FERMILAB-CONF-20-111-AE-SCD

# The updated DESGW processing pipeline for the third LIGO/VIRGO observing run

*Kenneth* Herner[1,*], *James* Annis[1], *Alyssa* Garcia[2], *Marcelle* Soares-Santos[2], *Dillon* Brout[3,**], *Noemi* Glaeser[1,4], *Nora* Sherman[2], *Richard* Kessler[5], *Robert* Morgan[6], *Antonella* Palmese[1], *Francisco* Paz-Chinchon[7], *Amber* Lenon[8], *Tristan* Bachmann[5], and

[1]Fermi National Accelerator Laboratory, Batavia, IL, USA
[2]Brandeis University, Waltham, MA, USA
[3]University of Pennsylvania, Philadelphia, PA, USA
[4]University of Maryland and Max Planck Institute, College Park, MD, USA
[5]Kavli Institute for Cosmological Physics and The University of Chicago, Chicago, IL, USA
[6]University of Wisconsin-Madison, Madison, WI, USA
[7]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[8]West Virginia University, Morgantown, WV, USA

**Abstract.** The DESGW group seeks to identify electromagnetic counterparts of gravitational wave events seen by the LIGO-VIRGO network, such as those expected from binary neutron star mergers or neutron star- black hole mergers. DESGW was active throughout the first two LIGO observing seasons, following up several binary black hole mergers and the first binary neutron star merger, GW170817. We describe the modifications to the observing strategy generation and image processing pipeline between the second (ending in August 2017) and third (beginning in April 2019) LIGO observing seasons. The modifications include a more robust observing strategy generator, further parallelization of the image reduction software and difference imaging processing pipeline, data transfer streamlining, and a web page listing identified counterpart candidates that updates in real time. Taken together, the additional parallelization steps enable us to identify potential electromagnetic counterparts within fully calibrated search images in less than one hour, compared to the 3-5 hours it would typically take during the first two seasons. These performance improvements are critical to the entire EM followup community, as rapid identification (or rejection) of candidates enables detailed spectroscopic followup by multiple instruments as soon as possible, leading to more information about the environment immediately following such gravitational wave events.

## 1 Introduction

The search for electromagnetic counterparts to gravitational wave (GW) events, such as those expected from binary neutron star or neutron star-black merge mergers, touches a number of topics in physics, including cosmology, the origin of heavy elements, and the neutron

*e-mail: kherner@fnal.gov
**NASA Einstein Fellow

This document was prepared by DESGW collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

star equation of state. Dozens of groups around the world use a variery of instruments to search for potential electromagnetic counters after receiving word of a GW event from the LIGO/Virgo Collaboration (NVC). One such group is DESGW, consisting of members of the Dark Energy Survey (DES), LIGO, and other members of the astronomical community, The DESGW group, active since the first LIGO observing season (01) in 2015, employs the Dark Energy Camera [1] to search for optical counterparts of GW events over multiple triggers in each observing season thus far. Figure 1 presents an outline of the steps the DESGW group takes from LIGO trigger to candidate announcement.

Our search and discovery pipeline [2] uses difference imaging to search for optical counterparts of GW events. The pipeline is in turn based on the Dark Energy Survey's single-epoch [3] and supernova difference imaging [4] pipelines. Ref. [2] primarily focuses on the details of the pipeline during the first two LIGO observing seasons (O1 and O2) from 2015-2017 and continues to form the backbone of the current pipeline. Here, we focus on modifications made between O2 and O3 (beginning in April 2019) in the areas of bringing additional parallelization to single-epoch processing, candidate identification, and public dissemination of candidates. These improvements aim to reduce the overall time from image acquisition to candidate identification, improve the purity of candidate selection, and facilitate rapid dissemination of candidates to the community.

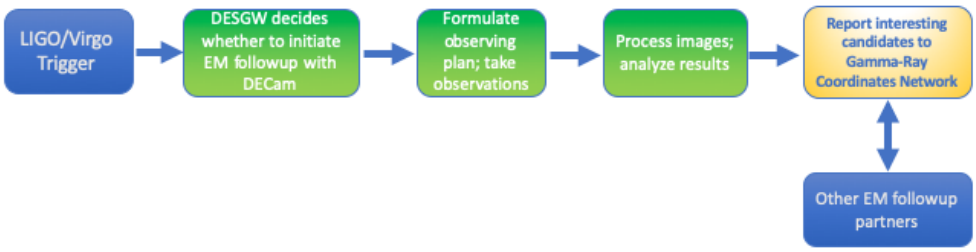


**Figure 1.** Overview of the DESGW gravitational wave followup program [2].

## 2 Single-epoch processing improvements

One of the main lessons learned during the O1 and O2 observing seasons was that there was a clear need to reduce the total time from search image collection to public release of EM counterpart candidate information. During real-time GW followup campaign it is important to released vetted candidates as soon as possible in order to enable spectroscopic followup by a variety of instruments around the world. During O1 and O2 our typical times to collect a search image, process it, and apply all section criteria to any candidates was typically between five and eight hours. During O3 our goal is to reduce this time to one hour. To meet this goal we have refactored part of our pipeline to make several steps run in parallel that were previously serial.

Our image processing pipeline begins with Single-Epoch (SE) processing on the raw telescope image. This step applies calibrations and corrections such as crosstalk correction, pixel correction, astrometric and photometric calibration, and object cataloguing. In earlier versions of our pipeline there was a single grid job that performed SE processing on all CCDs in entire search image. These jobs would typically take approximately one hour, but sometimes up to between two and three hours for a single image, taking up a significant part

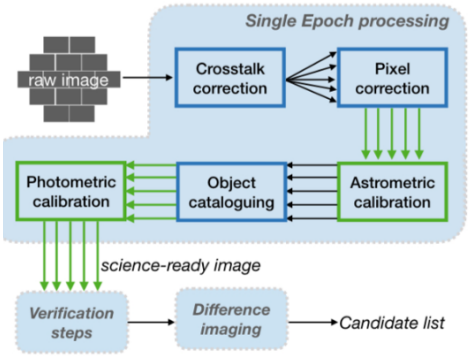


**Figure 2.** Diagram of the updated single-epoch processing pipeline. Green arrows indicate steps that were previously performed on all CCDs simultaneously and are now fully parallel per CCD in the updated O3 pipeline.

of the time from image acquisition to candidate identification (the difference imaging step is typically 30-40 minutes per job). Furthermore no difference imaging could proceed until SE processing was finished for all CCDs. In order to speed up this step we have refactored the pipeline so that the SE step now runs in parallel on a CCD-by-CCD basis, as shown in Figure 2. This arrangement enables a single job to perform both the SE and difference imaging step for a given CCD in each search image, minimizing overhead times. As shown in Figure 3, the time for the SE portion of the refactored pipeline is now typically under 15 minutes, with over 50% of the jobs completing in under 10 minutes. This represents roughly a factor of 5 improvement over the old pipeline with no loss of fidelity. Additionally in this approach CCDs that finish their SE processing faster no longer have to wait for "slow" CCDs to finish and can proceed immediately into the difference imaging step. Our tests show that the refactored pipeline does produce results within the one-hour time budget.

One drawback of the per-CCD SE processing approach is that it is more difficult to compute the astrometric solution without having the entire focal plane available. It is still possible, but it requires a denser astrometric catalog than the catalog used during the first two observing seasons. We now use the GAIA-DR2 catalog [5–7], enabling us to achieve sufficiently accurate astrometric solutions in a single CCD to effectively run difference imaging.

## 3 Candidate identification improvements

To expedite the identification of potential GW counterparts in the DECam images, we have designed and implemented a machine-learning-based photometric classification algorithm called `KN-Classify` [8]. `KN-Classify` utilizes an “individualized training set approach,” in which the training set is simulated immediately following observations and incorporates the exact seeing, sky-brightness, exposure times, optical filters, and time-spacing of exposures into the simulated photometric data used for training. Using models from the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) [9], we simulate time-series photometric observations of of a variety of astrophysical sources as they would appear in the DECam data using the SuperNova ANAlysis software suite (SNANA) [10]. The optimal light curve features for kilonova detection are selected algorithmically. The classifier itself is a Random Forest classifier [11] for which the number of estimators, maximum depth

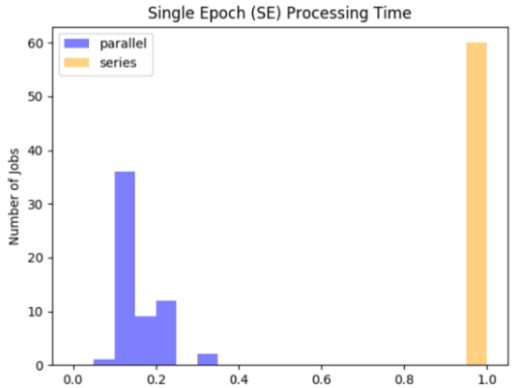


**Figure 3.** Purple: Wall time for the SE stage of the updated parallel workflow for a single exposure. There are 59 parallel jobs, one for each CCD used in the analysis. Yellow: wall time for SE processing of the same exposure using the O1/O2 pipeline. In the O1/O2 pipeline there was one SE job per exposure; we normalize it to 59 here for illustrative purposes. Since all 59 jobs are independent in the O3 pipeline, the effective time for SE processing to complete is simply that of the longest single job, providing an improvement factor of three to five depending on the individual exposure. The O3 pipeline jobs also immediately move to the next stage of processing in each CCD, whereas in the O1/O2 pipeline no diffimg jobs for a given exposure would start until all SE processing was complete.

of the constituent decision trees, class weights, and decision tree optimization criterion are all chosen to optimize kilonova identification.

`KN-Classify` is a powerful classifier and the individualized training set approach makes it robust against varying or poor observing conditions. In Figure 4, we display the performance of `KN-Classify` for a follow-up with various combinations of seeing, moon-illumination (a proxy for sky-brightness), and time between the LVC GW detection and the start of DECam observations. As a metric, we show the fraction of the time a KN is placed as one of the ten most probable candidates in the follow-up (out of $\sim 20,000$ candidates) since we would anticipate being able to target $\sim 10$ objects with spectroscopic instruments to obtain an unambiguous characterization. This particular follow-up assumes DECam collected only a total of four exposures (2 each in $r, i$ with 1 hour between each $r, i$ set) in each region of interest in the sky. Thus, with only four photometric data points, `KN-Classify` is capable of placing a true KN in the list of the ten most probable candidates more than 99% of the time in arbitrary observing conditions.

We also match each candidate against a galaxy catalog. For O3, we have improved our catalog by including coadd photometry with validated photometric z data from three different algorothms. The new catalog also contains spectroscopic redshift information for a much larger fraction of galaxies, which we constantly update as soon as new spectroscopic data become available.

We have also added tools to automatically query the Transient Name Server to check for detections of our counterpart candidates at other facilities prior to the GW event time. Such detections indicate that the candidate is highly unlikely to be related to the GW event.

## 4 Web page design

We have redesigned our web pages for each followup campaign and individual candidate. Figure 5 shows an example page listing all candidates for a given LVC trigger, including

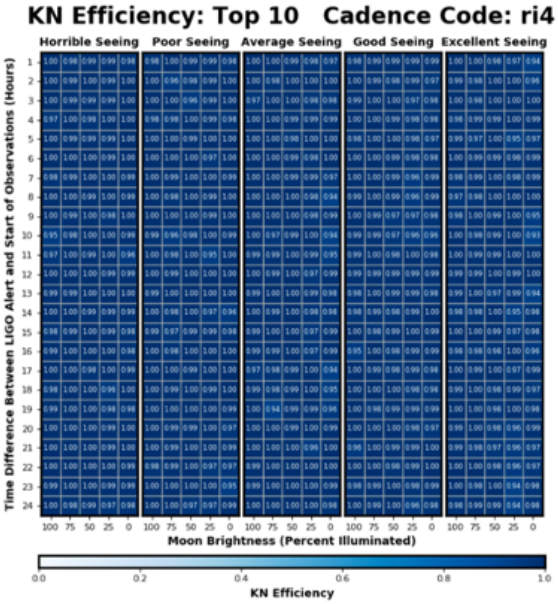


**Figure 4.** Fraction of the time `KN-Classify` ranks a KN as one of the ten most interesting candidates out of $\sim 20,000$ for various observing conditions. The observing cadence used for this plot was two sets of exposures in $r, i$ separated by 1 hour.

position, maximum ML score, and paths for accessing images. There is also a link for each candidate's dedicated page. Figure 6 shows an example of an individual candidate page. These pages show each candidate's complete record of observations including ML score, flux, magnitude, and observing conditions, the image cutouts, and a light curve showing the candidate's evolution over time.

## 5 Conclusions

The DESGW group has made a number of improvements to its already-successful search and discovery pipeline and deployed them during the third LIGO observing season. These improvements include additional pipeline automation, refactoring to make the pipeline more parallel, more accurate astrometric solutions, . Together, these improvements reduce the time from image acquisition to candidate identification by a factor of approximately five with no increase in the false positive rate.

Funding for the DES Projects has been provided by the DOE and NSF (USA), MEC/MICINN/MINECO (Spain), STFC (UK), HEFCE (UK). NCSA (UIUC), KICP (U. Chicago), CCAPP (Ohio State), MIFPA (Texas A&M), CNPQ, FAPERJ, FINEP (Brazil), DFG (Germany) and the Collaborating Institutions in the Dark Energy Survey. This research uses services or data provided by the NOAO Science Archive. NOAO is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the National Science Foundation. R. Morgan thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE- AC02-07CH11359 with the U.S. Department of Energy, Office of

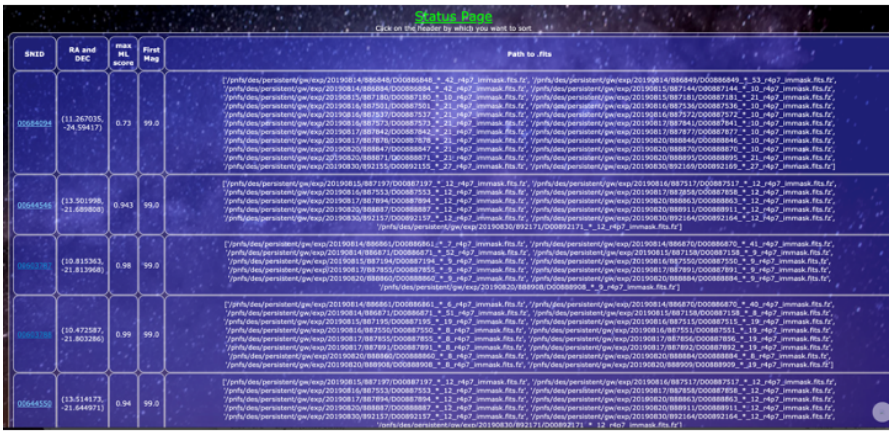


**Figure 5.** Example web page listing EM counterpart candidates for a given gravitational wave event. One can sort the page based on position and maximum machine learning score. By clicking on a candidate ID, one can view a more detailed candidate-specific page (see Fig. 6).

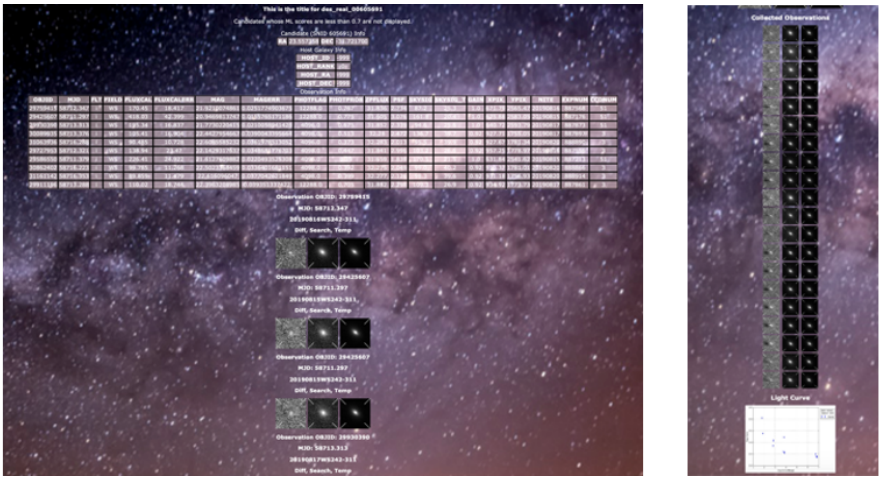


**Figure 6.** Example web page for a single EM counterpart candidate. The page provides a sky location, the time, magnitude, and observing condition information for each observation of the candidate (left panel). Also shown are image cutouts or "stamps" from the template, search, and difference image for each observation (left panel). Continuing down the page (right panel), one finds the collected observations presented together along with a light curve plot for the candidate.

Science, Office of High Energy Physics. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

## References


[1] B. Flaugher *et al.*, Astron. J. **150**, 150 (2015).
[2] K. Herner *et al.*, arXiv:2001.06551 (2020).
[3] E. Morganson *et al.*, Publ. Astron. Soc. Pac. **130**, 074501 (2018).
[4] R. Kessler *et al.*, Astron. J. **150**, 172 (2015).
[5] T. Prusti *et al.*, Astron. & Astrophys. **595**, A1 (2016).
[6] A. G. A. Brown *et al.*, Astron. & Astrophys. **616**, A1 (2018).
[7] L. Lindegren *et al.*, Astron. & Astrophys. **616**, A2 (2018).
[8] R. Morgan *et al.*, in preparation.
[9] T. Allam Jr *et al.*, arXiv:1810.00001 (2018).
[10] R. Kessler *et al.*, Publ. Astron. Soc. Pac. **121**, 883 (2009).
[11] L. Breiman, Machine Learning. **45**, 1 (2001).