MIT Open Access Articles

The TESS science processing operations center

Massachusetts Institute of Technology

PROCEEDINGS OF SPIE

The TESS science processing operations center

Jon M. Jenkins, Joseph D. Twicken, Sean McCauliff, Jennifer Campbell, Dwight Sanderfer, et al.

SPIE.

The TESS Science Processing Operations Center

Jon M. Jenkins[a], Joseph D. Twicken[b], Sean McCauliff[c], Jennifer Campbell[c], Dwight Sanderfer[a], David Lung[d], Masoud Mansouri-Samani[e], Forrest Girouard[f], Peter Tenenbaum[b], Todd Klaus[e], Jeffrey C. Smith[b], Douglas A. Caldwell[b], A. Dean Chacon[d], Christopher Henze[a], Cory Heiges[g], David W. Latham[h], Edward Morgan[i], Daryl Swade[j], Stephen Rinehart[k], and Roland Vanderspek[i]

[a]NASA Ames Research Center, Moffett Field, CA, USA
[b]SETI Institute/NASA Ames Research Center, Moffett Field, CA, USA
[c]Wyle Labs/NASA Ames Research Center, Moffett Field, CA USA
[d]Millenium Engineering/NASA Ames Research Center, Moffett Field, CA USA
[e]SGT, Inc./NASA Ames Research Center, Moffett Field, CA USA
[f]Logyx LLC/NASA Ames Research Center, Moffett Field, CA USA
[g]General Dynamics/NASA Goddard Space Flight Center, Greenbelt, MD USA
[h]Harvard-Smithsonian Center for Astrophysics, Cambridge, MA USA
[i]Massachusetts Institute of Technology, Cambridge, MA USA
[j]Space Telescope Science Institute, Baltimore, MD
[k]NASA Goddard Space Flight Center, Greenbelt, MD USA

ABSTRACT

The Transiting Exoplanet Survey Satellite (TESS) will conduct a search for Earth's closest cousins starting in early 2018 and is expected to discover $\sim$1,000 small planets with $R_p < 4\ R_\oplus$ and measure the masses of at least 50 of these small worlds. The Science Processing Operations Center (SPOC) is being developed at NASA Ames Research Center based on the *Kepler* science pipeline and will generate calibrated pixels and light curves on the NASA Advanced Supercomputing Division's Pleiades supercomputer. The SPOC will also search for periodic transit events and generate validation products for the transit-like features in the light curves. All TESS SPOC data products will be archived to the Mikulski Archive for Space Telescopes (MAST).

Keywords: transit photometry, science pipelines, TESS mission, exoplanets, high performance computing

1. INTRODUCTION

The Transiting Exoplanet Survey Satellite (TESS) was selected by NASA's Explorer Program to search for Earth's closest cousins starting in early 2018. Slated for launch in December 2017, TESS will identify the best small ($<4\ R_\oplus$) planets for detailed follow-up and characterization during the next several decades. This is accomplished by conducting an all-sky transit survey of F, G and K dwarf stars between 4 and 12 magnitudes and M dwarf stars within 200 light years.[1] TESS will observe at least 200,000 stars over the course of the mission and is expected to discover $\sim$1,000 small planets less than four times the size of Earth.[2] At least fifty of these small worlds will orbit stars sufficiently bright to allow for determination of their masses. Because these stars are typically $10\times$ closer and $100\times$ brighter than those observed by the *Kepler* Mission, they are much more amenable to follow-up observations and characterization. Indeed, the James Webb Space Telescope should be able to characterize the atmospheres of many of the TESS discoveries.

The TESS science pipeline is being developed by the Science Processing Operations Center (SPOC) at NASA Ames Research Center based on the highly successful *Kepler* Science Operations Center (KSOC) pipeline.[3] Like

---

the *Kepler* pipeline, the TESS pipeline will provide calibrated pixels, simple and systematic error-corrected aperture photometry, and centroid locations for all 200,000+ target stars observed over the 2-year prime mission, along with associated uncertainties. The temporal resolution of the primary TESS science data are significantly enhanced over that for *Kepler*, with a nominal cadence of 2 min compared to 29.4 min for *Kepler*.[4] For efficiency, the science data will be processed on the NASA Advanced Supercomputing (NAS) Division's Pleiades supercomputer.[*] The pixel and light curve products are modeled on those of *Kepler* and will be archived to the Mikulski Archive for Space Telescopes (MAST). In addition to the nominal science data, the 30-minute Full Frame Images (FFIs) simultaneously collected by TESS will also be calibrated by the SPOC. There is no proprietary data period for any TESS data products: all data will be made available to the community at the same time they are made available to the science team when they are archived to MAST.

As shown in Fig. 1, TESS will spend ∼27.4-days observing a 24° by 96° swath of sky extending from ∼6° above the ecliptic equator to the ecliptic pole in the anti-Sun direction. TESS orbits the Earth in a lunar-resonant 2:1 elliptical orbit, with nominal perigee and apogee of 17 $R_\oplus$ and 59 $R_\oplus$. Each field of view (FOV) is observed for two consecutive ∼13.7-day orbits, after which TESS rotates eastward by ∼27° in order to cover the southern hemisphere during the first year of observations. Kozai-like perturbations to the spacecraft's orbit due to the Moon induce variations in the orbital period of ∼5%, and the orbit is stable over thousands of years.[5] After the southern hemisphere is observed, the spacecraft then flips over to cover the northern hemisphere during the second year. While most of the stars are observed for only ∼27.4 days, the pole-ward camera is centered on the ecliptic pole, allowing a ∼450 square degree area in each hemisphere to be observed continuously for a full year.

The TESS pipeline will search through all light curves for evidence of transits 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 are modeled on the *Kepler* transit search products (tabulated numerical results, time series products, and pdf reports) all of which will be archived to MAST.
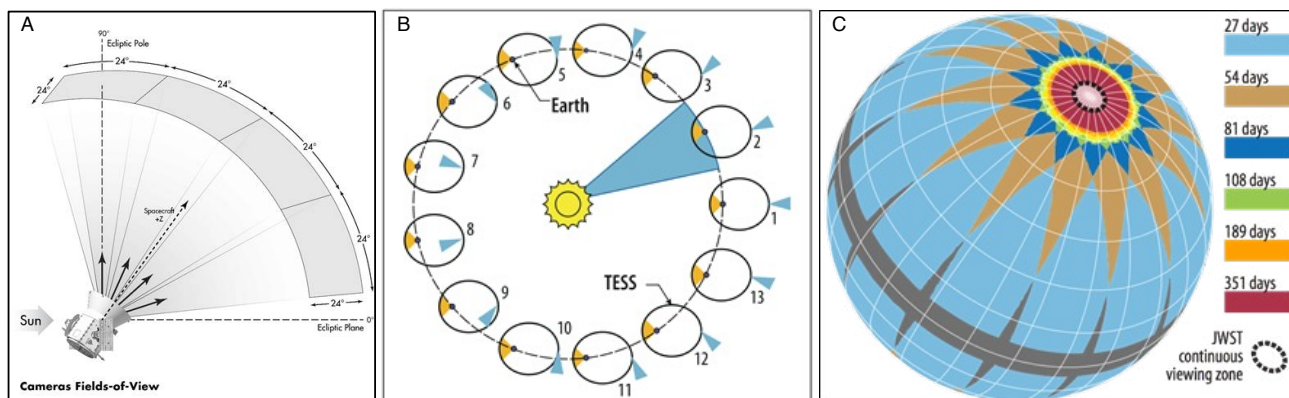

Figure 1. The TESS FOV is composed of four cameras, each with a 24° × 24° FOV arrayed in a column that spans from 6° above the ecliptic to 12° beyond the ecliptic pole (panel A). TESS will collect data for each pointing for two orbits, ∼27.4 days, and then rotate ∼27° so that the center of the next FOV is along the anti-Sun direction at the middle of the observing period (panel B). The entire southern hemisphere will be observed after 13 pointings over a one-year period and TESS will then flip over to repeat the observations for the northern hemisphere. Most stars are observed for only ∼27 days, but the pole-ward camera will observe ∼450 square degrees continuously for one year (panel C).

Several factors pose key challenges to scaling the *Kepler* science processing facility to meet the requirements for TESS, namely the faster turn-around time for processing the original science data (27.4 days for TESS vs. 120 days for *Kepler*), the much higher on-board storage capacity (385 GiB vs. 17 GiB), and the much higher data rate afforded by TESS' closer orbit to Earth. Table 1 lists the relevant quantities determining data rate and total data volume for the original data collected on board each spacecraft once it's downlinked to the ground and

---

[*]Pleiades currently has 227,808 computer cores and 828 TiB of memory (http://www.nas.nasa.gov/hecc/resources/pleiades.html).

decompressed. Note that the data rate for the TESS science data alone is $\sim13\times$ that of *Kepler*, and is $\sim25\times$ that of *Kepler* when the FFIs are considered. It's important to consider the fact that the processed science data expands the total data volume footprint considerably. Even though the *Kepler* original data volume amounts to ~1.6 TiB, the MAST is holding ~19 TiB of *Kepler* data products, for an expansion factor of ~12. In addition, data volume presents memory management challenges as well as increased storage requirements. Fortunately, data storage has become more affordable in the ten years since the original data storage system was purchased for *Kepler*. Data transfer rates have also improved significantly on the ground.

Table 1. TESS and Kepler Data Rates.

	Kepler	TESS	TESS/Kepler
# Stars	165000	15000	0.09
Samples day^{-1}	49	720	14.69
Pixels Per Star	32	100	3.13
Days FOV^{-1}	93	27.3	0.29
Mission Duration (years)	4	2.0	
Target Star Pixels day^{-1}	258,720,000	1,080,000,000	4.17
Background Target Pixels	378,000	0	0.00
Background pixels day^{-1}	18,522,000	0	0.00
Collateral Pixels	280,056	3,908,864	13.96
Collateral Pixels day^{-1}	13,722,744	2,814,382,080	205.09
All Science Pixels day^{-1}	290,964,744	3,894,382,080	13.38
All Science Pixels GiB day^{-1}	1.08	14.51	13.38
All Science Pixels GiB FOV^{-1}	101	396.37	3.93
All Science Pixels GiB mission^{-1}	1,583	10,591	6.69
FFI pixels	97,370,112	71,017,728	0.73
FFI Samples day^{-1}	0.03	48	1,488.00
FFI Pixels day^{-1}	3,140,971	3,408,850,944	1,085.29
FFI GiB day^{-1}	0.01	12.70	1,085.29
Sciencei+FFI GiB day^{-1}	1.10	27.21	24.83
Science+FFI GiB FOV^{-1}	101.89	743.33	7.30
Science+FFI GiB mission^{-1}	1,600	19,861	12.42

This paper provides an overview of the TESS science pipeline and describes the development of the SPOC remaining before launch in late 2017. The paper presents an overview of the ground segment in Sec. 2, then describes the architecture of the SPOC in Sec. 3. The approach developed for estimating the software development costs of the SPOC is discussed in Sec. 4. Secs. 5 and 6 describe how the SPOC pipeline is run on Pleiades and the operations concept for the SPOC. The paper concludes with a summary in Sec. 7.

2. THE TESS MISSION GROUND SEGMENT

The TESS ground system is illustrated in Fig. 2, and includes all the elements necessary to manage the spacecraft and instrument, collect, analyze, and interpret the science data, and to deliver it to a permanent archive.

The TESS spacecraft transmits the compressed, raw science data and engineering data through the **Deep Space Network (DSN)**[†] to the **Payload Operations Center (POC)** at MIT. The **Flight Dynamics Facility (FDF)** at Goddard Spaceflight Center reconstructs the spacecraft orbit and sends the resulting ephemerides to the POC. The **Navigation and Ancillary Information Facility (NAIF)** at JPL delivers SPICE kernels with ephemerides for all the solar system objects it tracks, including the Sun-Earth-Moon system, to the POC.

The **POC** is the hub of the TESS ground segment, distributing information and data products throughout the ground segment as needed to support science operations. It passes the original science and engineering

---

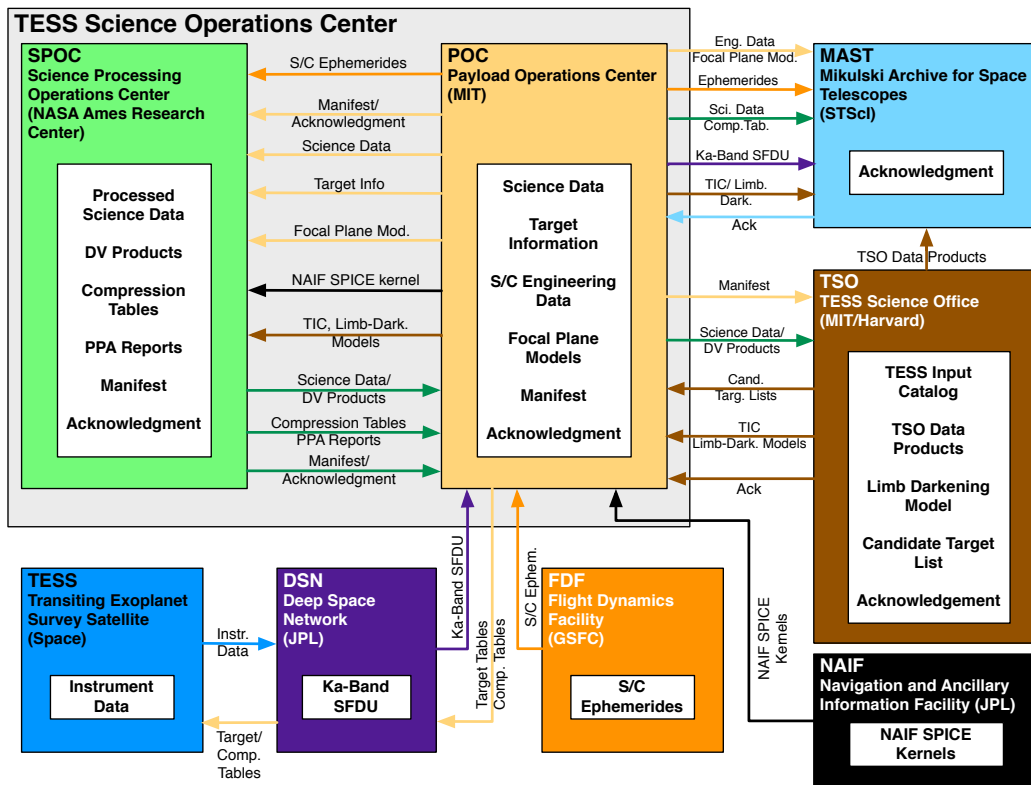[†]The DSN is operated by JPL: http://deepspace.jpl.nasa.gov/.

Figure 2. The TESS ground segment is composed of the Science Processing Operations Center at NASA Ames Research Center, the Payload Operations Center at MIT, the Mikulski Archive for Space Telescopes at STScI, the TESS Science Office in Cambridge MA, the Navigation and Ancillary Information Facility at JPL, the Flight Dynamics Facility at GSFC, and the Deep Space Network. This system manages the TESS science instrument to collect the science data and all the information necessary to carry out TESS's main objective: searching for Earth's nearest cousins. The diagram illustrates the flow of the information and data across the system.

data to the permanent archive at the **Mikulski Archive for Space Telescopes (MAST)** located at the Space Telescope Science Insitute (STScI) in Maryland, along with the archival science data products and all the information necessary to process and analyze the science data, including ephemerides, and limb-darkening models. The POC manages the science instrument and develops focal plane models critical for processing the science data, including items such as the flat field, the pixel response function (PRF) and focal plane geometry (FPG). The POC accepts candidate target lists from the **TESS Science Office (TSO)**, and formulates the flight target lists that determine which pixels are stored and downlinked from the spacecraft at the nominal 2-minute science cadence.

The POC decompresses and delivers the original science and engineering data to the **Science Processing Operations Center (SPOC)** at NASA Ames Research Center in Moffett Field CA, formatted into Google protocol buffers, along with the focal plane models, target lists, and the target star information. The SPOC processes the original science data to produce calibrated pixels, light curves and centroids for all target stars, and associated uncertainties. The SPOC also calibrates the FFIs and provides world coordinate system (WCS) information for each frame. The SPOC develops compression tables based on commissioning data and sends these to the POC for formatting and uplinking to the spacecraft to achieve a compression performance meeting or exceeding 12 bits per science pixel. The SPOC searches the light curves for transiting planets and constructs a suite of diagnostic tests to make or break confidence in the planetary nature of each transit-like signature. The pixel, light curve, and transit search data products are sent to the POC, which transmits them to the permanent archive at MAST, as well as to the TSO. Together, the POC and the SPOC comprise the **TESS Science Operations Center (SOC)**.

The **TESS Science Office (TSO)** in Cambridge MA manages the TESS Input Catalog (TIC), a catalog of up to 1×10^{10} stars with their celestial coordinates, right ascension and declination (α, δ), their apparent TESS magnitudes, and other stellar parameters. The TSO is composed of personnel from both the Harvard Smithsonian Center for Astrophysics (CfA) as well as MIT. The TSO prioritizes the potential target stars for each Field of View (FOV) and sends the list to the POC to be merged with the Guest Investigator (GI) target stars selected through a competitive process. The TSO vets the potential transiting planets identified by the SPOC transit search engine to compile a catalog of TESS Objects of Interest (TOIs), building their own diagnostics and vetting products in addition to those furnished by the SPOC. The TSO also coordinates the funded TESS follow-up observation program.

3. THE SCIENCE PROCESSING OPERATIONS CENTER

The main function of the TESS SPOC is to process the science data for the TESS mission, store the processed data for a limited time, and report the results. This is accomplished via the science analysis pipeline. Other functionality provided by the SPOC will include reporting, archiving, and updating target lists and models. The SPOC pipeline will receive raw data from other ground segment elements routed through the Payload Operations Center (POC). The high-level architecture of the SPOC is shown in Fig. 3.
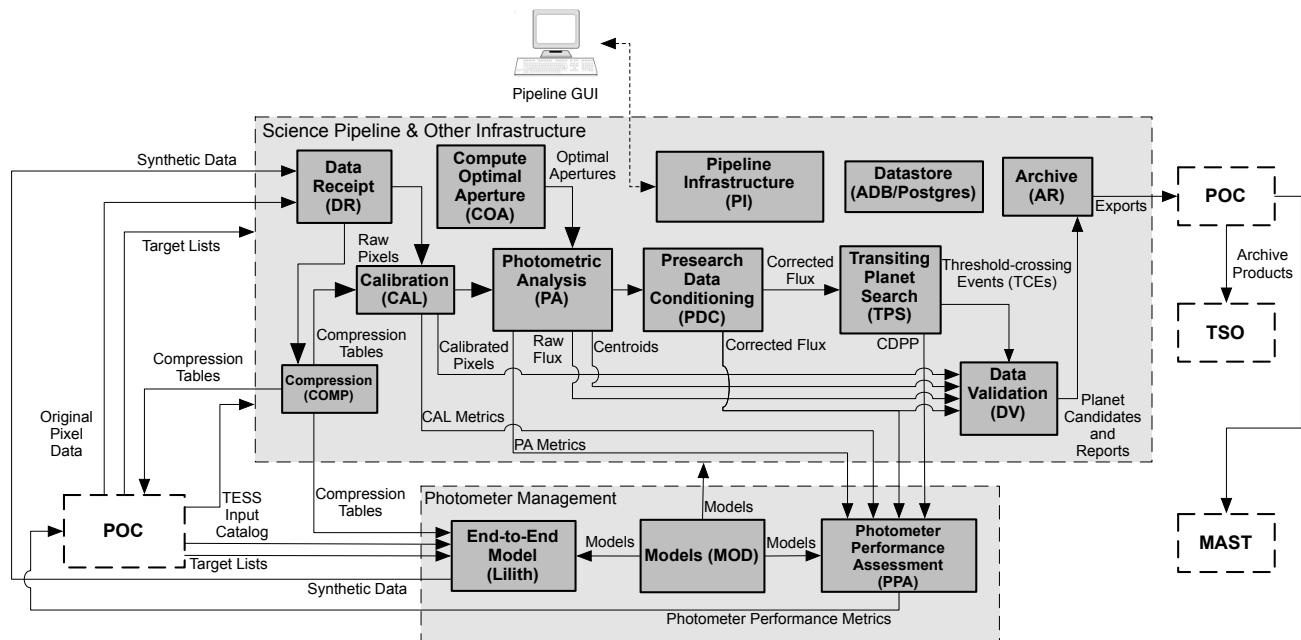

Figure 3. Architecture diagram of the TESS Science Processing Operations Center indicating the 14 major components comprising the TESS science pipeline and photometer management functions.

The SPOC is modeled on the KSOC, which consists of several elements: 1) a pipeline infrastructure coded in Java that ingests the science data, controls the science pipeline, writes the archive data products to files in the archive file format, 2) the science pipeline, which processes the raw pixels, extracts photometry and astrometry for each target star, identifies and removes systematic errors, identifies transiting planet signatures, and performs a suite of diagnostic tests on such signatures, 3) a target management system that contains a catalog of target and field stars and their characteristics, and a module that identifies the pixels of interest for each target star and collateral data needed for calibrating the target star data.

Pipeline Infrastructure (PI): The pipeline infrastructure provides fully automated distributed processing of science data and sequencing of modules based on the results of previous modules. Features applicable to TESS include:

- A customizable unit-of-work that controls how the data are distributed across the cluster

- Configuration management and versioning for algorithm parameters and pipeline configurations

- A graphical user interface for the configuration, execution, and monitoring of pipeline jobs.

- Scalability from a developer workstation to a large cluster of computing nodes and to Pleiades

PI was developed for the *Kepler* Mission, but is designed to be a generic, reusable platform suitable for developing science processing and analysis pipelines for a variety of missions. PI provides a plug-in architecture for pipeline modules, parameter definitions, and the unit-of-work definition, and contains no *Kepler*-specific code.

PI also provides a generic mechanism for running any pipeline module on large computing clusters that support the PBS (Portable Batch System) interface. *Kepler* has used this capability to process nearly all its science data on NASA's Advanced Supercomputer (Pleiades) and TESS uses the same mechanism.

Data Receipt (DR): This module provides data ingest and automated pipeline launch capabilities. Like PI, DR is divided into two main components, 1) a generic layer that handles watching for new files, dispatching to the proper handler, and launching pipelines, and 2) a plug-in layer for specific data types.

Data Store (DS): The Data Store is a transactional database management system for arrays, sparse arrays and binary data. It is composed of both a custom array data base (ADB) as well as a PostgreSQL database. The vast majority of data for use by the pipeline modules is stored in this database. TESS has higher scalability needs compared with *Kepler*.

Archive (AR): This software component generates the data products archived at the MAST and made available to the science team and greater astronomical community. The TESS archival products mirror the standard *Kepler* archival products, which include calibrated pixels, simple aperture photometry, systematic-error-corrected photometry, astrometry (centroids), and associated uncertainty estimates. The products include target pixel files that contain the pixel data, both original and calibrated, for each target organized as image data. These files also include background flux information and information about cosmic ray hits detected by the pipeline (if cosmic ray cleaning is enabled). The products of the transit search and the suite of DV diagnostic tests are archived as xml files along with pdf reports. These pdf files contain a wealth of information regarding each transit-like signature that is processed through DV. The TESS archive data products are modeled on those of *Kepler*, providing a set of familiar and easy-to-use data products to the transit photometry and astrophysics community.

Support Libraries (SLIB): The pipeline infrastructure depends on a set of support libraries that have been customized and augmented specifically for the KSOC. Most of this software is written in Java, but there is some MATLAB software in this collection to provide, for example, the capability to retrieve data and model information from the Data Store from an interactive MATLAB session on a KSOC development workstation. Much of this software underlies the automated test suites that are exercised on a nightly basis to check the codebase integrity. While not a module per se, this code is important to consider in estimating the software development cost of the SPOC as discussed in Sec. 4.

Compression (COMP) Pixel data compression tables are generated in the COMP module. Compressing the flight science data reduces the storage requirements aboard the spacecraft and reduces the requisite downlink time. The compression tables are sent to the POC/MOC for uplink. There are two stages of data compression:[6]

1. The data values representing co-added exposures for each pixel are requantized to make the quantization noise approximately a fixed fraction of the intrinsic measurement uncertainty. Small quantization steps are assigned to small data values with inherently low shot noise, and large quantization steps are assigned to large data values with inherently high shot noise.

2. The difference between each requantized pixel value and the prior value are entropically encoded via a length-limited Huffman code string. Significant compression is achieved because the scene does not change markedly from cadence to cadence in most cases. Huffman coding and subsequent decoding is lossless.

For operational and algorithmic simplicity, nominal 2-min science data and FFIs each have their own requantization and Huffman coding tables.

3.1 Photometer Management

This suite of software contains the models describing various aspects of the photometer that are relevant to data processing as well as modules that examine flight data to monitor the health and status of the photometer at different timescales.

Photometer Performance Assessment (PPA): This component assesses the health and performance of the instrument based on the science data sets collected each orbit, identifying out of bounds conditions and generating alerts. PPA is implemented with multiple pipeline modules: PPA Metrics Determination (PMD), PMD Aggregator (PAG), and PPA Attitude Determination (PAD). Various metrics are generated as the science pipeline processes the science data, including photometric precision, brightness, black level, background flux, smear level, dark current, cosmic ray counts, outlier counts, centroids, reconstructed attitude, and the difference between the reconstructed and nominal attitudes. These metrics are tracked and trended by PPA and the results are persisted to the Data Store and are used to generate a pdf report that is available for review by TESS project personnel. PAD combines the astrometric data collected for each CCD readout channel to construct a high fidelity record of the pointing history for each camera for each 2-min data collection interval and compares them against the nominal attitudes. The output of PPA is used to identify and set data anomaly flags required for the second and final processing of the science data to produce the archive data.

Lilith: Lilith is a suite of software that generates synthetic flight-like data for TESS with a high degree of fidelity. It is based on the *Kepler* End-to-End model (ETEM) used to develop and test the entire *Kepler* ground-segment, in particular the science pipeline, which required high-fidelity simulated data. ETEM was indispensable in testing the entire *Kepler* ground segment as well as for designing, implementing, and testing the KSOC. Lilith plays a similar role for TESS. Lilith simulates the astrophysics of planetary transits, stellar variability, background and foreground eclipsing binaries, artifacts such as sudden pixel sensitivity dropouts (SPSDs), cosmic rays, and other phenomena.

Models (MOD): This module consists of a set of database tables, persistence classes, and associated handling code that determine the correct values of a variety of instrumental effects for the purposes of data calibration at any time during the mission. The models are also used for managing the targets. These include the 2-D bias voltage image, the pixel-level gain model, and the point spread function (PSF). One of the most critical model sets maintained in MOD is the model describing the mapping from celestial coordinates to focal plane coordinates (pixels), called `RaDec2Pix`.

3.2 Science Analysis Pipeline

The science pipeline calibrates the original data from TESS and produces the archival data products. It conducts the transit search and constructs diagnostics used to prioritize and rank the planetary candidates for follow-up observations. Minor to modest changes in the KSOC code are being made to accommodate the differences between *Kepler* data and TESS data.

Calibration (CAL): This module operates on original spacecraft pixel data to remove instrument effects and other artifacts that pollute the data. Traditional CCD data reduction is performed (removal of instrument/detector effects such as bias and dark current and flat field), in addition to pixel-level calibration (correcting for cosmic rays in collateral data and variations in pixel sensitivity), TESS-specific corrections (removing smear signals which result from the lack of a shutter on the cameras), and additional operations that are needed due to the complexity and large volume of flight data. CAL operates on both 2-min science cadence data and FFIs, and produces calibrated pixel flux time series, uncertainties, and other metrics that are used in subsequent pipeline modules.

Compute Optimal Apertures (COA): This module identifies the pixels of interest for extracting photometric measurements from the CCD images for each target star in the SPOC pipeline. It was also used for providing target tables to the flight segment for *Kepler* which had very tight margins for the pixel data stored onboard and returned to the ground: only 32 pixels per star were available, on average. By contrast, TESS allocates 100 pixels per target star, on average, relaxing the requirements on predicted pointing accuracy.

The POC will use *Kepler*'s compute optimal apertures (COA) algorithm, which determines which pixels need to be used for extracting the photometry and astrometry from each target star's "postage stamp." COA relies

on a high fidelity model of the photometer and the spacecraft attitude control system, along with a complete catalog of the target stars and field stars provided by the TESS Input Catalog and updates therein.

Photometric Analysis (PA): This module measures the brightness of the image of each target star on each frame. It also fits and removes background flux due to zodiacal light and the diffuse stellar background, identifies and removes cosmic rays all target star apertures (if enabled), and measures the photocenter or centroid of each target star on each frame. PA also uses PSF fitting to measure precisely the location of $\sim$200 bright, unsaturated target stars on each CCD readout area in order to establish the pointing and focus of each camera.

Presearch Data Conditioning (PDC): This component performs an important set of corrections to the light curves produced by PA, including the identification and removal of instrumental signatures caused by changes in focus or pointing, for example, and step discontinuities that result occasionally from radiation events in the CCD detectors. PDC also identifies and removes isolated outliers and corrects the flux time series for the fact that a) not all the flux in the photometric aperture is due to the target star (i.e., crowding), and b) the photometric aperture does not contain all the flux from each target star.

Transiting Planet Search (TPS): This component implements a wavelet-based, adaptive matched filter algorithm to detect signatures of transiting planets. TPS stitches the $\sim$28-day light curves together for stars observed on consecutive sectors prior to searching for planets. TPS also provides estimates of photometric precision, a key performance diagnostic for transit survey missions, on timescales of transits. The metric used for both the TESS and *Kepler* missions is the combined differential photometric precision (CDPP).[7]

Data Validation (DV): This component performs a suite of diagnostic tests on each transiting planet signature identified by TPS to make or break confidence in its planetary nature. These include a comparison of the depth of the even transits to the odd transits, an examination of the correlation of changes in the photocenter (centroid) of the target star to the photometric transit signature, a statistical bootstrap to assess confidence in the detection, difference image centroiding to rule out background sources of confusion, and a ghost diagnostic test to rule out optical ghosts of bright eclipsing binaries as the source of the transit-like features. These tests can determine if the transit signature is likely to be due to a background eclipsing binary whose diluted eclipses are masquerading as transits of a planetary body. DV also calls TPS to search the residual light curve for evidence of additional transiting bodies after fitting and removing the first planetary transit signature from the light curve. This process is repeated until TPS fails to identify another transit signature.

3.3 Commissioning Tools

Four of the *Kepler* commissioning tools are relevant to TESS and have been provided to the POC as source code.

Focal Plane Geometry (FPG) and **Pixel Response Function (PRF)**: FPG and PRF were used to determine the detailed sky to pixel mapping coefficients and the shape of the PSF, respectively, across each of the 84 individual CCD readout channels. The FPG coefficients include terms for pincushion distortion. PRF constructed five individual pixel response functions for each CCD readout area in order to capture non-uniformity in the focus and PSF. PRF models at intermediate locations are obtained by interpolation. The pipeline uses these model waveforms to monitor the locations of the brightest, unsaturated 200 stars on each channel to reconstruct pointing and capture distortion due to focus changes. Both the POC and the SPOC will construct FPG and PRF models to reduce the risk attendant in developing these crucial models during the time critical commissioning phase.

Pixel Overlay On FFIs (POOF): This tool allows the user to retrieve *Kepler* FFIs and overlay the aperture masks from target tables on the images, along with information about the stellar targets themselves. This enabled the validation of target tables early in the *Kepler* Mission and is a tool that is still used to examine the FFIs to diagnose issues with time series pixel data.

Data Goodness (DG): The DG tool allows the user to examine diagnostic statistics on *Kepler* FFIs collected during commissioning that were collected, but not fully analyzed during the commissioning. This tool is still in use to verify the quality of *Kepler* FFIs collected during normal science operations.

3.4 Hardware

3.4.1 Cluster Worker Machine

Cluster worker machines are used to serialize data inputs for MATLAB executables that run on Pleiades and to parse the outputs generated from those pipeline modules. Worker machines can run these algorithms locally, but at a much smaller scale. This is done for less algorithmically intensive modules such as PPA and AR where the additional complexity of running on Pleiades is not warranted.

Cluster worker machines can be moved between different clusters in order to provide for some redundancy. We are procuring worker machines with 24 cores (48 including hyper threads) and 768 GiB of RAM. Temporary task file storage will be on NFS or directly attached via a storage area network (SAN). This means local storage on the worker machine is not a bottleneck for either storage capacity, performance or availability. Two 10 GiB Ethernet network interfaces will be present on the worker machine.

Analysis indicates that one worker machine is sufficient for the SPOC processing needs. Additional workers may be purchased to use as fail-over machines as worker machines fail.

3.5 Cluster Datastore Machine

Each cluster also has a dedicated datastore machine that has a similar specification to the cluster worker machines with the addition of two 8 GiB fibre channel host bus adapters. PostgreSQL and ADB will share this machine. The *Kepler* SAN has a theoretical limit of 2 GiB sec^{-1}, which is sufficient to copy the entire contents of the storage array in about 10 minutes. Practically, other parts of the architecture are the limiting factor. For *Kepler*, the SAN is often not the bottle neck as the number of concurrent I/O operations is limited by the number of disks rather than by the network.

ADB will be allocated 64 GiB of RAM with PostgreSQL using the remainder. The version of ADB used by *Kepler* uses most of its memory to cache b-tree indices and for temporary buffers. Index blocks are used to locate array blocks on disk; this scales with the number of independent arrays in working memory which is typically largest during processing for CAL. With TESS ADB we will have fewer arrays at any one time, but they will be larger so buffers will be larger. Also, the ADB allocator will be finer grained so there will be multiple entries per array and the total memory required will be about $5\times$ *Kepler* (32 GiB) or 160 GiB. PostgreSQL tends to be bottle necked on disk I/O rather than processor power. Compared with *Kepler*, PostgreSQL will have three times the CPU resources and eight times the memory available.

Kepler used Oracle instead of PostgreSQL. While Oracle performed well for *Kepler*, it also requires a small number of certified database administrators to order to keep it up and running. As a lower-cost solution, we are using PostgreSQL, which is a very reliable open source database, and requires less down time for patch updates.

3.5.1 Data Storage

Data storage is handled via a SAN. This is a dedicated network for the transmission of blocks of data to and from datastore machines. We will use a storage array with $\sim$200 7.2k RPM hard disks. The storage array presents the view of one or more virtual block devices to each host known as volumes or LUNs. Each volume is in fact a combination of disks placed in a RAID 6+0 configuration. This allows for each LUN to be striped across all the drives in the array and so each can access the full number of I/O operations. The storage array can provide approximately 15k I/O operations per second. A volume can also be snapshot which is a point-in-time copy of a base volume. Modifications to snapshots have a copy-on-write policy which means space is only allocated for modifications. Failed disks can be replaced with reserve space on the remaining working disks. At a minimum, two disks can fail without a loss of data. In practice, many more disks can fail without data loss.

The assumption here is that TESS I/O characteristics will be somewhat better than *Kepler*'s in that reads will be longer and thus highly contiguous. Linear, contiguous reads are closer to the ideal of spinning disks. In practice, disk work loads are difficult to predict and so the exact requirements for our storage system won't be known until testing is completed. We expect that I/O will be no more than eight hour for processing FFIs and significantly less for cadence data.

4. ESTIMATING THE SOFTWARE DEVELOPMENT COST OF THE SPOC

We developed a plan for providing the science processing facility for TESS during the original Explorer Program proposal phase and updated it during the concept study phase. This section describes our cost estimation approach and results.

4.1 Heritage Design Basis

The TESS Mission benefits from heritage software and lessons learned from the *Kepler* Mission and by using the *Kepler* science pipeline as the basis for TESS data reduction. By the time of TESS' launch, the *Kepler* team will have operated the data reduction pipeline for over eight years. Leveraging the KSOC design and development process, as well as actual code results in significant schedule and cost savings for the TESS project and presents a lower risk approach to this challenging aspect of the mission. In addition, the *Kepler* codebase that serves as the starting point for the TESS data reduction pipeline is extremely mature and well understood.

The manner in which TESS collects data is very similar to that of *Kepler*. Both TESS and *Kepler* sum images onboard the spacecraft and return postage stamps containing the images of target stars along with collateral data necessary to calibrate the data and allow for a full analysis on the ground. Both missions require that the pixel level data be calibrated to remove on-chip artifacts, that photometry and astrometry be extracted from the calibrated pixels, that instrumental signatures in the light curves be identified and removed, that the calibrated light curves be searched for signatures of transiting planets, and that identified planetary signatures be assessed automatically for their credibility in order to remove astrophysical false positives and other false alarms (artifacts). Both missions require the performance of the flight instrument to be monitored to track and trend performance. The high data rates for both missions, and the complexity of the data reduction process requires a sophisticated software infrastructure in order to process the data in parallel on a computer cluster, and to manage the data sets so that the processing history and provenance of each data set is maintained. A project of this size requires a formal, full lifecycle software development approach compliant with NPR 7150.[‡]

Kepler KSOC Heritage for the TESS SPOC: All software components are based on the design and codebase of the KSOC. The use and operational environment for the SPOC is identical to the KSOC. All changes are being made to "make play" and are required in order to support TESS' functional requirements. The effort represented on Table 2 only includes software developer's time, and is given in person-months (FTEM). The percent re-use based on software lines of code (SLOC) is given in percent for the Java and for the MATLAB software codebases.

4.2 Difference Between *Kepler* Baseline and Proposed TESS Design

We estimate that more than 67% of the *Kepler* pipeline infrastructure codebase, written in Java, is re-purposed for TESS. The scientific computations are performed in compiled MATLAB, which greatly simplifies the task of identifying and fixing code defects and anomalies since the code can be run interactively to isolate the problem. Over 74% of *Kepler* MATLAB code is re-used for TESS. Most code changes are related to the differences in the data rates and focal plane organization (e.g., number of cameras and CCDs, CCD format) between *Kepler* and TESS.

The effort required to retool the KSOC pipeline to support TESS is a small fraction of the cost to develop the KSOC, and risk is lowered given that SPOC development personnel are drawn primarily from the KSOC. The approach used to assess the heritage of the KSOC for TESS is as follows:

1. The SOC 6.2 codebase released in August 2010 was selected as the baseline for establishing costs as this was the first KSOC release that contained the full functionality required by TESS.

2. The original development cost per SLOC was established by tracing the FTEs that participated in the full lifetime cycle of design and implementation of the KSOC codebase through to the delivery of the SOC 6.2 release. Since the KSOC pipeline consists of Java software for the infrastructure elements and MATLAB code for the science processing, the software and associated costs were computed separately for Java and for MATLAB.

[‡]See http://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7150&s=2B.

3. The percent re-use was estimated for each software component by mapping of TESS functional requirements onto the *Kepler* functionality and identifying modifications required to support TESS. This resulted in estimates for the number of lines of code of Java and of MATLAB in each pipeline component that would need to be re-worked based on TESS requirements.

4. Efficiency factors (0.5 for Java, 0.75 for MATLAB) were adopted given that a) the TESS pipeline would not be built from scratch, b) existing *Kepler* SOC team members would be employed in this effort, and c) lessons learned from the *Kepler* development effort as well as re-use of design would significantly simplify the process. The different factors were chosen for Java and MATLAB principally because the Java code changes are likely to be straightforward based on changes to ICDs and data formats. The MATLAB code changes will likely be due to different behavior or performance characteristics of the TESS photometer and detectors, and will require more design and development time to identify the changes needed to support TESS. These factors do represent adoption of moderate risk for the estimated efforts and are at the 70% confidence level.

5. The effort to update and modify procedures and processes to support TESS based upon those of *Kepler* was estimated and used to assess the support needed by operations engineering staff.

6. The hardware required for the SPOC was estimated assuming that TESS would primarily rely on Pleiades for standard operations, just as the KSOC has relied on Pleiades for almost all science data processing throughout the *Kepler* extended mission.

This historical analysis resulted in an estimate of the development effort required for both Java (pipeline infrastructure) and MATLAB (scientific computing) when aggregated across all the components, listed in Table 2. The total software development effort for the KSOC through SOC 6.2 included 479 and 455 FTE-months for the Java and MATLAB codebases, respectively. In contrast, the TESS SPOC software development effort will take only 14% of that required for *Kepler*.

Table 2. SPOC Subsystem Heritage Design Basis Table.

Subsystem/ Component	Java % Re-Use	Java Effort (FTE-months)	MATLAB % Re-Use	MATLAB Effort (FTE-months)
PF	80	5.9	n/a	n/a
DR	50	4.1	n/a	n/a
FS	50	12.0	n/a	n/a
AR	60	5.8	50	1.4
SLIB	76	20.0	90	2.8
COA	50	3.5	70	4.9
CM	54	7.8	80	1
PDQ	100	0.0	100	0
PPA	70	2.0	90	4.1
Lilith	50	3.7	80	12
MOD	50	4.9	50	3.9
CAL	50	2.1	70	7.9
PA	70	0.9	75	5.7
PDC	80	0.4	70	3.8
TPS	80	0.3	80	1.7
DV	80	0.9	80	12.0
PRF	50	0.6	90	3.0
FPG	50	0.6	90	1.0
DG	100	0	100	0
POOF	100	0	100	0
Total Effort		75.9		65.2

5. RUNNING THE SPOC PIPELINE

The SPOC science processing pipeline is actually configured as multiple pipeline segments based on the dataset types that they process and the frequency at which they run, as indicated in the typical pipelines depicted in Fig. 4. Each pipeline is a directed graph of pipeline modules.
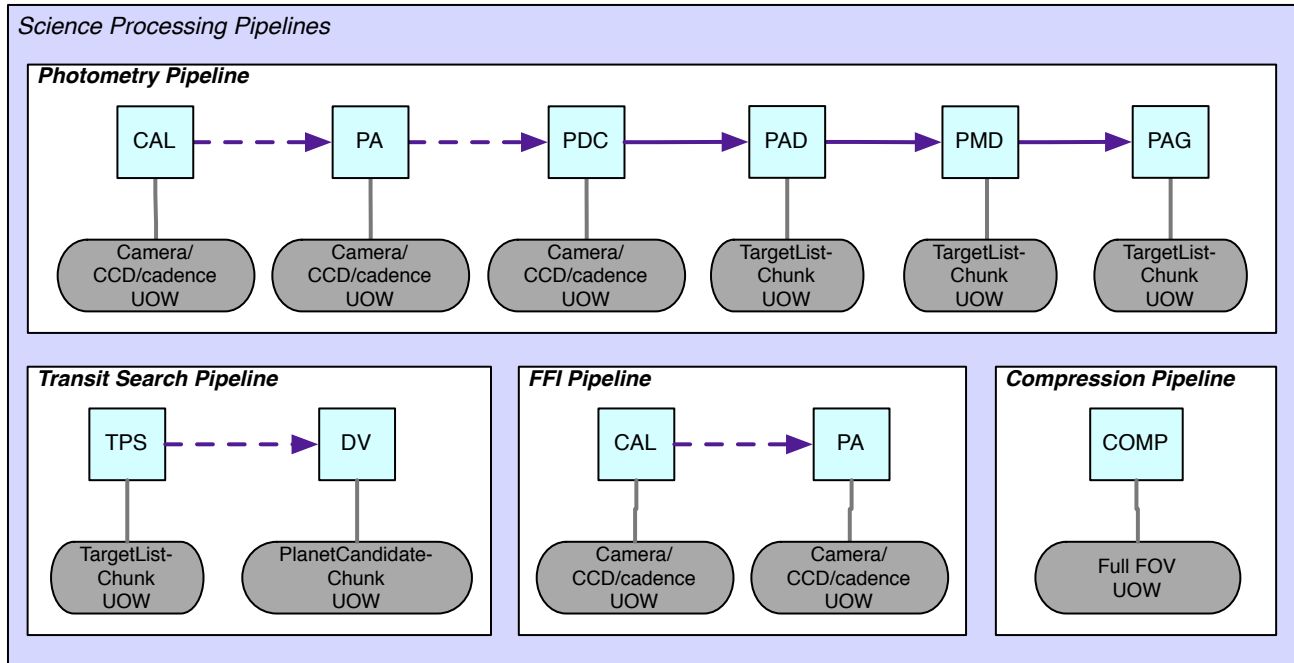

Figure 4. There are four major pipelines relevant to TESS science operations at the SPOC. Top panel: The photometry pipeline for the 2-min data. Left bottom panel: The transit search pipeline for the 2-min data. Middle bottom panel: The FFI pipeline. Right bottom panel: the compression table generation pipeline. See Sec. 6 for details about these pipelines.

Organizing these processing steps as separate pipelines provides flexibility without complicating the pipeline configuration. This flexibility also allows other scenarios to be implemented. The set of pipeline segments and the modules which they contain are completely configurable using the pipeline GUI. The set of available modules (the module library) is also configurable. This architecture enables modules to be easily updated, added, or removed without code changes (other than to the modules themselves).

5.1 Unit of Work

Because the pipeline algorithms can be very computationally intensive, and because of the large data volumes involved, the pipeline will be run on multiple machines. To this end, the pipeline can run on a cluster of worker machines or a set of remote machines. Specifically, the set of remote machines we use are the Pleiades supercomputer. When a new pipeline is launched, the work must be divided into units that can then be distributed to individual worker machines. This unit-of-work can be configured to bin the input data by cadence, CCD output, and/or targets. The design goal is to use the unit of work as a tuning knob to maximize concurrency. This knob is adjusted based on how many machines are available and how long a unit of work takes to process.

For execution on Pleiades, a unit-of-work is further broken into sub-tasks so that work can be distributed across the many nodes and cores in Pleiades. Additional tuning parameters control how many sub-tasks can execute on each node so as to best take advantage of the memory and cores present in each node. Table 3 lists units of work and subtasks for several pipeline modules.

Table 3. Units of work to the subtask level for several different pipeline modules.

Pipeline Module	Binned by
CAL	cadence interval, observing sector, CCD output, CCD row(s)
PA	observing sector, CCD, individual targets
PDC	observing sector, CCD
TPS	one or more observing sector(s), individual targets
DV	one or more observing sector(s), individual targets

5.2 Coordination of Work

5.2.1 Local Cluster

When a new pipeline is launched, a pipeline task is created for each unit of work for each module in the pipeline. Pipeline tasks are scheduled for asynchronous execution using a distributed message queue also known as Message Oriented Middleware (MOM). At the start of execution, a message is placed on the queue for each pipeline task. Once the messages are on the MOM queue, the next available worker machine will pull the next message off the MOM queue and execute the pipeline task corresponding to that message. Any worker is able to execute any pipeline task because each worker machine has access to all of the science modules and the pipeline infrastructure services. This design allows worker machines to be easily added for increased processing throughput.
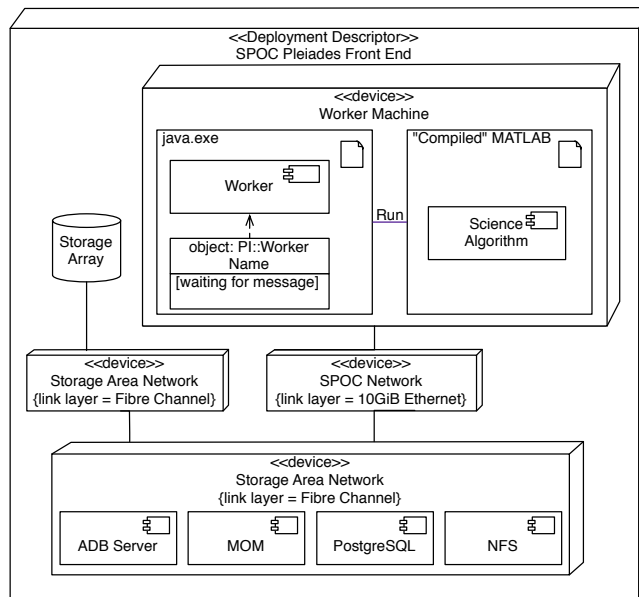

Figure 5. Pipeline deployed in a local cluster. Multiple worker machines and PI::Worker instances may be used in this configuration to scale up processing to 10s of independent tasks.

5.3 Remote Execution

The most distinguishing feature of remote execution over local execution is the use of third-party authentication and connection tools. This is used to execute pipeline tasks on Pleiades. In this deployment case the pipeline worker processes remain local and files are transmitted over secure shell (ssh) or secure network file system (NFS). A remote queuing system (RMOM), a kind of second level MOM, allocates super computer nodes to subtasks. Pipeline modules can generate a dependency graph that expresses the dependencies between subtasks. The RMOM obeys this dependency graph and so as many independent subtasks can run on at least as many available processing nodes (Fig. 6). There are additional parameters that determine the number of concurrent subtasks that can execute on a processing node. This is usually limited by the memory-to-core ratio of the type of subtask being executed. While not limited to coordinating MATLAB processes, these are the types of processes that are executed on the supercomputer nodes.
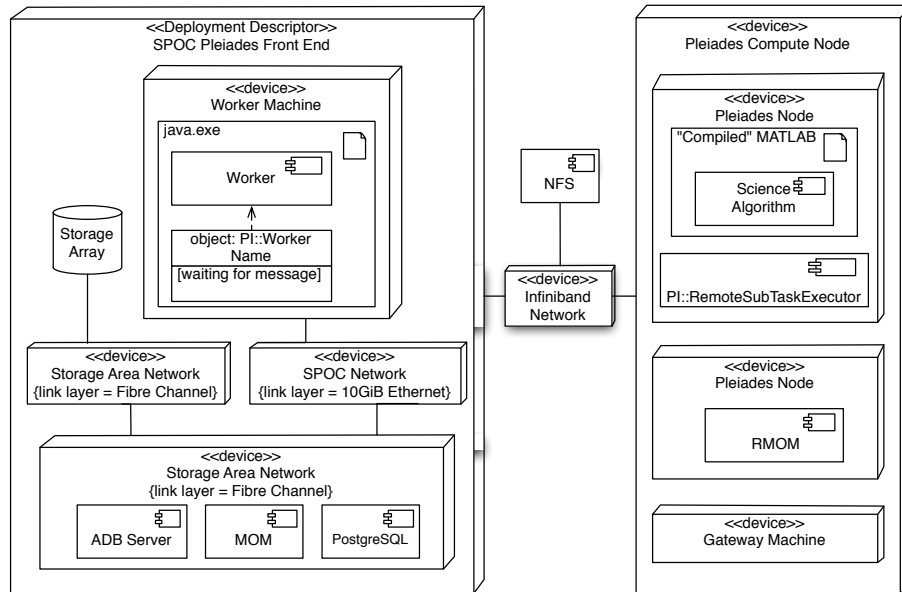
Figure 6. Pipeline deployed on Pleiades. In this deployment, the local cluster is used to generate inputs and outputs for sub-tasks on Pleiades. Remote processes manage the execution of science algorithm implementations. This can scale up to tens of thousands of independent sub-tasks.

5.4 Triggers

New pipeline instances are launched using pipeline triggers. These triggers are part of the pipeline configuration and are created by the pipeline operator using the pipeline GUI. Triggers can also be used to launch pipelines manually, as in the case of reprocessing, or automatically on a particular schedule or when the input data become available. These data-available triggers allow the various pipeline types to be chained together so that complete, end-to-end science processing can be automated. For example, the photometry pipeline can be configured to run when new data are delivered from the spacecraft.

5.5 Data accountability

Data accountability is a crosscutting feature of the pipeline infrastructure and is baked into the pipeline infrastructure, datastore, data receipt, and each science pipeline module. Tracked data are assigned a unique originator ID that determines its origin. The pipeline infrastructure manages the sets of parameters used for each pipeline module. These parameter sets are locked into an immutable state once a pipeline trigger has been fired. So by using pipeline instance IDs it's possible to backtrack through the data accountability trail to the actual parameter values used to process any piece of data.

6. OPERATING THE SPOC PIPELINE

In this section we describe the science data processing from an operations standpoint. The 2-min cadence data and FFIs for each orbit of a given sector are downlinked to the DSN when the spacecraft is near perigee and transferred to the POC, which packages the data and places it in a staging folder. The Data Receipt process polls the POC staging folder for new data continuously. When DR detects that new data have appeared, it pulls the files over to the SPOC for ingest into the data store and science processing. The processing occurs in two stages for nominal science data to produce the light curve products: single-orbit processing, and single sector processing. The transit search is conducted after the sector processing is complete. FFIs are processed separately. After the processing is complete, the archival data products are reviewed for quality and then delivered to the POC for distribution to the TSO and MAST, and for the final project-level review and compilation of data release notes.

One of the overarching, driving requirements for the SPOC is to keep up with the deluge of data downlinked from the spacecraft every ∼2 weeks. This means that each processing task must take no longer than ∼27.4 days,

and that there must be an adequate number of independent processing environments to allow for simultaneous processing of completed sectors, as well as new, incoming data from the next FOV. Fig. 7 illustrates how the SPOC will manage the computational environments in order to meet the critical 27.4-day processing throughput requirement. All original data from the spacecraft, input models, parameters, and configuration settings are stored in the datastore on the primary cluster, OPS1, hosted on an HP 3Par storage system. OPS1 is used for all archival processing activities. The image of the OPS1 datastore is "snapshotted" onto the secondary cluster, OPS2, which is used for preliminary processing of each orbital data set. Snapshotting is a very time-efficient process, as the data are only written to new disk space when write commands are executed. In this way, the task of copying the datastore image occurs as needed concurrent with the processing activities on the target cluster. The OPS1 datastore image is also snapshotted periodically to a third cluster, TEST, which is used for test, integration, and V&V activities as the SPOC codebase is updated during science operations to accommodate changes in spacecraft and/or instrument behavior. Figure 8 traces the SPOC operations activities from data
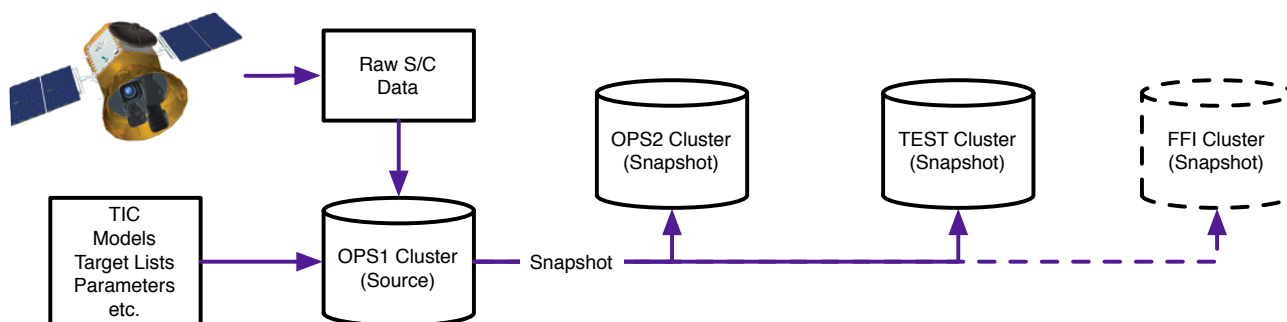

Figure 7. The data processing tasks are distributed over two operations clusters, OPS1 and OPS2, with a third cluster, TEST, used for test activities. All original spacecraft science and engineering data, models, algorithm parameters and configuration settings used for archival tasks are stored in OPS1. These are snapshotted (copy on write) to OPS2, used for preliminary processing activities, and to TEST. Also shown here is an optional cluster devoted specifically to processing FFIs that the TESS Project may elect to support in order to minimize the latency in processing and delivery of the calibrated FFIs.

receipt to the delivery of the archival data products to the TSO and MAST via the POC. The SPOC operations staff configures each pipeline run and then makes the resulting data available to the Data Analysis Working Group (DAWG) for review and manages the export of the archival data products to archival format and transfer to the TSO and MAST via the POC. The DAWG reviews preliminary processing products to identify issues with the data products or instrument or spacecraft environment, and to update parameters for archival runs. The DAWG reviews the archival products prior to export to aid in the compilation of data release notes. Once the data reach the POC, representatives from the TSO, the SPOC and the POC review the quality of the data products prior to their release to the archive, and assemble data release notes describing the character and quality of the data, highlighting issues peculiar to each data set, such as loss of fine point, coronal mass ejections (CMEs), and Earth or Moon eclipses. The pipeline processing activities are detailed below.

Figure 9 depicts the processing timeline. Data from a single orbit are processed as soon as they are available, regardless of how complete they are, in order to assess the preliminary data quality and health and status of the instrument, and to set up parameters for the final (archival) processing. This processing takes place in the OPS2 cluster using a snapshot of the data ingested into the primary OPS1 cluster. FFIs are processed at this time. Any missing data for the sector will be downlinked and transferred to the SPOC during the next downlink opportunity ~2 weeks later. At this time, the full sector can be processed on the OPS1 cluster to produce the archival light curve products as well as to run the planet detection pipeline. Multi-sector planet searches are conducted for targets that have been observed for previous consecutive sectors in addition to the current sector. Following the sector and multi-sector processing, the light curve, centroid, and planet search products are reviewed internally while they are written to the archive formats and then exported to the POC for final project review and for distribution to the MAST and TSO. The following sections provide further details of the processing.
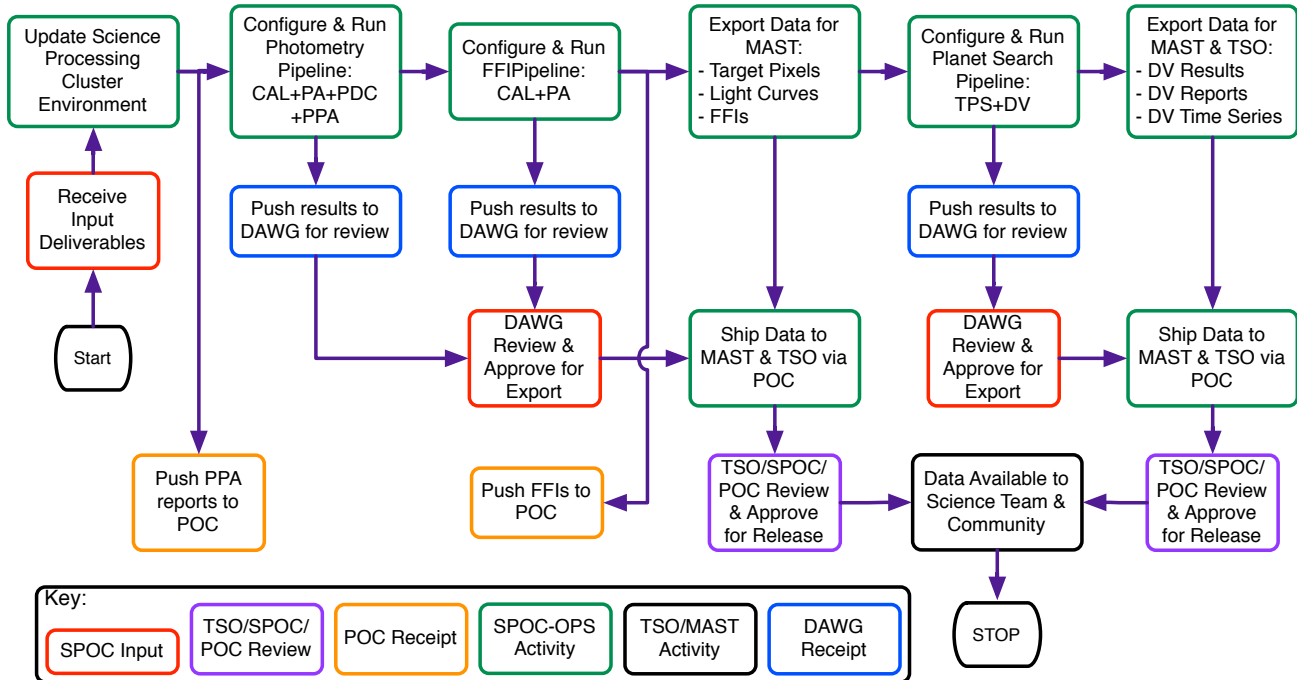
Figure 8. The SPOC science processing operations activities. This flow chart traces the data received from the POC through the nominal science data and FFI processing to deliver archive products to the MAST and TSO. The flow chart includes the data review activities, as well as the operations activities at the SPOC, and those of the TSO, MAST and POC relative to science data processing.
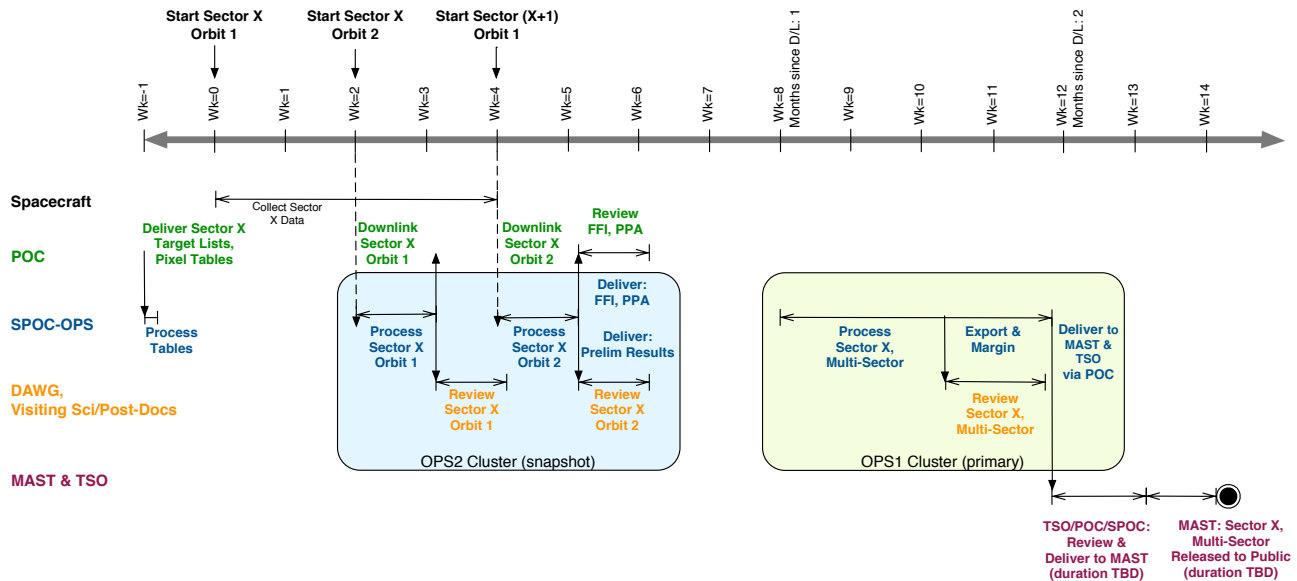

Figure 9. The science processing timeline is illustrated in this figure.

6.1 Single-Orbit (Half-Sector) Processing

At the end of each orbit, the MOC downlinks data from the just completed ~13-day orbit as well as any data that failed to come down during previous downlinks, which it then transfers to the POC. The raw data are decompressed at the POC, packaged into google protocol buffers, and staged for the SPOC to pull over the network. We assume that it will take ~5 days for the data to arrive at the SPOC. The SPOC processes the single-orbit data to identify data anomalies and allow for adjustment of algorithm parameters for the full

sector (archive) processing. This processing also enables the SPOC to monitor the instrument and spacecraft performance as reflected in the quality of the science data and the compression performance. The resulting performance reports are reviewed by the DAWG and are also delivered to the POC. The SPOC single-orbit (half-sector) processing includes Cadence and FFI processing downlinked at the end of each orbit, regardless of gaps in the single-orbit of data, along with the supporting spacecraft engineering data.

6.1.1 Process Cadence data through the Photometry Pipeline

Upon reception of a single-orbit of Cadence data, operations personnel configure the pipeline for processing the data, which includes specifying the range of data and the targets to process.

The single-orbit Cadence data are processed through the Photometry Pipeline, as depicted in Fig. 4. The data are first calibrated by CAL, which corrects for bias/dark, gain/linearity, and other on-chip artifacts. The pipeline then runs PA on the calibrated data, generating photometric results (light curves and centroids) for each target. PDC is run next, performing systematic error correction on the light curves and identifying and removing outliers. PDC calls TPS functions internally to generate photometric precision metrics to use in PPA for monitoring instrument performance. Finally, the PPA generates and tracks and trends a number of performance metrics including background, black/bias, and dark levels, dynamic range, brightness, stellar centroids, high-resolution camera attitudes, photometric precision, and compression performance. We assume that the single-orbit pipeline will take 5.7–8.9 days for 13 days of Cadence data. This time estimate (and all those that follow) are projections based on actual *Kepler* processing experience on Pleiades, scaled to TESS data volumes. The lower limit is the expected value with no contingencies. The upper value includes margin.

Once these activities are complete, the DAWG (including visiting scientists and post-docs) will review the results of the pipeline and the instrument performance. The PPA reports are passed to the POC for review of the instrument performance.

6.1.2 Process FFI Data

FFIs provide a record of the full focal plane over time that can enable a wealth of complementary science by the community, as well as be used for long-baseline analysis of the instrument performance and in associated data correction/calibration/false positive identification algorithms.

The CAL portion of the FFI pipeline (see Fig. 4) performs the same calibrations as it does in the Photometry Pipeline. The PA portion of the FFI pipeline calculates the motion polynomials of specific/pre-defined (bright unsaturated) stars in order to support generation of the World Coordinate System (WCS) coordinate coefficients by the FFI exporter. We are currently assuming that processing through the FFI pipeline will take 1.5–2.3 days for 13 days of FFI data. Once these activities are complete, the DAWG (including visiting scientists and post-docs) will review the results.

6.2 Single-Sector (Two-Orbit) Processing

Single-sector processing includes Cadence and FFI processing with the most up to date models, parameters, and other configuration changes for the purpose of delivery to the MAST Archive, TSO and POC. Once all the Cadence and FFI data have been downlinked and delivered to the SPOC, the data will be processed through the nominal single-sector pipeline. Any data not downlinked at the end of an orbit will be re-transmitted at the next orbit. Due to potential gaps in the downlink of the second orbit of data, it may take one full orbit after the completion of the science data collection for the SPOC to receive the full sector of data. If there are gaps, then the single-sector processing will wait until all data has been downlinked and delivered to the SPOC.

6.2.1 Process Cadence Data through the Photometry Pipeline

While the pre-processing of the final orbit of raw data is in progress at the POC, the SPOC operations and science personnel prepare the SPOC science pipeline system for data ingest and subsequent processing.

Once all Cadence data have been delivered and ingested at the SPOC, they are processed through the same Photometry Pipeline as the single-orbit data, with a few configuration updates. For example, some pipeline parameters are adjusted to optimize the software performance when working on a full sector's worth of data. Data anomaly flags are a way to identify cadences that the science pipeline should process differently from its

defaults, such as when the spacecraft is not in fine point or when it experiences anomalous behavior due to external events such as CMEs. The data anomaly flags are updated as needed for the single-sector processing. We assume that the photometric pipeline processing will take 11.4–16.5 days for 27 days of Cadence data.

6.2.2 Process FFI data

As with the single-orbit processing activity, the FFIs are downlinked by the POC and delivered to the SPOC every ~13 days. The FFIs are processed through the FFI pipeline as depicted in Fig. 4. The FFIs are processed through CAL and then a subset of bright, unsaturated stars are processed through PA to reconstruct the pointing and populate WCS coordinate information in the exported files. If new SPOC software has been deployed to production since the first-orbit FFIs were processed, those FFIs will be re-processed with the new software to produce a uniformly processed set of FFIs for the full sector. We assume that the FFI processing will take 2.4–3.4 days for 27 days of data. Upon completion of processing the full sector of FFIs, the FFIs will be exported and delivered to the POC for archiving at MAST.

6.3 Transiting Planet Search Processing

After each sector of data is processed through the Photometry Pipeline (Sec. 6.1.1), the Planet Search pipeline can be run. As the Observatory slews from one sector to the next, a set of ~5,000 targets will be remain in the field of view, generating a sub-set of targets that are multi-sector targets. The light curves for each sector for these multi-sector targets are collected and stitched together by TPS prior to searching for planets. The transit search identifies transit-like features called Threshold Crossing Events (TCEs). These TCEs are then passed to the DV, which fits a limb-darkened transit model to each TCE, and performs a suite of validation tests on each TCE. DV also calls TPS to identify additional TCEs in each light curve. We assume that the Planet Search pipeline processing will take 0.4–0.7 days for single-sector data, and 1.9–3.1 days for the 13-sector targets. The single sector targets and multi-sector targets may be run separately. Once the single and multi-sector targets have been processed, the products are exported for delivery to the TSO and MAST via the POC.

6.4 Archive products to the MAST and TSO

Upon completion of the single-sector or multi-sector data processing, the DAWG will review and approve the data for export. The SPOC will then export the data for archive at the MAST and delivery to the TSO. The SPOC will deliver the photometric, transit search, and supplemental products to the MAST for each single-sector and/or multi-sector data set. The first four sectors of science data will be delivered within six months of the beginning of science operations. Thereafter, the SPOC will deliver one sector of data to the MAST per month. The SPOC will deliver the Planet Search pipeline results of each single-sector and/or multi-sector data set to the TSO via the POC monthly.

6.4.1 Export Single-Sector products for the MAST archive

SPOC operators export the single-sector target pixel and light curve data, FFI data, and other supporting pixel products via the various export tools. The photometry pipeline export products include:

1. Target Pixel files: Raw and calibrated pixel data

2. Target Light Curve files: Photometric analysis and systematic error-corrected time series data

3. FFI files: Raw, calibrated, and uncertainty images of the CCDs

4. Collateral Pixel files - Raw and calibrated collateral pixel data

Once the exports are completed, the SPOC validates the export files and generates the associated data notification messages. The SPOC Science Lead approves the export prior to pushing the data to the POC via secure file transfer protocol (SFTP). This archiving process should take 0.7–0.9 days to complete. The single-sector photometric data products are delivered to the POC for final review and distribution to the TSO and MAST at monthly intervals.

6.4.2 Export Planet Search products for the MAST archive and TSO

Once the transit search pipeline has been run on single-sector and/or multi-sector targets, SPOC operators export the DV products. These include:

1. DV results files: TCE ephemeris and other results from DV, one per target

2. DV full reports: Multi-page reports containing data and plots, one per target

3. DV summary reports: Single-page reports containing plots and data on each TCE, one per planet candidate

4. DV Time Series data: Several types of flux time series data produced by DV, one per target

Exporting the planet search results follows the same procedure as for exporting single sector data products above (Sec. 6.4.1). This archiving process should take 0.5–0.6 days to complete for single sector data, and 1.0–1.2 days for multi-sector data. The single-sector transit search data products are delivered to the TSO within ∼27 days of receipt of the full data set at the SPOC, and are delivered to the MAST within four months. The multi-sector transit search data products are delivered to the TSO and MAST at monthly intervals.

6.4.3 Other products for the archive

The SPOC is responsible for delivery of several ancillary files to the MAST. These include files used by the SPOC pipeline as well as supplemental products of the pipeline, such as co-trending basis vectors (CBVs), the set of systematic trends present in the ensemble flux data used to correct instrumental effects in the light curves. These products are reviewed by the DAWG and approved by the SPOC Science Lead as necessary before delivery to the POC via SFTP for archival to MAST.

7. SUMMARY

The TESS Mission cleared its Critical Design Review in August 2015 and was confirmed to proceed with integration and test activities in December 2015. Development is proceeding well across the TESS Ground Segment and the second Ground Segment Integration Test (GSIT) is scheduled for November 2016. This test will validate and verify the photometric pipeline for TESS and the calibrated pixel and light curve products. The final release of the SPOC codebase is expected in April 2017, well in advance of the December 2017 launch date, permitting ample time for SPOC personnel to participate in operations readiness tests and to conduct further stress testing of the hardware and software. The SPOC development has adhered closely to the original proposed costs and schedule, proving the value and validity of the software development costing approach adopted for the proposal effort described in Sec. 4. TESS will enjoy the benefits of a science analysis pipeline based primarily on the time-tested *Kepler* Science Operations Center at a small fraction of the software development cost incurred by *Kepler*. The search for Earth's nearest cousins commences in 18 months. The TESS SPOC will be ready to open NASA's next chapter of exoplanet discovery by identifying the worlds that will be characterized in detail by the James Webb Space Telescope and other ground-based and space-borne telescope assets in the next years and decades to come.

ACKNOWLEDGMENTS

REFERENCES

[1] Ricker, G. R. et al., "Transiting Exoplanet Survey Satellite (TESS)," *JATIS* **1**, 014003 (2015).

[2] Sullivan, P. W. et al., "The Transiting Exoplanet Survey Satellite: Simulations of Planet Detections and Astrophysical False Positives," *Astrophysical Journal* **809**, 77 (2015).

[3] Jenkins, J. M. et al., "Overview of the *Kepler* science processing pipeline," *Astrophysical Journal Letters* **713**, 87 (2010).

[4] Van Cleve, J. E. and Caldwell, D. A., "Kepler instrument handbook," Tech. Rep. KSCI-19033, NASA Ames Research Center (2009).

[5] Gangestad, J. W., Henning, G. A., Persinger, R. R., and Ricker, G. R., "A High Earth, Lunar Resonant Orbit for Lower Cost Space Science Missions," *Adv. Astron. Sci.* **150**, 13 (2014).

[6] Jenkins, J. M. and Dunnuck, J., "The little photometer that could: technical challenges and science results from the Kepler Mission," *Proc. SPIE* **8146**, 814602 (2011).

[7] Jenkins, J. M. et al., "Transiting planet search in the *Kepler* pipeline," *Proc. SPIE* **7740**, 77400D (2010).