UCRL-TR-228468

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

# The Large Synoptic Survey Telescope and Foundations for Data Exploitation of Petabyte Data Sets

Kem H. Cook, Sergei Nikolaev, Mark E. Huber

February 28, 2007

**Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

**Auspices Statement**

This work was performed under the auspices of the U. S. Department of Energy (DOE) by the University of California, Lawrence Livermore National Laboratory (LLNL) under Contract No. W-7405-Eng-48. The project (04-ERD-070) was funded by the Laboratory Directed Research and Development Program at LLNL.

# FY06 LDRD Final Report
# The Large Synoptic Survey Telescope and Foundations for Data Exploitation of Petabyte Data Sets
# LDRD Project Tracking Code: 04-ERD-070

# Kem H. Cook, Principal Investigator
# Sergei Nikolaev & Mark Huber, LLNL Co-Investigators



### Abstract

The next generation of imaging surveys in astronomy, such as the Large Synoptic Survey Telescope (LSST), will require multigigapixel cameras that can process enormous amounts of data read out every few seconds. This huge increase in data throughput (compared to megapixel cameras and minute- to hour-long integrations of today's instruments) calls for a new paradigm for extracting the knowledge content. We have developed foundations for this new approach.

In this project, we have studied the necessary processes for extracting information from large time-domain databasesthe systematics. In the process, we have produced significant scientific breakthroughs by developing new methods to probe both the elusive time and spatial variations in astrophysics data sets from the SuperMACHO (Massive Compact Halo Objects) survey, the Lowell Observatory Near-Earth Object Search (LONEOS), and the Taiwanese American Occultation Survey (TAOS).

This project continues to contribute to the development of the scientific foundations for future wide-field, time-domain surveys. Our algorithm and pipeline development has provided the building blocks for the development of the LSST science software system. Our database design and performance measures have helped to size and constrain LSST database design. LLNL made significant contributions to the foundations of the LSST, which has applications for large-scale imaging and data-mining activities at LLNL. These developments are being actively applied to the previously mentioned surveys producing important scientific results that have been released to the scientific community and more continue to be published and referenced, enhancing LLNL's scientific stature.


### Introduction/Scientific Motivation

The astronomical community sets priorities for large projects primarily through reviews conducted by the National Academies of Science (NAS). In the past two years three NAS studies have strongly endorsed a wide-field, large-aperture ground-based telescope capable of mapping the dark matter distribution in the Universe, detecting all asteroids capable of catastrophic impact with the Earth and probing the time-varying sky in depth for the first time. The project that has emerged from these NAS recommendations is the Large Synoptic Survey Telescope (LSST), an 8.4 m telescope with a 3+ degree field of view.

Such a facility has unprecedented capabilities. The LSST will provide continual monitoring of the entire accessible sky, repeated as often as every two to three nights. It will uniquely probe new aspects of our Universe, cutting across scientific discipline boundaries. Over the wide range of scientific disciplines to which these data are applicable (see below) there is no current or proposed facility that provides a comparable resource.

Boroson et al. (2000), Tyson (2000) and the NAS reports identify some key LSST scientific goals:

1 An unprecedented digital map of the sky to extreme depths allowing an exquisite determination of the distribution of the dark matter in the Universe by detecting and

quantifying weakly lensed objects. This map also provides an invaluable resource to support observations at all wavelengths.

2 The discovery of thousands of supernovae that will trace departures from the smooth Hubble flow at small redshifts and help measure cosmological parameters at large redshifts;
3 Photometric monitoring of 100 million stars to detect extrasolar planets by occultations and microlensing;
4 The discovery and orbit determination of 90% of Earth-crossing asteroids above a size of 300 meters over a period of ten years;
5 The exploration of rare, perhaps unknown, objects in color and temporal space;

The huge leap in capability represented by LSST exacts a price. The LSST will require a corresponding advance in scientific approaches to deal with the immense data and new technology that will result from completing the LSST. This ER project uses wide-field, time-domain surveys as precursor experiments and data sets to develop approaches for managing and mining LSST data while extracting important science from these precursors.

This project supports the national security mission by developing the LSST telescope and software for solving a variety of technical problems in imaging and data mining that are directly applicable to surveillance for nonproliferation. It also supports the DOE goal of understanding the nature and distribution of dark matter and the nature and evolution of dark energy.

LSST Project Status:

LLNL scientists and engineers have provided critical guidance and expertise during the early phases of this project. Due to early effort in the conceptual stages of the project, and to recognized Laboratory expertise, LLNL is poised to play a major role in both the construction and the science return from the LSST.

The LSST is expected to cost approximately $350M to build, with the funding to come approximately $150M from the DOE Office of Science, $150M from the NSF and $50M from other sources including large private contributions. A $15M proposal was accepted by the NSF for the design and development phase of the project. The Research Corporation has made a significant upfront investment in these early phases as well. DOE Office of Science expects to fund the LSST starting in FY '07.

The three years of this project were crucial in establishing the key scientific requirements, and in identifying the corresponding scientific advances, that will produce the core LSST science. The Laboratory is widely recognized, as evidenced by our prominent role within the project, as a key partner in producing these advances.

A major focus of our development activities has been the LSST precursor project: the SuperMACHO (SM) survey for microlensing toward the Large Magellanic Cloud (LMC). The 500 Gb dataset through 2005 has been transferred and stored at LLNL. Pipeline processing, generation of object photometry catalogs, and difference image analysis is currently underway of all 5 yrs of data (1.5 Tb). In conjunction with the ICE-LSST SI, we worked on pipeline and algorithm development and have released a new version of our fast photometry code—Pixie. We have serendipitously discovered new light echoes, which serve as historical markers of supernovae, novae, or AGB stars. Director's Discretionary time using the 8m Gemini telescope has revealed one set echoes to be from a type Ia supernova event over 500 yrs ago. Spectroscopic follow-up was a large part of the SM program this year (granted 18 nights at the 6m Magellan telescope). The follow-up spectroscopy provided a substantial filter to remove the significant number of supernova contaminants and added the bonus of secondary science; a few new active galactic nuclei behind the LMC have been clearly identified as well as a set of Be variables and 2 possible cataclysmic variables. We plan similar observations with follow-up spectroscopy for next year. Recently awarded HST time for the project will play a significant part in confirming that the detected flux excursions arise from LMC stars, and providing measurements for determining the microlensing optical depth towards the LMC.

We have been active in 'stocking' the LSST collaboration's precursor data set, where we have had to bring our data sets into conformance with LSST standard headers and generate a world coordinate system for each image. Part of the LONEOS imaging data was prepared and packaged for uploading to NCSA and the MACHO data has been brought into conformance and has been uploaded. All of these precursor data sets are contributing to the design and testing of a prototype LSST database.

## Research Activities, Results, Technical Outcome

We investigated the RR Lyrae distribution in the halo of the Milky Way. We mined the Lowell Observatory Near Earth Object Survey data for RR Lyrae variable stars to probe the structure of the Milky Way halo, discovering two distinct RR Lyrae populations that likely trace two distinct production mechanisms responsible for the Milky Way's halo (Miceli et al, paper in preparation).

Obtained the LONEOS2 data and began long duration reductions to identify RR Lyrae stars, extended our analysis of the Milky Way halo (Huber et al., 2004).

Conducted a systematic study of the effect of the atmosphere on the apparent shapes of galaxies to better estimate weak lensing and the effects of the telescope on these measurements. Began initial work on a weak-lensing simulator (Seppala et al., 2004; Schlaufman 2004).

Finished bringing MACHO image data to LLNL to test the SuperMACHO pipelines and extract new science using difference image analysis, long process of reduction ongoing.

Designed a database schema for the SuperMACHO data as an LSST prototype, using it to extract new scientific information and investigate scaling and use scenarios (Nikolaev et al., 2004).

Loaded SuperMACHO data into the prototype LSST database; designed the schema for a prototype LSST database and developed typical queries (Becla et al., 2005; Nikolaev et al., 2006).

Refined the SuperMACHO reduction pipeline, using it to prototype an LSST pipeline system. Used SuperMACHO data to prototype high-throughput pipeline analysis systems for efficiently producing real-time alerts; finished the development of the LSST prototype pipeline and run it on SuperMACHO (Abdulla et al., 2005a; Abdulla et al., 2005b).

Developed a fast photometry code that made significant advances in optimizing image subtraction in order to report real-time detection of microlensing events in the SuperMACHO project. Incorporated our new photometry code into the SuperMACHO pipeline; finished data collection for the SuperMACHO survey and published first paper on setup (Rest et al., 2005a).

Continued to detect microlensing in real time in the SuperMACHO survey and won Hubble Space Telescope time to support SuperMACHO microlensing and follow-up of serendipitous discoveries such as light echo systems.

Continued to develop data-mining tools and expertise using MACHO, SuperMACHO, and LONEOS data. Developed new algorithms for finding unusual objects and tested new algorithms for characterizing light curves (Rest et al, 2005b, Garg et al, 2006 and 2007).

Used the data and tools acquired to probe the variable universe to better understand the physics of variable stars, the mass distribution of the Galaxy, and the local distance scale. Summarized discoveries from the SuperMACHO project and presented secondary science as a wide-field, time-domain survey of the LMC. In particular, using the LSST precursor database, extracted over 1000, newly discovered Delta Scuti variables in the Large Magellanic Cloud

(Cook, et al., 2005; Huber et al., 2005; Huber et al., 2006).

Discovered three new supernova light echoes in the Large Magellanic Cloud and published in *Nature* (Rest et al., 2005b).

Continued investigation of light echo systems in the LMC and elsewhere in the galaxy resulted in the discovery of another unique evolved star similar to V838 Mon.

Presented and published a paper on the rise time of type Ia supernovae, which constrains supernova explosion models (Garg et al., 2006; Garg et al., 2007).

Investigated unusual B-type emission-line stars and active galactic nuclei and other variable stars discovered with SuperMACHO (Huber et al., 2006, Huber et al, in preparation).

Identified and analyzed new sets of eclipsing binaries (Faccioli et al, in preparation), R Coronae Borealis stars (Zaniewski, A. et al., 2005), and long-period variables (Fraser et al., 2005) from the MACHO dataset.

Follow-up of a microlensing survey yielded the stellar shape of the microlens in microlensing event MOA 2002-BLG-33 (Rattenbury, et al., 2005).

Determined the large-scale extinction map of the Galactic Bulge by mining data from the MACHO Project photometry (Popowski, et al., 2003). Determined the microlensing optical depth towards the Galactic Bulge using clump giants mined from the MACHO Survey (Popowski, et al., 2005). Generated a catalog of microlensing events toward the Galactic Bulge mined from the MACHO Survey (Thomas et al., 2005).

Extracted a new catalog of type RR0 RR Lyrae in the Galactic Bulge from the MACHO survey (Kunder, et al., 2004).

Merged time-domain and multi-wavelength datasets (MACHO+2MASS) and mined Cepheid variables to investigate the geometry of the LMC disk (Nikolaev, et al. 2004a).

Using Cepheid variables mined from the MACHO data determined more empirical evidence for non-linearity of the period-luminosity relations as seen in the LMC Cepheids (Ngeow, et al., 2005).

Follow-up spectroscopy using RR Lyrae stars mined from the MACHO data as velocity probes, we found kinematic evidence for an old stellar halo in the Large Magellanic Cloud (Minniti, et al., 2003).

Using RR Lyrae mined from the MACHO Survey in the LMC, we made a frequency analysis of type RR0 and measured Fourier parameter of the RR1 to better understand the stellar pulsation physics and also to determine a distance measure to the LMC (Alcock, et al., 2003; Alcock et al., 2004). We also used these to determine additional properties of RR Lyrae and the structure of the Large Magellanic Cloud in its inner regions (Borissova, et al., 2004).

Using distant quasi-stellar objects mined from the MACHO Survey with unique mining tools, we obtained Hubble Space Telescope time and determined the proper motion of the LMC (Kallivayalil, et al., 2006).

Using unique data mining techniques and algorithms extracted a new catalog of eclipsing binaries from the MACHO data in the Large and Small Magellanic Clouds. Using this catalog and the MACHO data analyzed the circularization of eccentric orbits as a probe of stellar structure (Faccioli et al, two papers in preparation).

We participated in the follow-up of microlensing survey discoveries as part of the PLANET Collaboration resulting in the discovery of a cool sub-Neptune mass planet OGLE 2005-390Lb (Beaulieu, et al., 2006).

We developed new data mining tools to discovery new eruptive cataclysmic variables using the MACHO Database (Cieslinski, et al., 2004).

We used Hubble Space Telescope follow-up to determine the nature of the LMC microlensing event LMC-5, thus making an unambiguous measurement of the mass and distance of an M dwarf star which was too distant for parallax measurements and had no companion to help determine its mass (Drake, et al., 2004).

We initiated development of tools for accessing large datasets through the Virtual Observatory interface (Drake et al., 2006; Huber et al., 2007).

We continued as members of the Taiwanese-American Occultation Survey. Our collaboration continued data collection and developed a new class of techniques to look at wide-field, sub-second time-variable phenomena (Chen et al., 2006).

## Exit Plan

The LSST project as a whole maps nicely onto the LLNL S&T plan. There will be world-class science coming out of the precursor experiments and the LSST when it is finally inaugurated. The optical design and fabrication of the telescope are technically challenging forays into the collection and detection of radiation. The data analysis management and data analysis system for the LSST is a world-class challenge in data exploitation. This particular project exploited precursor surveys to generate science and insight into the design requirements for the eventual data management and exploitation system for the LSST.

We expect DOE Office of Science money to flow from HEP in about one year to support the science activities surrounding the LSST search for dark energy using Type Ia supernovae. We also expect DOE/LSST money to come to the lab to support the engineering efforts surrounding the design and construction of the telescope and camera. Using MACHO as a model, it is also likely that NASA money will come to the lab to support space-based follow-up of discoveries made in the precursor experiments and eventually the LSST.

## Summary

This ERD has been focusing on developing the foundations for extracting knowledge from PetaByte data sets so that we will be ready for the data torrent from the Large Synoptic Survey Telescope (LSST) which is a key element in LLNL's Nuclear, Radiative and Astrophysics S&T plan. Our approach is to utilize our unique expertise in, and access to, existing large data sets. This ERD is using wide-field, time-domain surveys as precursor projects to drive the needed development. In the process, we have produced significant near-term science, primarily mining the MACHO and SuperMACHO databases (19 refereed journal articles and 17 abstracts, reports and conference proceedings).

## Acknowledgements

We wish to acknowledge our principal collaborators: Ghalleb Abdulla (LLNL), Andy Becker (UW), Jacek Becla (SLAC), Andrew Drake (CIT), Lorenzo Faccioli (LBL), Oliver Fraser (UW), Arti Garg (Harvard), Stuart Marshall (SLAC), Chris Smith (CTIO), Christopher Stubbs (Harvard) and Doug Welch (McMaster U.)

## References

Abdulla, G., et al., 2005a. "Middleware for Astronomical Data Analysis Pipelines." Astron. Data Analysis Software Systems, Pasadena, CA, United States, Oct 24-Oct 27, 2005.UCRL-CONF-209265.

Abdulla, G. M., et al., 2005b. "LSST data pipeline prototyping plans and strategy." Technical

Report. UCRL-TR-204392.

Alcock, C., et al., 2003. "The Macho Project Large Magellanic Cloud Variable Star Inventory. XI. Frequency Analysis of the Fundamental Mode RR Lyrae Stars." Astrophys. J., 598, 597. UCRL-JC-153183-PT-11.

Alcock, C., et al., 2004. "The MACHO Project Large Magellanic Cloud Variable Star Inventory. XIII. Fourier Parameters for the First Overtone RR Lyrae Variables and the LMC Distance." Astronom. J., 127, 334. UCRL-JRNL-201664.

Beaulieu, J. P., et al., 2006. "Discovery of the distant cool sub-Neptune mass planet OGLE 2005-BLG-390Lb by microlensing." Nature, 439, 437. UCRL-JRNL-216921.

Becla, J., et al., 2005. "LSST Data Access Overview." Technical Report. UCRL-ABS-217857.

Borissova, J., et al., 2004. "Properties of RR Lyrae stars in the Inner Regions of the Large Magellanic Cloud." Astron. & Astroph., 423, 97. UCRL-JRNL-205052.

Cieslinski, D., et al., 2004. "Discovery of New Eruptive Cataclysmic Variables using the MACHO Database." Publ. Astron. Soc. Pacific, 116, 610. UCRL-JRNL-205040.

Chen, W., et al., 2006. "Search for Small Trans-Neptunian Objects by the TAOS Project." Near Earth Objects, our Celestial Neighbors: Opportunity and Risk, Prague, Czech Republic, Aug 14-Aug 23, 2006. UCRL-PROC-225991.

Cook, K. H., et al., 2004. "MACHO 97-BLG-56 and MACHO 99 BLG-8: Candidate Microlensed Stars in the Sgr Dwarf Galaxy." UCRL-ABS-201976.

Cook, K., et al., 2005. "The SuperMacho Project: Wide-field, Time-domain Survey of the Large Magellanic Cloud." Stellar Pulsation and Evolution, Monte Porizo,Italy, Jun 19-Jun 24, 2005. UCRL-PROC-223125.

Drake, A. J., et al., 2004. "Resolving the Nature of the LMC Microlensing Event LMC-5." Astrophys. J. Letters, 607, 29. UCRL-JRNL-203738.

Drake, A., et al., 2006. "Creating a VOTimeSeries." 207th American Astronomical Society Mtg., Washington, D.C., Jan. 8-12, 2006. UCRL-POST-218729.

Fraser, O. J., et al., 2005. "Long Period Variables in the LMC: Results from MACHO and 2Mass." Astron. J., 129, 768. UCRL-JRNL-205457.

Garg, A., et al., 2006. *Pre-maximum lightcurves of type ia supernovae from the superMACHO project*. 207th American Astronomical Society Mtg., Washington, DC, Jan. 8–12, 2006. UCRL-POST-218290.

Garg, A., et al., 2007. "Light Curves of Type Ia Supernovae from Near the Time of Explosion." Astronom. J. 133, 403. UCRL-JRNL-224121.

Huber, M. E., et al., 2004. "Utilizing Variability Signatures for Tracing Galactic Populations in Wide-Field Imaging Surveys." LLNL Postdoc Symposium, Dec. 13, 2004. UCRL-POST-208430.

Huber, M. E., et al., 2005. "Bleeding/Cutting Edge Secondary Science from SuperMACHO." LLNL Postdoc Symposium, Dec. 12, 2005. UCRL-ABS-217719, UCRL-PRES-217651.

Huber, M., et al., 2006. "Mining the Variables Out of the SuperMACHO Dataset." 207th American Astronomical Society Mtg., Washington, D.C., Jan. 8-12, 2006. UCRL-POST-217927.

Huber, M., et al., 2007. "Time-Series Data and the Virtual Observatory: The Faint Sky Variability Survey Dataset." 209th American Astronomical Society Mtg., Seattle, WA, Jan. 6-10, 2007. UCRL-POST-227082.

Kallivayalil, N., et al., 2006. "The Proper Motion of the Large Magellanic Cloud using HST." Astrophys. J., 638, 772. UCRL-JRNL-214724.

Kunder, A. M., et al., 2004, "RR Lyrae RR0 Stars in the Bulge." LLNL Summer Student Symposium, August, 2004. UCRL-POST-205824.

Minniti, D., et al.,2003. "Kinematic Evidence for an Old Stellar Halo in the Large Magellanic Cloud." Science, 301, 1508. UCRL-JC-155452.

Ngeow, C.-C., et al., 2005. "Further empirical evidence for the non-linearity of the period-luminosity relations as seen in the Large Magellanic Cloud Cepheids." Mon. Not. Royal Astron. Soc.,363, 831. UCRL-JRNL-213034.

Nikolaev, S. et al. 2004a. "Geometry of the LMC Disk: Results from MACHO and 2MASS." Astrophys. J., 601, 260. UCRL-JRNL-200780.

Nikolaev, S., et al., 2004b. "Conceptual Design of a Prototype LSST Database." Technical Report. UCRL-TR-207256.

Nikolaev, S., et al., 2006. "Evaluation of Potential LSST Spatial Indexing Strategies." Technical Report. UCRL-TR-225827.

Popowski, P., et al., 2003. "The Large-Scale Extinction Map of the Galactic Bulge from the MACHO Project Photometry." Astronom. J.126, 2901, UCRL-JC-152345.

Popowski, P., et al., 2005. "Microlensing Optical Depth towards the Galactic Bulge Using Clump Giants from the MACHO Survey." Astrophys. J., 631, 875. UCRL-JRNL-213625.

Rattenbury, N. J., et al., 2005. "Determination of stellar shape in microlensing event MOA 2002-BLG-33." Astron. & Astrophys., 439, 645. UCRL-JRNL-214971.

Rest, A., et al., 2005a. "Testing LMC Microlensing Scenarios: The Discrimination Power of the SuperMACHO Microlensing Survey." Astrophys. J., 634, 1103. UCRL-JRNL-204389.

Rest, A. et al., 2005b. "Light Echoes from Ancient Supernovae in the Large Magellanic Cloud." Nature, 438, 1132. UCRL-JRNL-212979,UCRL-JRNL-216557.

Schlaufman, K., 2004. "Simulations of Turbulence Induced Ellipticity Over Large Fields-of-View: the First Step Towards Enabling LSST Weak Lensing Science." Technical Report. UCRL-TR-207730.

Seppala, L., et al., 2004. "Optical Design for the 8.4m Large Synoptic Survey Telescope." 205th American Astronomical Society Mtg., San Diego, CA, Jan. 9-13, 2005. UCRL-ABS-207882.

Thomas, C. L., et al., 2005. "Galactic Bulge Microlensing Events from the MACHO Collaboration." Astrophys. J., 631, 906. UCRL-JRNL-213632, UCRL-JRNL-213064.

Zaniewski, A. et al., 2005. "Discovery of Five New R. Coronae Borealis Stars in the MACHO Galactic Bulge Database." *Astronom. J.*, **130**, 2293. UCRL-JRNL-213036.