Nancy Grace Roman Space Telescope

SSC Support for Roman ROSES 2022
Call for Proposals

Roman Science Support Center at IPAC

February 14, 2022

IPAC
California Institute of Technology

Table of Contents

1 Introduction

1.1 Scope

This document describes the support for community teams and researchers planning to use data from the Nancy Grace Roman Space Telescope as performed by the Science Support Center (SSC) at Caltech/IPAC. In particular, this includes processing of Wide-field spectroscopic and Galactic Bulge Time Domain Survey (GBTDS) data and their resulting data products and support for the CGI technology demonstration observations.

1.2 SSC Overview

Science Operations for the Roman Space Telescope are provided by the Mission Operations Center (MOC; Goddard Space Flight Center), the Science Operations Center (SOC; Space Telescope Science Institute) and the SSC (Caltech/IPAC), which are all part of the Roman Ground System. The SSC is responsible for science data processing and community outreach for Exoplanet Microlensing, and science data processing and community outreach for all spectroscopic data. The SSC is responsible for issuing the calls for Roman science investigation proposals, proposal ingest, the time allocation committee process, and grants management. Finally, the SSC is responsible for Coronagraph Instrument (CGI) observation planning, production of Level-1 CGI data products, the delivery of all CGI data products to the Roman archive, and CGI calibration and CGI outreach.

2 Roman Spectroscopic Data Processing

This section describes the SSC support for the Roman Space Telescope Wide-field spectroscopic data processing and the resulting data products.

2.1 Roles and Responsibilities

The tasks for spectroscopic data processing and analysis are distributed between the SSC, SOC, and the WFI science teams.

The SOC is responsible for standard calibration of all WFI data, producing calibrated individual direct imaging and spectroscopic exposures (Data Level-2), regularized and mosaic image and image data catalogs (Data Level-3 and 4), and archiving of all Roman data.

The SSC is responsible for the Roman Grism-Prism Data Processing System (GDPS). The Grism-Prism Data Processing System (GDPS) is a component of the Roman Science Support Center (SSC) at IPAC. The SSC is responsible for all processing of spectroscopic data beyond Level-2. The SSC will use the images and spectra to identify, extract, calibrate and fit the Roman grism and prism spectra, producing a wide variety of spectroscopic and derived science data products and returning these to the Roman archive. The data processing and the data products are designed to meet the relevant Roman science and mission requirements.

The SSC will be the main interface with the scientific user community for Roman spectroscopic data processing. The SSC will also provide educational and public outreach on Roman wide field spectroscopy. The SSC is responsible for creating and securing the scientific legacy of Roman spectroscopy in the form of reliably reduced, calibrated and well-documented data products.

2.2 Grism-Prism Data Processing System (GDPS)

The GDPS is based on the slitless grism spectroscopic pipeline under development for Euclid, and the SSC is collaborating with the Laboratoire d'Astrophysique de Marseille (LAM) to design and implement the GDPS. The GDPS will be run in an automated fashion on all Roman spectroscopic data.

2.3 Pipeline Processing and Data Flow

After transmission to the ground system, Level-0 Roman spectroscopic data will be transmitted to the SOC, which will process them to Level-1 (uncalibrated exposures including metadata and engineering telemetry) and Level-2 (exposures with detector-level signatures removed). The corresponding image data will be further processed to

Level-3 (image mosaics) and Level-4 (image catalogs) by the SOC. The Level-2 grism and prism data, along with the corresponding Level-3 and Level-4 imaging data, will be retrieved by the SSC for processing to Level-4 in the GDPS. The SSC will transfer all Level-4 spectroscopic data products, relevant calibration reference files, and appropriate documentation to the Roman archive. The operational interfaces and basic data flow for the GDPS are shown in Figure 1. Transmission from Roman to the ground stations is allocated up to 24 hours and transmission from the ground stations to the SOC is allocated an additional 24 hours.

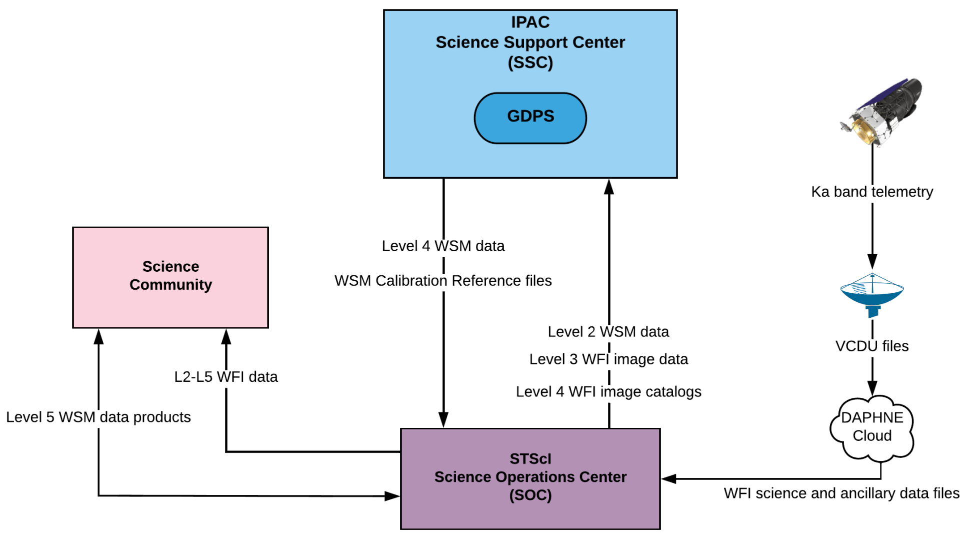

Figure 1: Operational interfaces and basic data flow for the Roman Grism-Prism Data Pipeline System (GDPS). Basic calibrated, Level-2 WFI spectroscopic data, together with Level-3 and 4 WFI image data (image and source detection catalogs) are retrieved by the SSC. Level-4 grism and prism science data products, along with calibration reference files, are delivered to the Roman archive by the SSC. Level-5 products, derived by users from the Level-4 data, will be submitted to the SOC for ingestion into the Roman archive.

2.4 Operational Elements of the GDPS

The GDPS will include a Calibration Data Pipeline (CDP) and a Science Data Pipeline (SDP) for all grism and prism data. The main task of the CDP is to produce the calibration reference files needed to reduce the grism and prism science data in the SDP. The main task of the SDP is to produce the grism and prism science data products, and to evaluate the performance of the WFI grism and prism data with respect to the performance requirements specified in the Roman Science Requirements Document.

The CDP and SDP will be made of a sequence of scripts, each of which invokes one or more independent executable modules that process the calibration and science data. The pipeline will run autonomously and will trigger messages and warnings if errors occur. The GDPS will be developed, tested and supported by scientists and engineers at the SSC, with input from the Roman Project Infrastructure and Wide Field Science teams.

2.4.1 Science Data Pipeline (SDP)

The SDP is made up of two parts. The first part, the G2DP, is designed to calibrate the 2D grism and prism science data and produce 1D spectra for each identified source. The second part, the G1DP, is designed to analyze and extract basic information from the 1D spectra, namely redshifts and emission/absorption line parameters, for each target. The requirements to produce the 2D and 1D spectra, and the basic measured parameters, are common to both the grism and prism data.

The algorithms for data reduction and spectral fitting will be based on those currently under development and testing for use in the Euclid SIR (spectroscopy 2D processing element) and SPE (spectroscopy 1D processing element) functions. The basic functions of the G2DP and G1DP are shown in Figure 2 and described briefly, below.

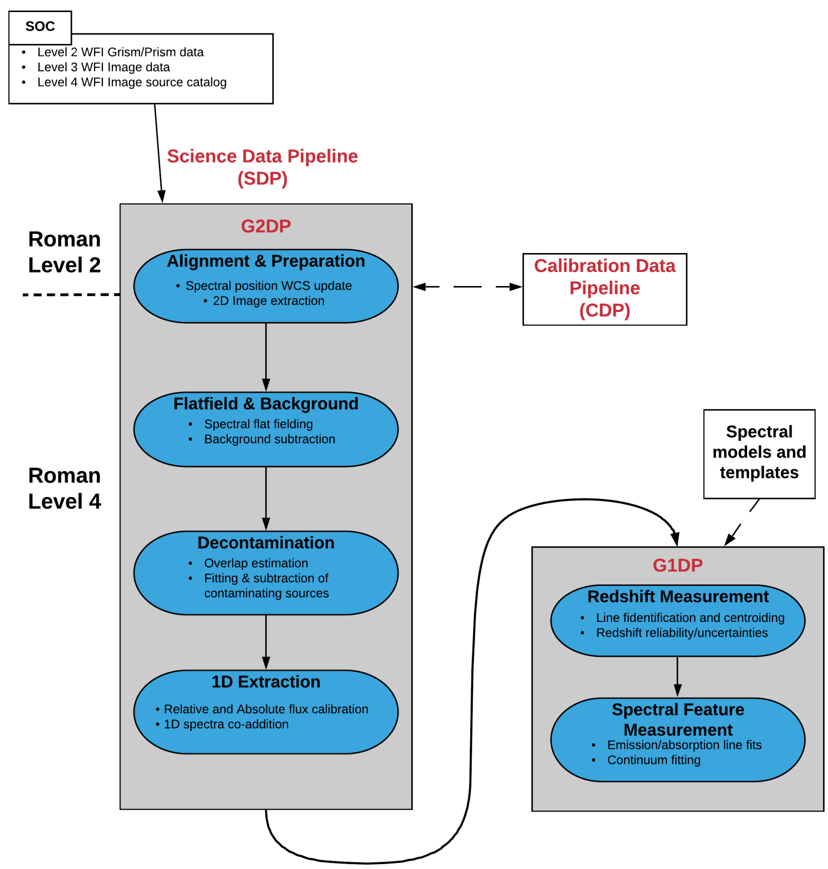

Figure 2: *Basic processing steps of the G2DP (left) and G1DP (right) components of the Science Data Pipeline (SDP). Imaging and basic spectroscopic data products from the SOC enter the G2DP at the top left. Rough divisions between Level-2 and Level-4 Roman data levels for the grism and prism data are indicated on the left. Calibration reference files derived in the Calibration Data Pipeline (CDP) from dedicated calibration observations, are used in the SDP and enter at the upper*

right. The 1D extracted spectra from the G2DP are the primary inputs to the G1DP. Additional reference files (e.g., source type spectral templates) that are inputs for the G1DP, and which are not derived directly from Roman observations, are indicated on the right.

The basic steps of the SDP are:

- Identification and position measurement of all spectra
- Creation of 2D image cutouts for all detected sources
- Spectral flat fielding
- Background identification and subtraction
- Identification and removal of contaminating sources from the 2D spectra
- Creation of 2D cutouts and 1D spectra
- Relative and absolute flux calibration of the 1D spectra
- Fitting of the 1D spectra to produce redshifts and other spectral feature parameters

Level-3 and Level-4 image data products are inputs to the SDP and are used in conjunction with the Level-2 spectroscopic data to identify and extract sources from the grism and prism data. Calibration reference files produced in the CDP are used as inputs to the SDP to calibrate the grism and prism science data.

Roman slitless spectroscopic data will suffer from cross-contamination from sources located close enough in the sky that their spectral footprints overlap. A typical galaxy will be dispersed by the grism across 90 arcseconds of sky and cover more than 4000 pixels. With expected spectroscopic source densities approaching ~105 deg-2 (at H < 22.5 AB mag), overlapping spectra will be common. To illustrate this, a simulated H-band image and corresponding set of grism spectra across ½ of one WFI chip is shown in Figure 3. Spectral images for the prism will be much shorter, but overlap of spectra will still occur and have to be tracked during pipeline data processing.

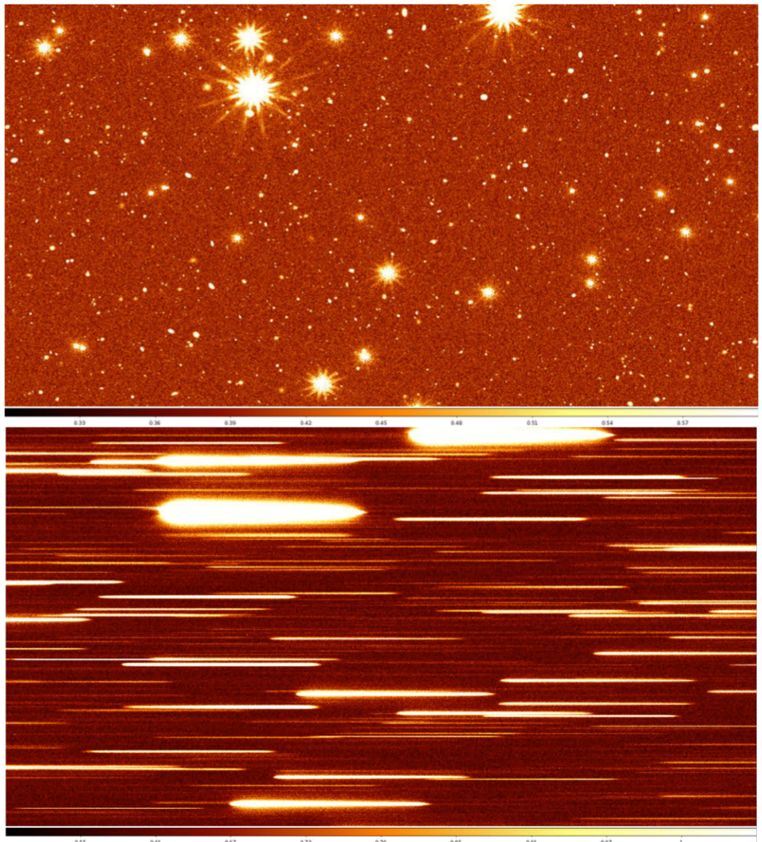

Figure 3: Simulated H-band image and corresponding grism spectra for roughly 28 square arcminutes, equivalent to about half of a single Roman detector. The simulated exposure times are 400 sec for the direct image and 350sec for the grism image. The image contains about 500 objects with H < 22.5 mag AB, roughly the dividing line for significant grism continuum detection in a single 350 sec image. For comparison, we expect roughly 50-100 line emitters with line flux > 1×10^{-16} ergs/cm^2/s in the same region.

The G1DP involves two primary steps - redshift and spectral feature measurement. The redshift measurement is based on a noise weighted least-square fitting of a model to the extracted and calibrated 1D spectra. In addition to redshift, the pipeline returns the best fit model parameters: galaxy template, emission line template, interstellar absorption on the continuum and emission lines (independently) and intergalactic absorption. Errors in the fitting are propagated in all procedures and used to derive an estimate of the error in the measured line characteristics. Estimation of higher order astrophysical parameters (e.g. extinction, star formation rate, star formation history, etc.) will be done by the Wide Field Science teams.

The SDP will be implemented to allow input of test and/or simulated data to calculate the completeness and reliability of the detection and characterization of isolated and overlapping sources in the 2D data, the effectiveness of source decontamination, and the measurement of spectral features and redshifts in the 1D spectra.

2.4.2 Calibration Data Pipeline (CDP)

The CDP produces the reference files used in the SDP to process grism and prism science data. Bright stars, star clusters, and planetary nebulae will be the primary targets observed by Roman for use in the CDP, augmented with observatory calibration sources such as the Relative Calibration System. Ground based measurements will provide the initial set of calibration reference files used by the SDP, updated with these on-orbit measurements. The basic steps of the CDP are the calculation of the direct image to grism/prism mapping, the spectral flat field, wavelength and zero-point calibration, and the relative and absolute flux calibration. These are briefly described below.

It is necessary in the science pipeline to be able to transform from a direct image to a dispersed image for every grism and prism observation. To define spectral extraction boxes for each of the thousands of sources found on each detector, a transformation is made from the objects identified in the direct images to the dispersed spectra. This mapping from direct image to dispersed image is initially determined early in the mission through observations of one or more densely populated stellar fields, where it is possible to identify the coordinates of the blue and/or red edges of the spectral trace of each source and compare it to its direct image equivalent. The CDP will derive the transformation using a sequence of dithered observations of a dense star field.

For wavelength calibration, unlike many grism surveys, Roman cannot use the zeroth-order transmission spectrum as the astrometric basis of the wavelength zero-point (the zeroth order is suppressed by design in Roman grism spectra and is not present at all for the prism). Instead, the sharp blue and/or red edges of the 1st order spectra will be used for this purpose. Using as a starting point a ground-calibrated solution for the dispersion of the first order spectrum, primary in-orbit wavelength calibration observations will be made of bright compact sources and star clusters. The CDP will determine the dispersion solution and wavelength zero point as a function of position in the focal plane. Verification and further refinements of the wavelength solution will also be possible during the mission using samples of bright emission line galaxies with accurately known redshifts derived from high spectral resolution, ground-based observations.

The large-scale illumination field pattern is recovered by making multiple observations of the same sources across the focal plane. Like the small-scale flat field pattern, the relative flux calibration corrections, applied to each extracted source, will be chromatic and will necessitate the creation of a spectral cube to quantify the effect. Because the creation of this relative flux calibration cube is based on many repeated observations of the same stars in one or more fields, this is sometimes called “self-calibration”. Absolute flux calibration will be performed by observing a set of spectro-photometric calibration stars. These spectro-photometric standards will allow us to convert from observed to physical units, producing a 1-dimensional sensitivity function (and associated uncertainty/covariance) that can then be applied in the SDP during spectral extraction.

2.5 Data Product Quality Assessments

The G2DP, G1DP and CDP will include automated data quality assessment (DQA). These assessments will be developed in consultation with the Project Infrastructure Teams. The results of the QA will be evaluated by GDPS science staff at the SSC using algorithms developed by the SSC and LAM. A subset of grism and prism DQA products will be included in the GDPS data products transferred to the Roman Archive.

2.6 Data Volume

Spectroscopic data products delivered to the Roman archive, including Level-4 grism and prism spectra, catalogs, calibration reference files, etc., are estimated to have a volume of about 733 Tb/yr. Roman spectroscopic data will be re-processed through the GDPS when appropriate, based on updated observatory or pipeline performance, or calibration solutions.

2.7 Data Products

A preliminary list of GDPS data products to be delivered to the Roman archive is given below. These products will be produced for all spectroscopic data. The pipeline processing steps for both grism and prism data are equivalent, and therefore the intermediate and archived data products will be similar in structure and format. Most products will be in ASDF format.

- Updated spectra location tables and WCS astrometric information for grism and prism L2 data.
- Background subtracted 2D spectral images.
- 2D decontaminated spectral cutouts.
- Spectral overlap parameters for each extracted source.
- 1D flux and wavelength calibrated spectra (individual and combined) and uncertainties for each extracted source.
- Spectroscopic redshifts, redshift probability density functions, emission/absorption line and continuum parameters, best fit template, and associated uncertainties for all 1D spectra.
- Calibration Reference Files (flats, wavelength solutions, zero points, etc.) derived in CDP and used in the SDP.
- Data Quality Assessment reports and data flags.

3 Exoplanet Microlensing and Galactic Bulge Time Domain Survey Processing

This section describes the SSC support for the Exoplanet Microlensing data processing data from the Galactic Bulge Time Domain Survey (GBTDS) and the resulting data products.

3.1 Roles and Responsibilities

The tasks for the Exoplanet Microlensing data processing are distributed between the SSC, SOC, and the relevant Project Infrastructure Teams (PIT). Figure 4 shows the flow of data to and between these elements of the project. Transmission from Roman to the ground stations is allocated up to 24 hours and transmission from the ground stations to the SOC is allocated an additional 24 hours.

The SOC is responsible for standard calibration of all WFI data, producing calibrated individual direct imaging and spectroscopic exposures (Data Level-2), regularized and mosaic image and image data catalogs (Data Level-3 and 4), and archiving of all Roman data.

The SSC is responsible for the Microlensing Science Operations System (MSOS), which includes the production of photometry and light curves, the identification and characterization of microlensing events, measuring the detection efficiency of the microlensing pipeline, production of relevant data products, including the time series light curves for all objects in the field, and user support. The SSC will produce both Level 3 and 4 data products for the GBTDS.

The SSC will be the main interface with the scientific user community for Exoplanet Microlensing and the high-level data products. The SSC will support observers, including relevant Project Infrastructure Teams, archival researchers, and the astronomical community at large. The SSC will also provide educational and public outreach on Exoplanet Microlensing. The SSC is responsible for creating and securing the scientific legacy of Roman exoplanet microlensing in the form of reliably reduced, calibrated and well-documented data products.

The PIT will produce Level 5 data products relevant to the planet demographic goals described in the Roman Science Requirements Document.

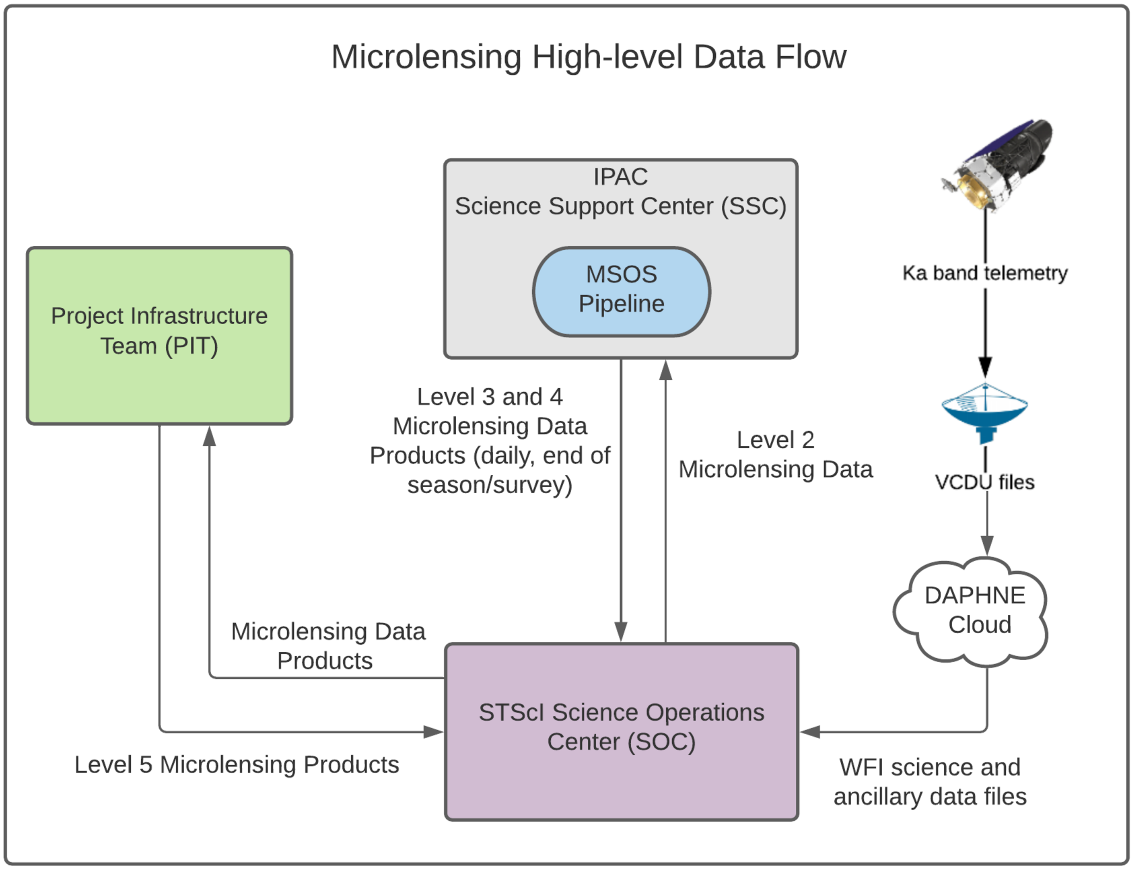

Figure 4: The high-level data flow for GBTDS/Exoplanet Microlensing Investigation Data

3.2 Microlensing Science Operations System (MSOS)

The MSOS has been designed to meet the relevant exoplanet demographic requirements as described in the Roman Science Requirements Document.

3.3 Roman Exoplanet Microlensing Investigation

The high-level science goal of the Roman Exoplanet Microlensing Investigation is to utilize the gravitational microlensing technique to produce a statistical census of exoplanetary systems, with focus on planets from the outer habitable zone to free floating planets, including analogs of all the planets in our Solar System with masses greater than that of Mars. Here we summarize the properties of the survey specified in the Roman Design Reference Mission as needed to understand the structure of the MSOS processing. The MSOS pipeline will accommodate the properties of the final GBTDS.

The survey will comprise 7 fields in the Galactic Bulge and will include of 6 seasons spread along the 5 years of Roman prime mission. Each season will consist of nearly continuous observations for ~60-70 days each during the Roman visibility window towards the Galactic Bulge (around the spring and fall equinox). Observations of each field will be taken every 15 minutes with the wide band filter and every 6 hours with each of two additional filters. The science requirements are based on the above survey monitoring ~240 million star-years with an expected star density down to J=23 of ~ few hundred million stars per square degree.

3.4 Pipeline Processing and Data Products

The MSOS pipeline takes as input Level 2 calibrated data, creates a catalog of objects, and measures the astrometry and photometry for each of those objects for each epoch of the survey (nominally every 15 minutes), along with additional ancillary products. The second stage of the MSOS identifies and analyzes microlensing events, including planetary events, using MSOS photometry products as input. Finally, the MSOS will also quantify the pipeline detection efficiency, using simulated data. Figure 5 shows a schematic of the MSOS components.

The MSOS pipeline will have three modes: daily, end-of-season, and end-of-survey. The three timeline modes reflect the cadence of the GBTDS and the corresponding plan to deliver the science data products to the Roman archive. Each mode and its resulting products are described in more detail below. The MSOS pipeline will run and generate data products in an automated mode. The pipeline code will be distributed to the community.

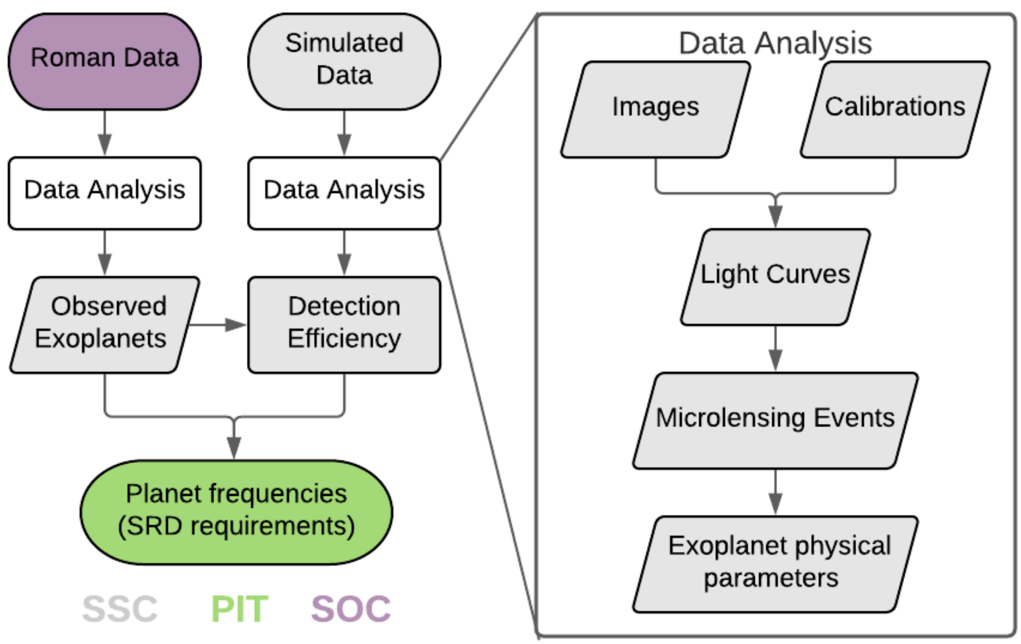

Figure 5: High-level processing diagram for the MSOS pipeline. More details are given in the text.

3.4.1 First 30 days

During the first 30 days of the first season, the daily pipeline will not operate, as the necessary catalog data will not be available. The initial point spread function (PSF) models will be in part based upon high-fidelity models of the optics and in part of upon detector models developed during the Commissioning Phase. The data reduction plan for the first 30 days is:

1. Create a super-sampled reference image using data from every 8 days.
2. Establish preliminary PSF models
3. Use PSF fitting on the reference images to create the object catalog of known stars. This could be assisted using a catalog of known stars in the survey field if it exists.
4. Once there are multiple difference images, use difference imaging on the reference images to improve the object catalog and check for new events.

The SSC will not deliver data products to the SOC during the first 30 days of the first microlensing survey season. Once the above steps are complete, the SSC will process the first 30 days of data following the daily pipeline mode.

3.4.2 Daily Pipeline

The science goal of the daily pipeline is to provide well-calibrated data products during each microlensing season. It will run every day during the microlensing observing season and will complete within 12 hours after receiving the relevant data products from SOC. The archival products produced will be returned to the SOC within 48 hours. Additional processing will occur every 8 days as described below.

To preserve the critical temporal information in the microlensing survey, the daily pipeline will start with the Level 2 calibrated data from SOC. As noted below, the daily processing steps are common across all modes of the pipeline. These steps are:

1. Update the PSF model
2. Run PSF fitting on the stars in the current object catalog
3. Update light curves with results of PSF and DIA (difference imaging analysis)
4. Produce data quality metrics
5. Return data to the archive

After every 8 days the daily pipeline will execute the following steps, in addition to the daily steps:

1. Create a new super-sampled reference image
2. Update the PSF model for the reference image
3. Use DIA to compare most recent reference image to previous reference image to improve knowledge of stars in the current object catalog, check for new events and update the object catalog
4. Generate photometry (using PSF fitting) for any new events
5. Measure and report the higher-order order moments of the PSF (dispersion, skewness, and higher order moments, if possible, using algorithms specifically suited to Roman's complicated PSF)
6. Flag objects for which the above algorithms (PSF fitting, moment measuring) fail.
7. Return data to the archive

The data products created by the MSOS daily pipeline for each identified source include: flux, position, PSF fit shape measurements, and quality flags. The MSOS will deliver Level 3 and Level 4 products of the daily pipeline to the SOC within 2 days of receipt of the last input data needed for those products.

3.4.3 End-of-Season Pipeline

At the end of each microlensing observing season, the final reference image will be used to generate the list of new sources to be added to the object catalog. Then the daily pipeline will be run again to produce all the products from the daily pipeline in both the regular and 8-day modes. The photometry section of the pipeline will be re-run on applicable previous seasons when substantial improvements due to parallactic sampling are made in the source positions and proper motions. The goal is to complete this processing within 2 months of receiving the relevant data products from SOC and the requirement is 6 months. In addition, the end-of-season pipeline will:

1. Create super-sampled images in each filter of entire field.
2. Identify variable objects from light curves to produce the variability catalog.
3. Time-trend all time-variable objects, including intrinsic proper motion and intrinsic moment changes.
4. Identify microlensing events.
5. Refine photometry analysis for microlensing events.
6. Characterize microlensing model, including number of lensing bodies and source components.
7. Evaluate physical parameters of microlensing events; models for previously identified events will be refined.
8. Characterize pipeline detection efficiency, including completeness and reliability metrics via injection and recovery of events.

The additional data products produced by this mode are the variability catalog, and a catalog of microlensing events with model fits, physical parameters, and quantified measures of the pipeline detection efficiency. These data products will be sent to the SOC for inclusion in the Roman Archive. The SSC will deliver Level 3 and Level 4 data products of the end-of-season pipeline to the SOC within 6 months of receipt of the last input data needed for those products.

3.4.4 End-of-Survey Pipeline

At the end of the microlensing survey, a final run of the pipeline will run on the complete data set to produce all products from the end-of-season pipeline, and will complete within 6 months after receiving the relevant data products from the SOC. In addition to the end-of-season products, the following data products are also produced by the MSOS pipeline at SSC:

1. Relative source-lens proper motion for all robustly identified sources (after the second observing season for each of the spring and fall telescope orientations.)
2. Parallax estimates for all catalogued sources; note that this is the traditional "distance" parallax due to the annular motion of the sun (as opposed to the "microlensing" parallax), which, for the majority of sources (including

microlensing source stars), can potentially be measured for both the source and the lens.

The MSOS shall deliver Level 3 and Level 4 data products of the end-of-survey pipeline to the SOC within 6 months of receipt of the last input data needed for those products.

3.4.5 Pipeline Completeness and Reliability

The MSOS pipeline for the image and photometry analysis and the following identification and characterization of microlensing events, and in particular of planetary events, will be completely automated. This allows an evaluation of the completeness (false negative rate) and reliability (false positive rate) of the pipeline itself. This analysis will be carried out with simulated data, both at the catalog level and at the image level. The more computationally expensive and resource demanding image-level simulation will be performed for a subset of data so to assess its results versus the catalog level simulation. The completeness is evaluated as the ratio of recovered “true” signal versus the total number of simulated ones and the reliability is the fraction of false positive identified within the simulation. The reliability and completeness modes will be included in the distributed version of the MSOS pipeline and the output products will be transferred to the Roman Archive within 6 months of the receipt of the last input data needed.

3.4.6 Data Product Quality Assessments (QA)

All modes of the MSOS pipeline will include automated data quality assessments. The QA data products will be included in the microlensing data products transferred to the Roman Archive.

3.4.7 Data Volume

The Level 2 data volume during the microlensing survey is 4.2 TB per day. The total survey image data volume for 6 seasons is 1.9 PB. The Level 3 and 4 data products produced by MSOS will be smaller in data volume, with an estimated volume of 0.5 Tb per day during the season. The assumed size for the source catalog at the end of the survey is 500 TB. This corresponds to a daily rate of 1.1 TB during the microlensing observing.

3.5 Data Products

Table 1 summarizes the data products produced by the pipeline and how often they are updated.

Data product		Update Frequency
Reference frame		
	Wideband filter	Every 8 days
	Other filters	Once per season
Object catalog		

	ID, position, flux, FWHM	Every 8 days
	Proper motion, trig. parallax	Once per season
Light curves		
	Per exposure PSF photometry	Daily
	Per exposure DIA photometry	Daily
	Updates for microlensing events	Once per season
Variability catalog		Once per season
Microlensing event catalog		Once per season
Pipeline detection efficiency		At end of survey
Reddening and extinction maps		At end of survey

4 CGI Technology Demonstration Support

4.1 Roles and Responsibilities

The CGI Operations Support (COS) and CGI Data Management System (CDMS) teams are part of the CGI Ground System support for Roman at SSC. The Teams are responsible for commanding CGI, assessing and trending health and safety of CGI, Level 1 data processing, validating and delivering Level 1-4 data, operation of High Order Wavefront Sensing and Control (HOWFSC) with Ground in the Loop (GITL), and I&T and commissioning support.

The Coronagraph Technology Center (CTC) is made up of the technologists responsible for carrying out the technology demonstration in Phase E, Level 2 to 4 data processing, and maintaining (1) the CGI functional and operations testbeds, (2) CGI flight software and (3) HOWFSC/GITL ground software.

The Community Participation Program (CPP) refers to the scientists selected from the community to work with the CTC to carry out the technology demonstration.

4.2 CGI Data Flow

This section provides a high-level description of the CGI data flow, from downlink to archiving. The Roman ground stations are responsible for capturing telemetry from the spacecraft science recorder and a housekeeping (HK) recorder. The science recorder captures all science observation (image) data along with "unfiltered" (sampled as produced) instrument and spacecraft telemetry, while the HK recorder contains instrument and spacecraft telemetry sampled more coarsely to allow faster download.

The science recorder data stream contains all astronomical target data from EXCAM and the Low Order Wavefront Sensor (LOWFS) data, as well as other requested telemetry, at an unfiltered cadence. As with data from the HK recorders, these data are in raw packetized (level 0) form. However, the science recorder data will be downlinked from the spacecraft through the ground station to the DAPHNE cloud. The data will then flow to the SOC and then to the SSC via the Roman archive. All CGI data will be stored in the data analysis environment (DAE) maintained by the SSC for higher-level data processing. Each of the types of higher processed data (L1-L4) will be transferred back to the Roman archive. This data flow is shown in Figure 6. The SSC is also responsible for delivering all CGI Integration and Test (I&T) and commissioning data to the SOC for archiving.

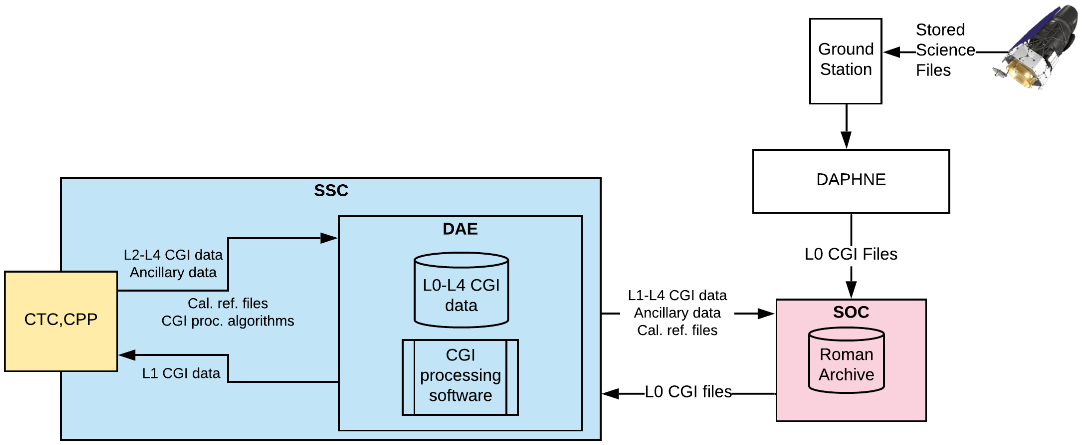

Figure 6: Illustration of the CGI data flow from the science recorder during nominal operations. Level-0 CGI data flow from the spacecraft to SSC via the Roman Ground System. The SSC processes L0 data, and also delivers higher level CGI data (L1-L4), ancillary data and calibration reference files, to the SOC. The CTC and CPP are shown partially within the SSC box because it is expected that much of the higher level CGI data processing will occur within the DAE on premises at SSC.

4.2.1 Data Processing Levels

The primary astronomical observation data from CGI will be taken with its EXoplanetary systems Camera (EXCAM). The CGI data levels are defined from Level 0 to 4. The Level 0 EXCAM data are raw packetized "telemetry" files received from either the MOC (in the case of HOWFSC) or the SOC at the SSC. The CDMS processes these data to Level 1 (raw, uncalibrated images, with a metadata header). These data are stored in the Data Analysis Environment for CTC and CPP to process to Level 2 (calibrated images), Level 3 (photometry, astrometric, and wavelength corrected images), and Level 4 (PSF subtraction and spectral extraction).

4.2.2 Ancillary Technology Demonstration Data Products and Calibration Reference Files

Ancillary CGI data will be processed by the CTC and CPP, validated by the CDMS, and sent to the SOC for archiving. The ancillary data includes:

- Data from mechanisms and camera settings
- Status from each wavefront control system at the time of the observation
- Calibrations generated autonomously on board
- Subset of housekeeping data on the science recorder
- Low-order Camera (LOCAM) images

Calibration reference files will also be generated from observational and/or modelling data by the CTC. These files may include, but are not limited to:

- Flat fields

- Darks
- Bad pixel maps
- Astrometric calibration files (including pixel scale solution, distortion correction, satellite spot calibration, etc)
- Photometric calibration files (including flux conversion file, etc.)
- Instrumental polarization calibration files

A limited set of calibration files generated on the ground will be sent up to the spacecraft for use during CGI observations. The data are created by the CTC and uploaded by the MOC. Currently, the flat fields are the only image files anticipated to be needed for on-board calibration.

4.2.3 Data Volume

The peak CGI data rate spacecraft allocation is 60 Mbps, with a CBE (current best estimate) of 4.6 Mbps. Over a 24 hr period, the CBE CGI data volume is about 0.430 Tb, with a spacecraft allocation of 3 Tb for Level 0 data. The Level 1 to Level 4 data volume for an average observation is expected to be 2.6 Tb. For the full technological demonstration, the expected data volume will be 34 Tb.

4.2.4 Data Analysis Environment

The SSC will operate and maintain a Data Analysis Environment (DAE) combining processing resources and disk space, to support the data processing needs of the CGI technology demonstration, including HOWFSC/GITL and the L2-L4 data processing of the CTC and CPP. The DAE will support compliant software developed by the CGI team at JPL, CTC, or Community Participation Program teams. Access to the DAE will be in person (at SSC premises) or remotely via two factor authentication. A diagram of the prototype DAE structure is depicted in Figure 7.

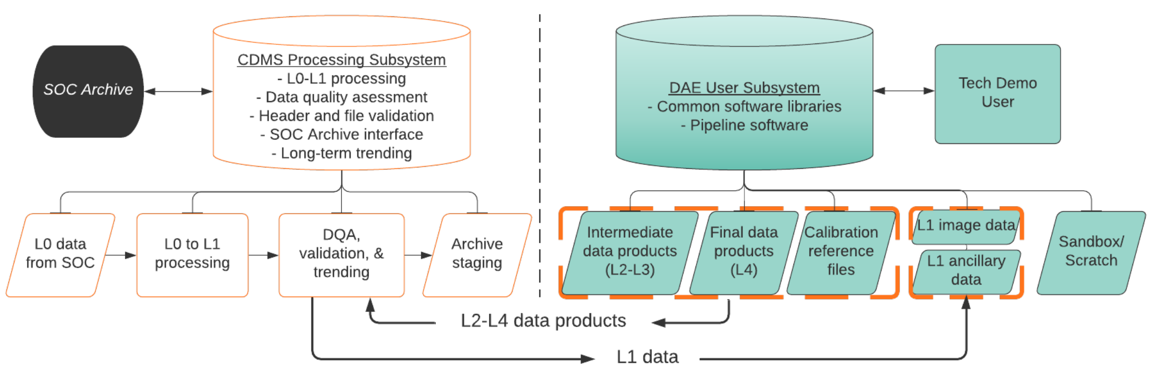

Figure 7: Diagram of the Data Analysis Environment structure. (Left): The CDMS processing subsystem, connected to the Roman archive, houses data and carries out automated SSC data management tasks. (Right:) The DAE user subsystem is accessible to Tech Demo users for higher level processing.

4.2.5 Data Quality Assessment

During Observatory I&T, commissioning and nominal operations, the SSC will perform Data Quality Assessment (DQA) of the CGI data. Any anomalies in the data will be flagged. A DQA product will be included with the data products transferred to the SOC for ingestion into the Roman archive. The DQA team will regularly produce summary reports on anomalies, calibration, and other instrument data trending.

4.3 CGI Operations Overview

During the CGI Technology Demonstration, 90 days of observations spread over the first 18 months of the Roman mission, nominal operations will be conducted in a mostly autonomous mode, referred to as “event-driven.” For event-driven operations, the CTC, together with the Participating Scientists, will design CGI observations and calibrations, creating observation specifications using Command Product Generation Software (CPGS), developed by COS. The CTC, together with the CPP, will design CGI observations and calibrations using this tool. Data from CGI will be processed to Level 1, checked for quality, and hosted in the DAE for CTC and CPP members to perform further processing. CDMS will validate higher level data products and send them to the Roman archive. COS and CDMS will also generate regular health and safety data reports base on engineering data.

4.3.1 Performance Monitoring

Target and calibration observations, HK, and ancillary data obtained during Phase E (Nominal Operations) will be used to derive statistical trends of several indicators of CGI performance. Listed below are a representative set of items that will be regularly monitored during operations.

- Bad pixel monitoring (dead and/or varying pixels)
- Dark current measurements (using non-illuminated pixels)
- Charge transfer effects (QE response vs. source flux)
- Absolute flux calibration (based on L3 data)
- Deformable mirror performance
- Detector gain variations (from pulse height distributions)
- Cosmic ray flux
- LOWFS performance (Zernike coefficients, total flux, dark current in LOCAM data)
- HOWFSC algorithm performance

5 Appendices

5.1 Appendix A: List of Acronyms

CDP	Calibration Data Pipeline
CDMS	CGI Data Management System
CGI	Coronagraph Instrument
COS	CGI Operations System
CPGS	Command Product Generation Software
CPP	Community Participation Program
CTC	Coronagraph Technology Center
DAE	Data Analysis Environment
DAPHNE	Data Acquisition, Processing, and Handling Environment
DIA	Difference Image Analysis
DQA	Data Quality Assessment
EXCAM	EXoplanetary Systems Camera
FOV	Field of View
GBTDS	Galactic Bulge Time Domain Survey
GDPS	Grism-Prism Data Processing System
GITL	Ground in the Loop
GSFC	Goddard Space Flight Center
HK	House Keeping
HLC	Hybrid Lyot Coronagraph
HOWFS	High-Order WaveFront Sensor/Sensing (context dependent)
HOWFSC	High-Order WaveFront Sensing and Control
HOWFSC/GITL	High-Order WaveFront Sensing and Control with Ground in the Loop
I&T	Integration and Testing
JPL	Jet Propulsion Laboratory
LAM	Laboratoire d'Astrophysique de Marseille
L0	Level 0 data
L1	Level 1 data
L2	Level 2 data
L3	Level 3 data
L4	Level 4 data
LOCAM	Low-Order wavefront sensor CAMera
LOWFS	Low-Order WaveFront Sensor
Mbps	Mega-bits per second
MOC	Mission Operations Center
MSOS	Microlensing Science Operations System
Pb	Petabyte
PSF	Point Spread Function
QA	Quality Analysis
SDP	Science Data Pipeline
SOC	Science Operations Center
SPC	Shaped Pupil Coronagraph
SSC	Science Support Center
STScI	Space Telescope Science Institute
Tb	Terabit(s)

TB	Terabyte(s)
WFI	Wide Field Instrument