columns, and quantum efficiency variations are all simulated. The sky background is added as a post-processing step with the sky background generated based on SEDs for the full moon and the dark sky (though the sky background can also be ray-traced at additional computational cost). The simulator generates about one million photons per second on a typical workstation and even faster on a GPU. In Figure 6 we illustrate how we can evaluate the impact of individual components of the simulator on, for example, the PSF by turning on or off these effects. In the example shown in Figure 6 we progressively add in the effects of errors in tracking, diffraction, mirror misalignment, charge diffusion in the detector, and then (as this simulation was implemented for a ground-based system) the impact of the atmosphere and pixelization of the images. Each of these effects could have been evaluated in isolation or as part of the complete optical system.



Figure 6 The image simulation framework is flexible enough to switch off and on different optical components. From left to right and top to bottom we show the resulting PSF as we progressively add more components to the optical path (including perturbations in the optical surfaces and a six-layer atmosphere).

Our goal is to take this prototype framework and develop a general system capable of simulating any ground-based or space-based imaging system. As noted previously, the prototype implementation has been developed for as part of the LSST project. To generalize the code we will do the following:

To generalize this work we propose to:

* Implement a generalized interface that is agnostic to the parameters that are used in defining the telescope design. In such a way we can separate the ray-trace algorithms from the telescope description and enable a broad range of different designs to be incorporated at little extra effort. We have already started this work with the implementation of the Subaru telescope as shown in Figure 7. Glasses with different properties, as well as different surface descriptions had to be generalized for this purpose.
* Provide a simplified interface to the image simulation framework that is easily accessible by an external user (including predefined telescope designs and simple telescope models with representative characteristics such as aperture, camera and pixel size and focal length).
* Extend the properties of the surface coatings and detector physics to the entire UV/optical/IR bands.
* Improve the physics of the simulations that may impact some telescopes more than others. For instance, we have a Monte Carlo diffraction model using a numerical technique of Freniere et al. 1999. The effects of telescope diffraction, however, are smaller for larger telescopes, so this may require some additional algorithm improvement for space-based or smaller telescopes. We also have a generalized model of telescope tracking and surface perturbations representing the different thermal and mechanical perturbations that can affect the orientations and shapes of optical elements. For each telescope, however, there may be details that require a slightly different implementation. Camera defect models may need to be improved as well to accommodate different detector design details.
* Extend the atmosphere model for all sites. The current model of the atmosphere includes the structure function of the turbulence, the composition of important optical absorbers, and a profile of the wind velocity all as a function of height above the location of the telescope. The values of the parameters for these components of the atmosphere have significant effects on the PSF of images and also on the loss on light that affects the correct simulation of throughput and therefore photometric accuracy of the simulated images. The current model that sets these values is limited to the planned location of LSST. Some of this information is obtained from sources on the Internet that archive past information about the atmosphere for any location on the earth's surface. We will generalize the process of acquiring this web-based information so that the new extended simulator can locate telescopes at any place on the earth's surface. This extension of the process for setting the atmospheric parameters would support the high fidelity simulation of images for any existing or planned telescopes. It would also allow for the simulation of any planned telescopes at multiple possible locations that would aid the planning for future telescopes. Also, by simulation of existing telescopes we can use the existing archive of images to verify and improve the models of the atmosphere.

We propose to implement the several telescopes during this proposal. We have first started with the Large Synoptic Survey Telescope (LSST). As shown in Figure 7, we have made a prototype implementation of the optics, coatings, and camera of the 8-m Subaru telescope by working with Satoshi Miyazaki (NAO, Japan). Subaru is currently operating, so the simulations can be compared with real data. We also plan to implement the survey telescope, Pan-STAARS, a new survey telescope, the Dark Energy Survey Telescope (DES)., another large survey telescope. An ideal space-based telescope would be the future James Webb Space Telescope (JWST). Three telescopes that have completed large surveys and would complete our implementation would be the Sloan Digital Sky Survey (SDSS), the CFHT survey, and the UKIDDS infrared survey. In each case, we will obtain telescope designs and instrumentation details from the relevant experts, and continue to generalize the software infrastructure discussed above.

subarufull.pdf

**Figure 8:** Implementation of the Subaru Telescope optical design using the same photon simulation code, which is described by a single mirror (orange) and a set of complex lenses (shown in light blue near the top). The system successfully focuses to a few micron accuracy.

## 4. Development of a Sustainable Simulation System

The prototype software system described in the earlier sections of this proposal is designed to model the optical performance of the LSST (from cosmological simulations to resulting catalogs and images). The goal of this proposal is to extend this prototype to provide a general and sustainable simulation framework for the whole of the astrophysics community that is capable of modeling, to high fidelity, the properties of a new generation of imaging and, potentially, spectroscopic surveys.

***4.1 Existing Simulation Frameworks***

To the best of our knowledge there are no simulation frameworks in existence that can take a source catalog (stars, galaxies and solar system objects) and generate a series of observations with sufficient fidelity that we can utilize these to study the cosmological constraints of a given astronomical survey. Existing simulation frameworks either focus on the engineering aspects of the optical performance of the telescope (e.g. Zemax) or produce simplified analytic approximations for the performance of the optical system (e.g. the DES simulation framework; Abbott et al 2005). Engineering simulations characterize optical performance by ray-tracing the photons along the light path. They do this, however, for representative point sources; generating “spot diagrams” of the point spread function (PSF) of a telescope for a small number of positions on the focal plane. This is ideal for characterizing an optical system, but does not incorporate the other physical effects that impact the properties of actual images. The interplay between the properties of the astronomical sources (e.g. galaxy shapes, cosmology, stellar densities, background emission from the sky) and the resulting measured characteristics (e.g. the distortion of the galaxy shapes used in characterizing gravitational weak lensing) are not included within these simulation systems. Ray-trace algorithms are optimized for small scale engineering questions but do not scale to questions concerning the simulation of the wide-field cosmological properties of the telescope. Engineering simulations are ideal for designing an optical system but are not designed (nor used) for understanding and optimizing the scientific quality of a survey telescope.

Current astronomical simulations (that cover representative volumes of the universe) deal with issues of scalability by simplifying the optical model for the telescope. For example, the Dark Energy Survey (DES) simulations (Abbott et al 2005) do this by assuming analytic models for the point-spread function and apply these models to the outputs of cosmological simulations. This increases the speed of the simulation but at the cost of fidelity. Variations in the PSF due to the surface height of the detector, changes in the effective throughput of the system due to scattering off the optical surfaces, the impact of tracking and guiding error, optical perturbations, are all difficult to model accurately within these analytic approximations yet, as we noted previously, it is these very effects that will limit the performance of these telescopes – we are in an era where the science is governed by systematics and not statistical noise.

A consequence of this is that, while simulations have been undertaken for some astronomical systems (either at an engineering level or in simplified ways) they do not have the fidelity to accurately model the limitations of proposed telescopes and surveys, nor to produce representative data sets which can be used to evaluate how well a telescope can reproduce its scientific goals. This explains why simulation frameworks are not shared amongst different programs (they do not have the fidelity to model a general optical system in any detail). When cosmology is entering an era of high precision, when billions of dollars are being invested in observational systems that will lead to thousands of breakthrough is astrophysics over the next decade, it is clear that we need to be able to model and understand the properties of these systems (both capabilities and limitations), in detail, prior to them coming on-line.

***4.2Sustainability and development of the simulation framework***

We will build upon our current simulation work in designing the generalized simulation framework. Computationally expensive ray-trace code is written in C/C++ (to optimize speed). Framework code is written in Python with simple interfaces to the C/C++ algorithms. This design minimizes the learning curve for users of the system while enabling power users to access the full sophistication of the software. SQL Alchemy (http://www.sqlalchemy.org/) is used to abstract database access so that querying of the databases is agnostic of the underlying database technology (in fact Postgres, MySQL, and Microsoft SQLServer have been used in the current implementation).

The simulation framework has been designed with a goal of portability. Core algorithm code (C/C++) has been compiled and run on a range of systems including: most major flavors of Linux and Mac OS X. All libraries used in the code (FFTW, http://www.fftw.org; FITSIO, http://heasarc.gsfc.nasa.gov/fitsio/) have long-standing and actively developed cross-platform libraries. The inner loops of the codes are well-optimized to increase maximal computing efficiency. with a large amount of the software architecture written for grid-based computing simulations. In addition, we have recently improved the efficiency of the code in multi-core environments and to enable rapid simulations of individual images the ray-trace code was branched and ported to run under GPUs using the CUDA compilers. This version of the code has been run on individual (consumer) graphics cards and on Tesla clusters.

We have made great strides in implementing the code on a variety of grid-based computing systems as our baseline simulation platform. Grid-based computing is ideal because each processor can work on an image on the piece of an image, and modern large-scale scripting languages (CONDOR or PBS) can handle a complicated hierarchy of executables and file transfers. The photon simulation code is currently 11 different codes each requiring access to different data and having a different execution scheme. We have used the CONDOR-based system on the DiaGrid led by Purdue (see the accompanying letter of support), which also has nearly seamless portals to the Teragrid and the Open Science Grid. We have also ported the code on a cluster using PBS-submission scripts at the University of Washington as well as a large grid of computers at the Stanford Linear Accelerator Laboratory. A variety of experience has allowed us to build robust automated scripts, and handle the variety of I/O bottlenecks, rare computing failures, and file organization issues. We have utilized about 4,000 processors multiple of the various grid-based systems during multiple runs, allowing us to generate data at rates as high as 0.4 TB per day, which we think exceeds data production rates of all existing telescopes (but not necessarily future telescopes). As part of this proposal, we will make the use grid-based computing a core- component of the software and will simplify the access to and instantiation of the individual algorithms for a general user.

All code currently developed is available in a publicly available SVN repository (http://dev.lsstcorp.org/trac/browser/imsims/) and includes tagging of periodic releases and a ticket tracking system. Sustainability of the code base will require separating both the telescope-specific data input interface and the user-specified catalogs interface from the core code that needs to be completely general. Precise documentation will be important for this as will a redefinition of the interfaces to make them less specific for the LSST.

**5. Project Plan**

In the preceding sections we describe both the need for this simulation framework in astrophysics and the prototypes system that we have developed to address these needs. The goal of this proposal it to take this prototype system and create a sustainable software product that can be used within the broader astronomical community (for a range of telescopes and instruments).

***5.1 Development Plan and Engineering Plan***

The development of the simulation framework will follow the timeline used for the LSST simulations. Each year will be broken into 4 intervals of 3 months. Two of these are devoted to feature development, and two to the creation of simulated data products, releases of the code and data to the community, and the evaluation of the simulator fidelity through comparisons with existing astronomical data.

Every six months a request for feature development will be made to the astronomical survey groups (including, but not limited to, the science collaborations of the DES, LSST, and PanSTARRS communities). This amounts to over 500 astronomers active in survey astrophysics. We will combine and prioritize these feature requests to plan the development work over the following three months. At the end of development window, work would transition into delivering the documentation and data associated with the current state of the simulation framework and the generation of representative data sets and tools that can be used to analyze these data. This development would be synchronized with the LSST work such that at the end of every six months an LSST data set would be released together with simulations of the corresponding parts of the sky with additional telescopes.

Prototypes of each of these components (design requests, code development, data generation, data release and feedback) are already in place through the LSST analysis. The access to the LSST communities provides a large and natural audience to seed the user and community interaction with this work (we currently have responded to 10s of requests from the LSST community for special case simulations and the inclusion of new astrophysical populations during the prototype work). We will build upon this work for the general simulation framework by expanding the community access to the prototype, and developing user forums for feedback to the released software.

For the three years of development we will; **Year One**: focus on developing the prototype code into a documented and unit tested framework, extend the framework to one additional telescope (Subaru), set up the data delivery and feedback mechanisms, and generate documentation for the system, **Year Two:** develop simplified interfaces to enable ingestion of external catalogs and telescope models, simulate and release data for the Dark Energy Survey telescope and implement features based on community feedback, **Year Three:** extend the number of cosmological and astronomical populations supported within the databases (including supernovae and gravitational weak lensing), and to improve the scalability of the system to enable large scale simulations as well as instant generation of individual images for small scale simulations.

All code will be accessible under a public code repository (currently svn but a transition to github will be considered at the end of the first year with the initial release). Metrics for tracking the development of this process include: the setup of a user forum and ticketing system in the first six months, the delivery of the code, documentation and representative data every 6 months, the release of the Subaru simulator at the end of year one, the release of the DES simulator at the end of year two, and tracking of the number of feature requests for the enhancements to the simulator by the user group.

***6. Broader impacts***

The primary long term goal of this proposed project is to leverage the current work on the LSST high fidelity image simulator and thereby create a new open source tool that would expand the use of the image simulator eventually including many NSF telescopes and cameras. We envision that the open source concept will encourage others to add the information to simulate other existing and planned telescopes and camera. Ideally this tool would evolve into a broad simulation tool that has a community role that is similar to the analysis packages of IRAF and IDL for image analysis. A common framework for simulation will enable scientists from the wider astronomical community to test a variety of science goals and analysis software in a controlled way for a wide range of planned and existing telescopes. In simulated images the truth of what is simulated is known. Studies can then be used to understand the advantages and deficiencies of analysis methods before they are applied to real astronomical images (and before millions of dollars are spent on the construction of these experiments). This work has the potential to impact all astronomers who will use the next generation of surveys. We will also configure the output of the image simulator to be fully compliant with the Virtual Observatory standards and other community defined and supported software standards.

Ideally the utility of this evolving framework would be the tool of choice for all phases of future projects including writing proposals, doing engineering trade studies for newly proposal designs and sites for telescopes, evaluating improvements and modifications of cameras and spectrometers on existing telescopes, and testing new science analysis pipelines which are becoming an ever larger portion of the work of modern astronomical research. Beyond astrophysics, techniques developed as part of this work (fast ray-tracing codes, use of large astronomical databases, databases for moving sources, and open-source software and community development) will have application within any data of the intensive fields in physics and biology that rely on extensive simulations.

**NEED MORE ON BROADER IMPACTS NEED SOMETHING ON EDUCATION – JOHN YOU HAVE LOTS OF UNDERGRADS> CAN YOU WRITE ABOUT THEM AND HOW THIS PROVIDES LOTS OF REAL WORLD EXPERINCE OF THE PHYSICS OF TELESCOPES**

**(I added the last general paragraph – it would be nice to say something less general – it is somewhat redundant with the previous paragraph – I was trying to make a distinction between the goal and the approach) -**

References

**“The Dark Energy Survey”, Abbott, T., et al, 2005,** In: *AIP Conf. Proc.* 842 (2006) pp.989-991 In: [17th International Conference on Particles and Nuclei](http://cdsweb.cern.ch/record/771496), Santa Fe, NM, USA, 23 - 30 Oct 2005, pp.989-991

“Searchable Sky Coverage of Astronomical Observations: Footprints and Exposures”, Budavari, T. and Szalay, A.S. and Fekete, G., 2010, PASP, 122, 1375