Lilith: A Versatile Instrument and All-Sky Simulator for use with Space-Based Astrophysics Observatories

Jeffrey Smith,[1 2] and Peter Tenenbaum,[1] and Jon M. Jenkins,[2] and Joseph D. Twicken,[1]

[1]*SETI Institute, Mountain View, CA, USA;* `jeffrey.smith@nasa.gov`

[2]*NASA Ames Research Center, Moffett Field, CA, USA*

Abstract.
To help facilitate the development of the Transiting Exoplanet Survey Satellite (TESS) data analysis pipeline, it was necessary to produce simulated flight data with sufficient fidelity and volume to exercise all the capabilities of the pipeline in an integrated way. As a generator of simulated flight data, *Lilith*, was developed for this purpose. We describe the capabilities of the *Lilith* software package, with particular attention to the interaction between the implemented features and the pipeline capabilities that it exercises. Full instrumental and astrophysics ground truth is available and can be used as a training set for TESS data analysis software, such as when training a machine learning classifier for planet candidates. Our intention is to continue to tune*Lilith* as real TESS flight data becomes available, allowing for an up-to-date simulated set of data products to complement the mission flight data products, thereby aiding researchers as they continue to adapt their tools to the TESS data streams. We discuss the execution performance of the resulting package, and offer some suggestions for improvements for instrument and sky simulators to be developed for other missions.

1. Simulated TESS Data

The Transiting Exoplanet Survey Satellite (TESS) (Ricker et al. 2014) is a space-based planet-finding NASA mission. TESS will identify the best small ($< 4R_e$) planets for detailed follow-up and characterization during the next several decades. This is accomplished by conducting an all-sky transit survey of ~200,000 stars including F, G and K dwarfs between 4 and 12 magnitudes and most M dwarfs within ~200 lightyears. For each 27.4-day period, TESS will observe a 24Âř by 96Âř swath of sky extending from near the ecliptic equator to the ecliptic pole. TESS is expected to discover ~1,000 small planets less than four times the size of Earth. At least fifty of these small worlds will orbit stars sufficiently bright to allow for determination of their masses. Because these stars are typically 10ÃŮ closer and 100ÃŮ brighter than those observed by the Kepler Mission, they are much more amenable to follow-up observations and characterization, such as by the James Webb Space Telescope (JWST).

To this end, the instrument will select small regions of pixels (approximately 10 pixels x 10 pixels) around each of up to 16,000 target stars, and produce images of these regions with high frequency (nominally 2-minutes) in order to support fine time resolution photometry. In addition, the full field of view (FOV) of the instrument (24 degrees x 96 degrees) will be captured at somewhat lower frequency (nominally 30-minutes).

Building upon the great success of the Kepler mission, the TESS Science Processing Operations Center (SPOC) team at NASA Ames Research Center adapted the Kepler Science Processing Pipeline (Jenkins 2017) for use with TESS (Jenkins et al. 2016). The pipeline runs on the NAS Pleiades supercomputer and provides calibrated pixels, simple and systematic error-corrected aperture photometry, and centroid locations for all target stars observed over the 2-year mission, along with associated uncertainties. The TESS pipeline will search through all light curves for evidence of periodic transit signals that occur when a planet crosses the disk of its host star. It will generate a suite of diagnostic metrics for each transit-like signature discovered, and extract planetary parameters by fitting a limb-darkened transit model to each potential planetary signature. The results of the transit search will be archived to the Mikulski Archive for Space Telescopes (MAST)[1].

To help facilitate the development of the SPOC science pipeline, it was necessary to produce simulated flight data with sufficient fidelity and volume to exercise all the capabilities of the pipeline in an integrated way. As a generator of simulated flight data, *Lilith*, was developed for this purpose. Using a physics-based TESS instrument and sky model, *Lilith* creates a set of raw TESS data which includes models for the CCDs, readout electronics, camera optics, behavior of the attitude control system (ACS), spacecraft orbit, spacecraft jitter and the sky, including zodiacal light, and the TESS Input Catalog. The model also incorporates realistic instances of stellar astrophysics, including stellar variability, eclipsing binaries, background eclipsing binaries, transiting planets and diffuse light. This data can then be passed to the SPOC pipeline providing full integration tests of the science processing all the way from raw pixel calibration to the generation of archivable data products.

The Kepler, and now TESS, pipelines are written in Matlab. We therefore chose Matlab for *Lilith* to leverage all the pipeline functionality that has been developed over the past 20 years of the Kepler mission pipeline development. Due to the somewhat slow execution performance of Matlab, key functionality requiring fast execution is written in C++ and called within Matlab via the “mexfile” interface. The shared codebase also helps facilitate fully integrating *Lilith* into the development and execution process: A pipeline run using *Lilith* and flight data is virtually indistinguishable to the pipeline operators.

2. The Importance of Realistic Simulated Data

One key importance to the TESS pipeline is that it is, in fact, a *pipeline*. The inputs to one component are the outputs of a previous component. It is therefore necessary to produce a data set that introduces all the test features for the entire processing pipeline into the raw pixel data. In this way, it is possible to see the effect of interactions between the elements, i.e., it exposes situations in which the actions of one element corrupt information in the test data in a way that causes a subsequent element to malfunction. This, then, requires an integrated simulation: one in which all phenomena from the

[1]https://archive.stsci.edu/tess/

pixel level (readout noise, etc.) to the astrophysical level (transit signatures, etc.) are generated.

There are two principle types of modelling performed in *Lilith*: 1) astrophysical and 1) instrumental. The fundamental process of Lilith star field rendering is a mathematical model of a point source that incorporates all the effects that contribute to its rendered appearance. This model is an extension of the Point Spread Function (PSF), and is known as a Pixel Response Function (PRF). The details of the rendering is beyond the scope of this paper but to summarize, we first generate an astrophysics based model of the star field, we then use physics-based model of the instrument optics, CCD and readout electronics to generate a sub-cadence PRF model and finally pass the astrophysical scene through the PRF to record the pixel values.

Our intention is to continue to tune *Lilith* as real TESS flight data becomes available, allowing for an up-to-date simulated set of data products to complement the mission flight data products, thereby aiding researchers as they continue to adapt their tools to the TESS data streams. Numerous studies have been performed with the Kepler data set to determine how well various astrophysical signals can be reliably extracted CITE THESE!. Crucial to these studies is a clear understanding of ground truth. For space-based all-sky surveys, such as TESS, there are rarely sufficient independent data streams to provide a ground truth reference. We are therefore compelled to provide realistic simulated data streams which when passed through the TESS pipeline can be compared to quantitatively measure signal reliability. Now that TESS is collecting data we will compare the flight data to our simulation and update both the instrumental model to adapt to reality. If we do indeed find inaccuracies in our astrophysics model then those will be updated as well.

Discuss:

- give examples of this being used with Kepler data (cite papers).
- Comparisons of signal properties between simulated and flight data.
- How to improve the realism of the simulated data.

3. Generating Training Data for Machine Learning

Full instrumental and astrophysics ground truth is available and can be used as a training set for TESS data analysis software, such as when training a machine learning classifier for planet candidates.

Discuss:

- Summer FDL work
- Large quantities of injected transits
- Compare contrast use of real flight data with injected planets versus fully simulated data

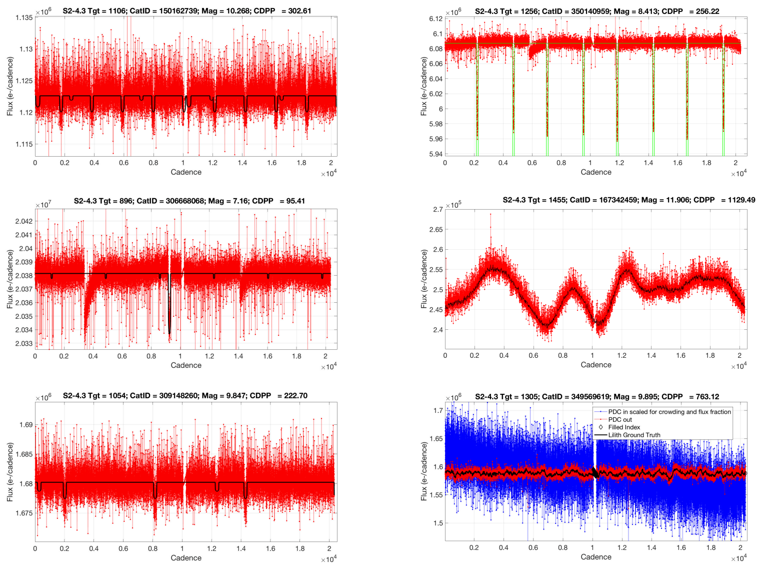

Figure 1. Example light curves generated by Lilith. UPDATE FIGURE!!!!!

4. Peformance

We discuss the execution performance of the resulting package, and offer some suggestions for improvements for instrument and sky simulators to be developed for other missions.

References

Jenkins, J. M. 2017, Kepler Data Processing Handbook: Overview of the Science Operations Center, Tech. rep., NASA

Jenkins, J. M., et al. 2016, in Software and Cyberinfrastructure for Astronomy IV, vol. 9913 of SPIE, 99133E

Ricker, G. R., et al. 2014, in Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, vol. 9143 of SPIE, 914320. `1406.0151`