

The Palomar Transient Factory

Peter Nugent^{a,b}, Yi Cao^c and Mansi Kasliwal^{c,d}

^a Computational Cosmology Center, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA; ^b Department of Astronomy, University of California, Berkeley, CA 94720-3411; ^c Astronomy Department, California Institute of Technology, MC 249-17, 1200 East California Boulevard, Pasadena, CA 91125, USA; ^d The Observatories, Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA

ABSTRACT

Astrophysics is transforming from a data-starved to a data-swamped discipline, fundamentally changing the nature of scientific inquiry and discovery. New technologies are enabling the detection, transmission, and storage of data of hitherto unimaginable quantity and quality across the electromagnetic, gravity and particle spectra. The observational data obtained during this decade alone will supersede everything accumulated over the preceding four thousand years of astronomy. Currently there are 4 large-scale photometric and spectroscopic surveys underway, each generating and/or utilizing hundreds of terabytes of data per year. Some will focus on the static universe while others will greatly expand our knowledge of transient phenomena. Maximizing the science from these programs requires integrating the processing pipeline with high-performance computing resources. These are coupled to large astrophysics databases while making use of machine learning algorithms with near real-time turnaround. Here we present an overview of one of these programs, the Palomar Transient Factory (PTF). We will cover the processing and discovery pipeline we developed at LBNL and NERSC for it and several of the great discoveries made during the 4 years of observations with PTF.

Keywords: Image Processing, Transient Detection, Machine Learning, Workflows

1. INTRODUCTION

The majority of synoptic optical surveys are tuned to maximize discoveries of selected source populations (typically microlensing, classical novae, or supernovae). While such surveys are crucial for specialized science, gaping holes in the time-domain phase space remain. The Palomar Transient Factory (PTF) is a next-generation transient survey which has systematically explored the variable sky on a variety of timescales. PTF has, simultaneously, *discovered* well-studied populations (e.g., classical novae, supernovae), poorly constrained events (e.g., luminous red novae, tidal disruption flares), and predicted, but here-to-fore not yet observed, phenomena (e.g., orphan afterglows of γ -ray bursts, supernova precursor explosions, etc.). Figure 1 summarizes the cornucopia of discoveries by PTF - all of which were enabled by tightly coupling HPC resources in the workflow pipeline for PTF. The goal of our research is to systematically *follow-up* these discoveries with photometric and spectroscopic observations *as close to real-time as possible*. The pressing need for this is clear: photometric light curves provide clues to the physics of the events while spectroscopy are significantly more definitive in determining both the nature of the event and its distance.

PTF is a comprehensive transient detection system including a wide-field survey camera, an automated realtime data reduction pipeline, a dedicated photometric follow up telescope, and a full archive of all detected sources. The survey camera achieved first light on 13 Dec 2008; it completed commissioning in 1 Mar 2009; and will finish its original survey 31 Dec 2012 with continued operations until the 2016 timeframe.^{1,2}

The transient detection survey component of PTF is performed at the automated Palomar Samuel Oschin 48-inch telescope (P48); candidate transients are photometrically followed up at the automated Palomar 60-inch

Further author information: (Send correspondence to PEN)
PEN: E-mail: penugent@LBL.gov Telephone: 1 510 486 6942

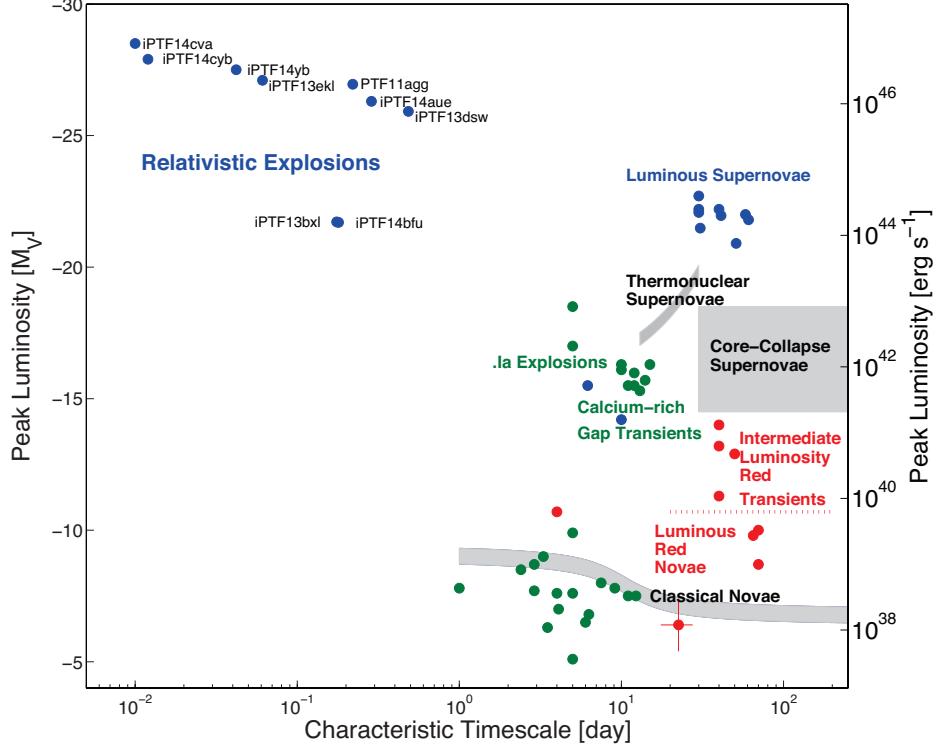


Figure 1. The phase space of optical transients.⁴ The x-axis is the characteristic time and the y-axis is the peak luminosity. The grey shaded areas highlight the locations of transients known prior to PTF. Note the many new classes of explosions, in particular the relativistic ones which have timescales of minutes.³

telescope (P60). This dual-telescope approach allows both high survey throughput and a very flexible followup program. PTF fields are scheduled in two main modes, a 3-5 day cadence survey and a dynamic cadence survey.

The 3-5 day cadence experiment runs yearly from March 1st until October 30th using 65% of the time in that period. The main goal of this experiment is to construct large samples of SNe Ia and core-collapse SNe. The experiment observes a footprint of approximately 8000 square degrees, with $|b| > 30^\circ$, ecliptic latitude $|\beta| > 10^\circ$, and with a median cadence of 4 days. At a given time, the active area for the search (including all overheads) is 2700 square degrees. The observations are conducted in almost all lunar phases in R -band. In each epoch we typically obtain two 60 s exposures separated by 60 min. The two images are used to remove cosmetic artifacts, cosmic rays and to find solar system objects. The dynamic cadence experiment (DyC) is designed to explore transient phenomena on time scales shorter than ~ 3 days and longer than one minute and to catch very young, nearby supernovae. The dynamic cadence is an evolving experiment. Every six months the PTF collaboration reviewed the results of the dynamic cadence experiment and tweaked sky, coverage, nightly cadence, etc., to optimize the survey.

PTF science programs are undertaken in the context of projects that are proposed to and recognized by the PTF Consortium. Individuals or groups affiliated with one or more PTF Consortium member institutions have the right to propose projects to be undertaken using PTF, and member institutions may endorse projects advocated by their personnel. Major projects are: Type Ia Supernova Project, Transients in the Local Universe, Core-Collapse Supernova Project, Hostless Transients, RR Lyrae, TDF's and AGN's among others. Once a target has been confirmed spectroscopically or photometrically, all collaboration data are turned over to the given science group which then decides the course of action for additional follow-up and the resultant publications.

2. OPERATIONS

Data taken with the camera are transferred to two automated reduction pipelines (see Figure 2). A near-realtime image subtraction pipeline is run at NERSC/LBNL and has achieved the goal of identifying optical transients within minutes of images being taken. The output of this pipeline is sent to UC Berkeley where a source classifier determines a set of probabilistic statements about the scientific classification of the transients based on all available time-series and context data.⁵

On few-day timescales the images are also ingested into a database at the Infrared Processing and Analysis Center (IPAC). Each incoming frame is calibrated and searched for objects, before the detections are merged into a database which is comprehensive, made public after an 18 month proprietary period and can be queried for all detections in an area of the sky.

Followup of detected transients is a vital component of successful transient surveys. The P60 photometric followup telescope automatically generates colors and light curves for interesting transients detected using P48. The PTF collaboration also leverages a further 15 telescopes for photometric and spectroscopic followup. An automated system collates detections from the Berkeley classification engine, and makes them available to the various follow up facilities, coordinates the observations, and reports on the results.

From the perspective of NERSC/LBNL operations are centered around three distinct areas: the running of the near real-time subtraction pipeline, the archiving of this data and the management and organization for the PTF SN Ia project and its various follow-up programs. The maintenance of the telescopes (P48 & P60) and their respective instruments for PTF are handled by Caltech and the PTF Consortium in addition to the IPAC processing pipeline.

2.1 Technical

The operations of the Palomar Transient Factory can be summed up as follows: find the transients, screen the transients, follow-up them up with additional resources and publish the research.

The survey operations are performed robotically by the P48 Observatory Control System (OCS), written in MATLAB. The OCS is responsible for controlling the camera, filter exchanger, shutter, telescope, dome, and focuser, on the basis of feedback information from the scheduling system, the camera and telescope, a weather station, and the data quality monitor.

The OCS is responsible for sequencing all PTF observations, starting with the bias frames before sunset, through focus images, and finally the science images. The PTF scheduler is responsible for selecting the next target for observation. The PTF schedule is not pre-defined; the next target is selected 30-90 s before each exposure starts, during the previous exposure.

Science images are transferred more than 400 miles via high-speed networks including the National Science Foundations High Performance Wireless Research and Education Network (HPWREN) and the DOE's Energy Sciences Network (ESnet) to the National Energy Research Scientific Computing Center (NERSC), located at Berkeley Lab.

Data are processed and analyzed at NERSC and follow-up observations are triggered first at the P60 (of which 50% of the time is allocated to the PTF Consortium) and then on a variety of telescopes run and maintained by national (including the University of California and NSF) and international facilities. This time is provided for free through competitive allocations at these facilities, however, in actual dollars, they represent the largest resource used by the PTF - and thus the most valuable. For example, the PTF Type Ia Supernova (SN Ia) project has been awarded over 400 nights of observing time since 2009 on everything from 1-meter (\sim 50 nights per year) to 8 & 10-meter class telescopes and the Hubble Space Telescope.

2.2 Data

At NERSC the Real-time Transient Detection Pipeline makes use of the IBM iDataPlex supercomputer *Carver*, a high-speed parallel filesystem and sophisticated machine learning algorithms to sift the data and identify events for scientists to follow up on. The pipeline for PTF was developed in 2008 as part of a NERSC-based program called DeepSky.⁶ This pilot program was the first experimentally-based data intensive project at NERSC which

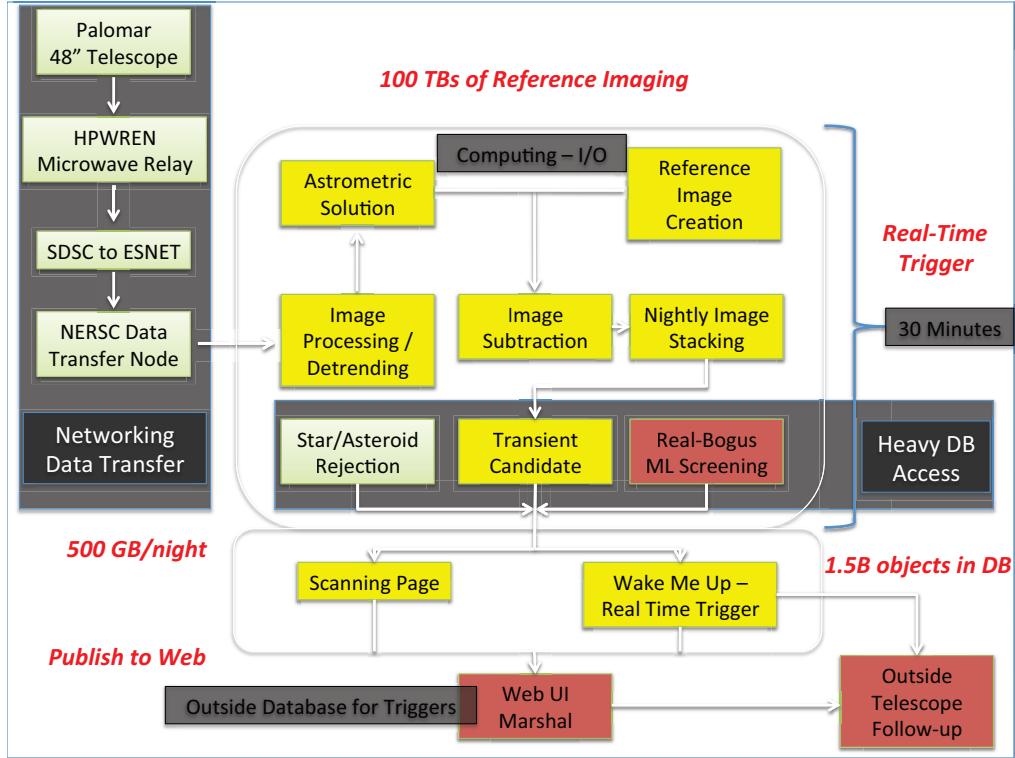


Figure 2. About 100 GB of raw optical imaging data is taken every night at Palomar Observatory, which is subtracted from existing reference images from the same part of the sky to look for new astrophysical transients, including supernovae, variable stars and other cataclysmic explosions. This data is processed at NERSC and produces nearly 500GB of subsequent new, reference and subtraction imaging data after detrending and alignment. These are scanned using machine learning algorithms to find the needle in the haystack, about 1 in 1M objects recorded is of interest. A nearly 1TB database containing over 1.5B objects is scoured to compare this event to previous detections at that location on the sky. The results are published to the web in < 40 min after an individual image is taken at Palomar.

made use of the NERSC Global Filesystem, the Tape Archive System, the various supercomputers and Science Gateway Nodes (where databases and access to the data and its products are made available on the web) to manage a pipeline from start to finish.

NERSC is the primary scientific computing facility for the Office of Science in the DOE. All research projects that are funded by the DOE Office of Science and require high performance computing support are eligible to apply to use NERSC resources. PTF has been awarded project status at NERSC for the past 4 years and given the success of the program and its ties to DESI there is every expectation that this program will be renewed in the years to come. These resources, which are maintained 24X7 by NERSC, are provided at no cost to the user unless they desire special services.

PTF currently uses an allocation of 150 khrs per year of time on *Carver* to process incoming data, build references, perform subtractions, run the classification codes and share the data over the web to the collaboration. NERSC provides, for free, up to 10TB of spinning disk space on their Global Filesystem (NGF). For PTF, the only special service required is to keep all the images (new's, references and subtractions) spinning on NGF. The first 100TB of this space were purchased through the NERSC pilot program and an additional 100TB were acquired in 2010 as part of LBNL's buy-in to PTF. Additional space will be acquired as needed.

The PTF real-time reduction and subtraction pipeline was designed from the start to take full advantage of the parallel high performance computing facilities at NERSC (<http://www.nersc.gov>). The major improvement to the hardware used in this pipeline over previous versions designed by the Supernova Cosmology Project⁷ and Nearby Supernovae Factory⁸ at LBNL is the use of the NGF. NGF is a 2PB shared filesystem with over

15 GB/sec bandwidth for streaming I/O which can be seen by all of the high performance computers. On the software front, the improvements include a tight coupling of a PostgreSQL database involved in tracking every facet of the image processing, reference building, image subtraction and candidate detection along with a complete re-write of the processing and subtraction codes, described below. This database now contains over 1 billion candidate sources.

Incoming P48 images are immediately backed up on the High Performance Storage System, a 60PB tape archive system at NERSC. Each night the P48 pushes between 50-100GB of data to NERSC depending on the length of the night. Processing begins by splitting the packed images apart by chip, applying crosstalk corrections, and performing standard bias/overscan subtraction and flat fielding. After this point, all operations are performed on a chip-by-chip basis in parallel. Catalogs of sources are created for each image via the Astromatic SExtractor code (<http://www.astromatic.net>), which then is fed to the astrometry.net (<http://astrometry.net>) code to perform an astrometric solution. At this point a comparison to the USNO catalog is made to determine the zero-point and 3σ limiting magnitude of the image along with a calculation of the seeing. The image is then loaded into the processed image database, storing all relevant information from the FITS headers.

After several images are obtained for a given PTF pointing, we are able to create a reference image. The reference images are created via the Astromatic software Scamp and Swarp after querying the processed image database in order to obtain the highest quality input images for a given pointing/chip combination. The references are assigned version numbers and (along with their relevant information) are stored in the reference database.

Subsequent new images are processed as above, and if a reference image exists for them, a subtraction is performed. First, the new image is astrometrically aligned to the reference by utilizing Scamp and the new and reference catalogs. The reference is then Swarped to the size and scale of the new image and a subtraction is performed using HOTPAnTS (<http://www.astro.washington.edu/users/becker/hotpants.html>). The subtraction is performed in both the positive and negative direction (reference minus new and new minus reference) to detect both positive and negative flux changes. Candidate transient sources are detected via SExtractor and along with other relevant parameters (location with-respect-to bad pixels, etc.) are stored in a candidate and subtraction database.

In addition, PTF takes advantage of DeepSky (<http://www.deepskyproject.org>). DeepSky was started in response to the needs of several astrophysics projects hosted at NERSC. It is a repository of digital images taken with the point-and-stare observations by the Palomar-QUEST Consortium and the Near Earth Asteroid Team. This data spans nine years and more than 15,000 square degrees, with 20-200 pointings on a particular part of the sky and cadences from minutes to years. For a large fraction of the survey DeepSky achieves depths of $m_R > 23$ magnitude. In total there are more than 11 million images in DeepSky. This historical data set compliments the Palomar Transient Factory survey by identifying known variable stars, AGN, and the detection of low surface brightness host galaxies of supernovae.

2.3 Classification & Follow-up

The Transients Classification Pipeline (TCP) is a parallelized, Python-based framework created to identify and classify transient sources in the realtime PTF differencing pipeline. The TCP polls the candidate database from that pipeline and retrieves all available metadata about recently extracted sources. Using the locations and uncertainties of the transient candidate objects, the TCP either associates an object with existing known sources in the TCP database or, after passing several filters to exclude non-stellar (e.g. known minor planets) and non-astrophysical events, generates a new source.

Once a transient source has been identified, the TCP generates “features” which map contextual and time-domain properties to a large dimensional real-number space. After generating a set of features for a transient source candidate, the TCP then applies several science classification tools to determine the most likely science class of that source. For rapid-response transient science, a subset of features — such as those related to rise times and the distance to nearby galaxies — are most useful. As light curves are better sampled and colors are obtained more features are used in the classification. The resulting science class probabilities are stored in a database for further data mining applications. An example of this classification process can be see in Figure 3.

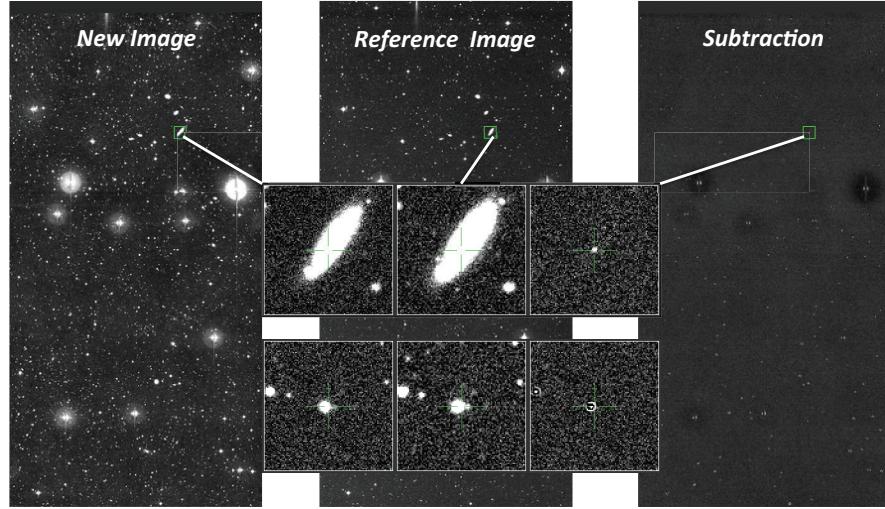


Figure 3. A triplet of a new image, a reference image (taken months earlier of the same part of the sky) and a subtraction. Inset are the discovery of a supernova in a nearby galaxy and one of the false-positive detections for this image. On this relatively *clean* subtraction there were over 250 candidate transients, of which only 10 were astrophysical in nature. 2 asteroids, 7 variable stars and a SN Ia caught within a day of explosion. The rest were image artifacts that were culled by the machine learning codes used to separate real from bogus candidates.

Sources with high probabilities of belonging to a science class of interest to the PTF group are broadcast to the PTF's Follow-up Marshal (hosted on the Science Gateway Nodes at NERSC) for scheduling of follow-up observations. The Marshal maintains all the information obtained by the PTF Consortium on every candidate object of interest it has tracked. As of 2014 this included over 2300 supernovae and their accompanying lightcurves from the P48 and P60 telescopes and over 4000 spectra obtained through other follow-up facilities (see <http://www.ptf.caltech.edu/iptf>).

After publication, data on PTF SNe Ia is released to the public through the Weizmann Interactive Supernova data REPOSITORY - WISEREP - an SQL-based database with an interactive web-based graphical interface (<http://www.weizmann.ac.il/astrophysics/wiserep>). The system serves as an archive of high quality SN spectra, including both historical (legacy) data as well as data that is accumulated by ongoing modern programs such as PTF.⁹ The archive provides information about objects, their spectra, and related meta-data. Utilizing interactive plots, they provide a graphical interface to visualize data, perform line identification of the major relevant species, determine object redshifts, classify SNe and measure expansion velocities. Guest users may view and download spectra or other data that have been placed in the public domain. The Weizmann Institute maintains this database. IPAC is in charge of releasing the PTF data acquired from the P60 and P48 telescopes. Currently, once the data has gone through the final processing by IPAC, the data are released to the public through a web-interface after an 18 month proprietary period.

3. CONCLUSIONS

In 2016, PTF will cease operations and make way for the Zwicky Transient Facility (ZTF) collaboration with a new camera being built at Caltech's Palomar Observatory that will be able to survey the entire sky in the

Northern Hemisphere each night, searching for supernovas, black holes, near-Earth asteroids, and other objects. The ZTF camera's field of view will encompass 47 square degrees, larger than 200 full moons. By contrast, the field of view of the Hubble Space Telescope is so small that a mosaic of 130 of its images of the moon would be needed to see it in its entirety. ZTF will be able to go wide and fast and cover most of the observable sky. It will shoot one frame every 30 seconds at 18 gigabits per frame. ZTF will be able to find a supernova less than 24 hours since its explosion *every* single night. This quick response is critical, as the light emitted in the first few hours after a supernova explodes contains a wealth of information that cannot be retrieved later. As PTF has demonstrated, this leads to incredible new scientific discoveries. The order of magnitude increase in data from ZTF (a factor of 2 in speed and 5 in volume) means that computational challenges will increase as well. By coupling the astronomical pipelines with ever expanding HPC resources, we are well on our way to delivering this capability for ZTF.

REFERENCES

- [1] Rau, A., Kulkarni, S. R., Law, N. M., Bloom, J. S., Ciardi, D., Djorgovski, G. S., Fox, D. B., Gal-Yam, A., Grillmair, C. C., Kasliwal, M. M., Nugent, P. E., Ofek, E. O., Quimby, R. M., Reach, W. T., Shara, M., Bildsten, L., Cenko, S. B., Drake, A. J., Filippenko, A. V., Helfand, D. J., Helou, G., Howell, D. A., Poznanski, D., and Sullivan, M., "Exploring the Optical Transient Sky with the Palomar Transient Factory," *PASP* **121**, 1334–1351 (Dec. 2009).
- [2] Law, N. M., Kulkarni, S. R., Dekany, R. G., Ofek, E. O., Quimby, R. M., Nugent, P. E., Surace, J., Grillmair, C. C., Bloom, J. S., Kasliwal, M. M., Bildsten, L., Brown, T., Cenko, S. B., Ciardi, D., Croner, E., Djorgovski, S. G., van Eyken, J., Filippenko, A. V., Fox, D. B., Gal-Yam, A., Hale, D., Hamam, N., Helou, G., Henning, J., Howell, D. A., Jacobsen, J., Laher, R., Mattingly, S., McKenna, D., Pickles, A., Poznanski, D., Rahmer, G., Rau, A., Rosing, W., Shara, M., Smith, R., Starr, D., Sullivan, M., Velur, V., Walters, R., and Zolkower, J., "The Palomar Transient Factory: System Overview, Performance, and First Results," *PASP* **121**, 1395–1408 (Dec. 2009).
- [3] Singer, L. P., Cenko, S. B., Kasliwal, M. M., Perley, D. A., Ofek, E. O., Brown, D. A., Nugent, P. E., Kulkarni, S. R., Corsi, A., Frail, D. A., Bellm, E., Mulchaey, J., Arcavi, I., Barlow, T., Bloom, J. S., Cao, Y., Gehrels, N., Horesh, A., Masci, F. J., McEnery, J., Rau, A., Surace, J. A., and Yaron, O., "Discovery and Redshift of an Optical Afterglow in 71 deg²: iPTF13bxl and GRB 130702A," *Astrophys. J. Letters* **776**, L34 (Oct. 2013).
- [4] Kasliwal, M. M., *Bridging the gap : elusive explosions in the local universe*, PhD thesis, California Institute of Technology (2011).
- [5] Brink, H., Richards, J. W., Poznanski, D., Bloom, J. S., Rice, J., Negahban, S., and Wainwright, M., "Using machine learning for discovery in synoptic survey imaging data," *Mon. Not. Roy. Astron. Soc.* **435**, 1047–1060 (Oct. 2013).
- [6] Nugent, P. E., Transient Factory, P., and Project, D., "DeepSky and PTF: A New Era Begins," in [American Astronomical Society Meeting Abstracts #213], *Bulletin of the American Astronomical Society* **41**, 469.10 (Jan. 2009).
- [7] Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Supernova Cosmology Project, "Measurements of Omega and Lambda from 42 High-Redshift Supernovae," *Astrophys. J.* **517**, 565–586 (June 1999).
- [8] Aldering, G., Adam, G., Antilogus, P., Astier, P., Bacon, R., Bongard, S., Bonnaud, C., Copin, Y., Hardin, D., Henault, F., Howell, D. A., Lemonnier, J.-P., Levy, J.-M., Loken, S. C., Nugent, P. E., Pain, R., Pecontal, A., Pecontal, E., Perlmutter, S., Quimby, R. M., Schahmaneche, K., Smadja, G., and Wood-Vasey, W. M., "Overview of the Nearby Supernova Factory," in [Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series], Tyson, J. A. and Wolff, S., eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **4836**, 61–72 (Dec. 2002).
- [9] Yaron, O. and Gal-Yam, A., "WISEREP - An Interactive Supernova Data Repository," *PASP* **124**, 668–681 (July 2012).