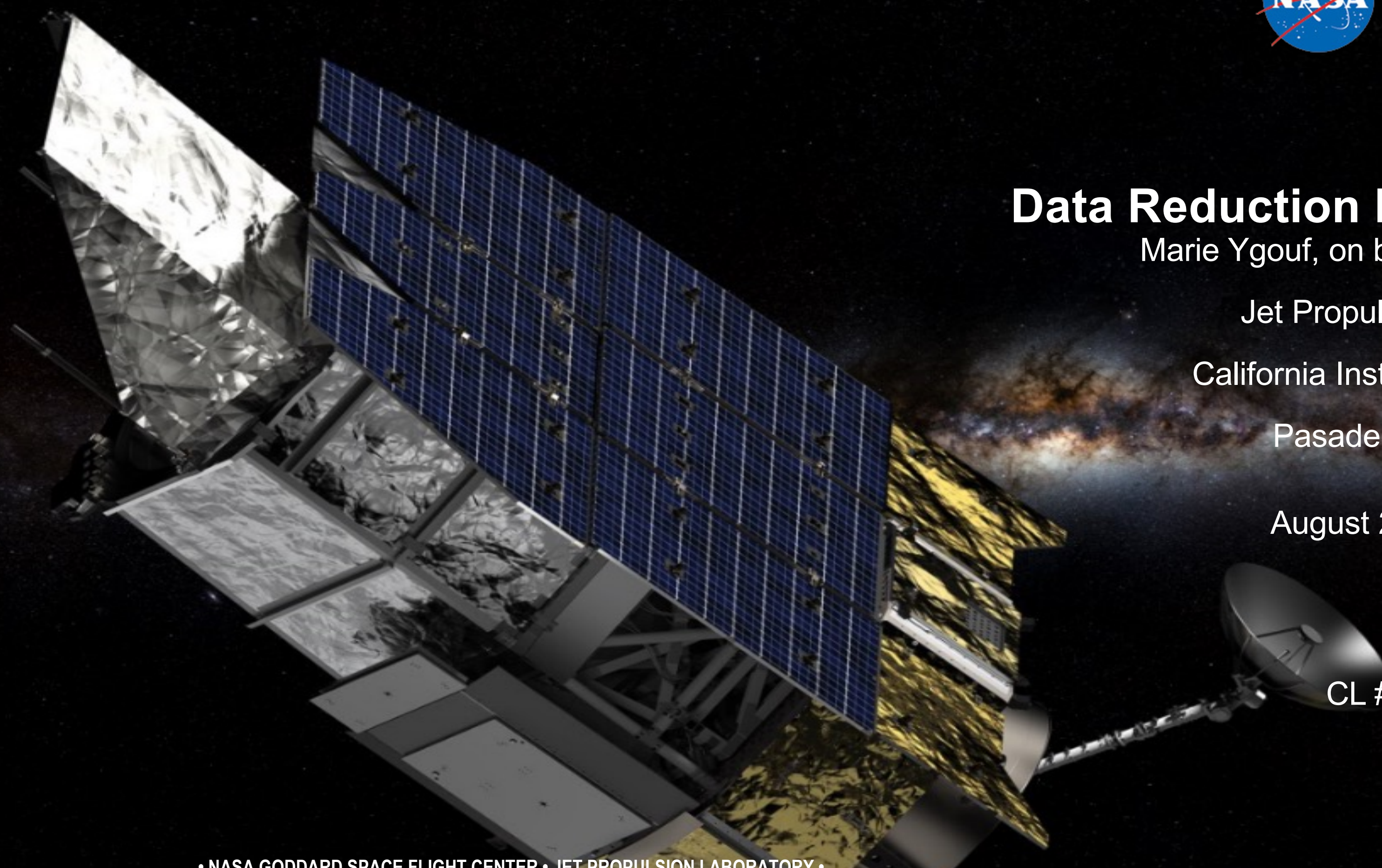


Jet Propulsion Laboratory
California Institute of Technology

Data Reduction Pipeline (DRP) Plans

Marie Ygouf, on behalf of the DRP team

Jet Propulsion Laboratory

California Institute of Technology

Pasadena, CA 91109

August 26 – 27, 2024

CL # 24-4934

• NASA GODDARD SPACE FLIGHT CENTER • JET PROPULSION LABORATORY •
• L3HARRIS TECHNOLOGIES • BALL AEROSPACE • TELEDYNE • NASA KENNEDY SPACE CENTER •
• SPACE TELESCOPE SCIENCE INSTITUTE • INFRARED PROCESSING AND ANALYSIS CENTER •
• EUROPEAN SPACE AGENCY • JAPAN AEROSPACE EXPLORATION AGENCY •
• CENTRE NATIONAL d'ÉTUDES SPATIALES • MAX PLANCK INSTITUTE FOR ASTRONOMY •

Jet Propulsion Laboratory
California Institute of Technology

Introduction – Stakeholders and Project Organization

ROMAN CORONAGRAPH

Community Participation Program (CPP):
- Multi-institutional team of astronomers and technologists; externally-funded
- Contacts:
  - Max Millar-Blanchaer & Jason Wang: DRP CPP working group co-leads

Project Science (PS) team:
- JPL and GSFC astronomers on the CGI team
- Contacts:
  - Marie Ygouf: PS Product Delivery Lead
  - Vanessa Bailey: Instrument Technologist / CPP co-chair

Coronagraph Technology Center (CTC):
- JPL personnel staffing the Functional Testbed and CGI Ground Software (GSW) support
- Contacts:
  - Tim Koch: CTC lead
  - Cynthia Wong: CTC GSW lead

Science Support Center (SSC):
- IPAC personnel supporting observation planning, operations, and data archiving
- Contacts:
  - Alexandra (Alex) Greenbaum
  - Jim Ingalls

- Roman Coronagraph Data Reduction Pipeline (DRP) Development:
  - ~20 active developers and consultants from 12 different institutions
  - Bi-weekly meetings

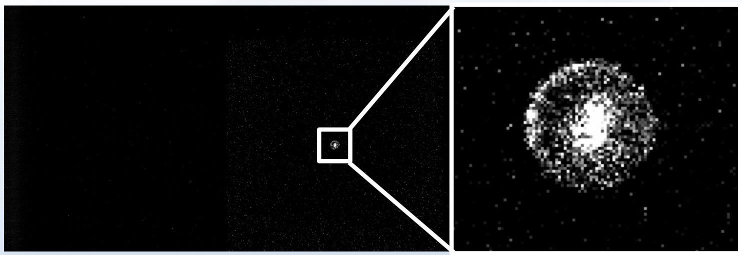
CSUN CALIFORNIA STATE UNIVERSITY NORTHRIDGE
NASA
JPL
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE
STScI SPACE TELESCOPE SCIENCE INSTITUTE
Goddard SPACE FLIGHT CENTER
THE UNIVERSITY OF ARIZONA
LAM LABORATOIRE D'ASTROPHYSIQUE DE MARSEILLE
Northwestern University
ipac

Introduction - Pipeline Objectives

Jet Propulsion Laboratory
California Institute of Technology

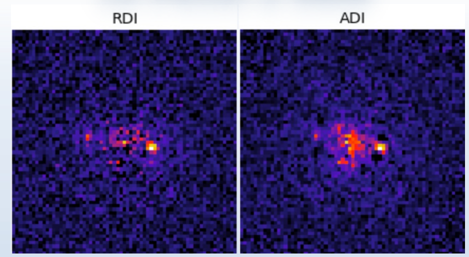
ROMAN CORONAGRAPH

Going from Level 1 data



To Level 4 data

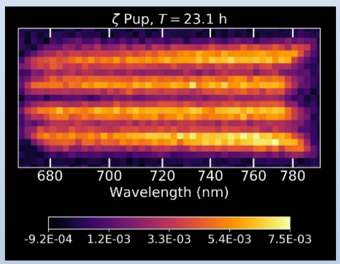
Imaging



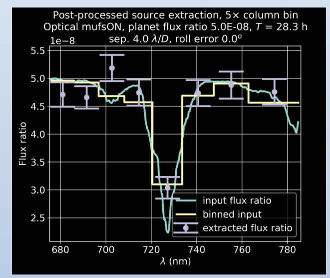
Ygouf et al. 2021

Scope

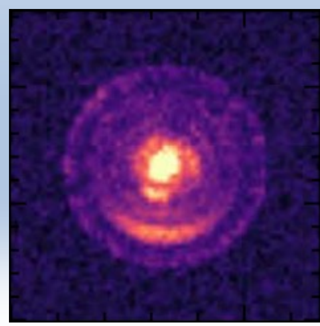
Produce a tool that processes Roman Coronagraph EXCAM raw Level 1 frames and deliver L2 - L4 data products to the Roman Science Support Center (SSC) for public archiving



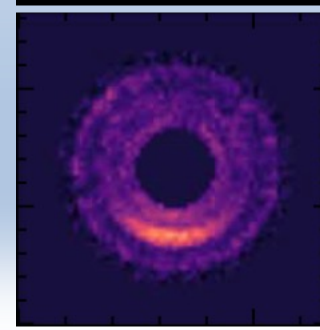
Spectroscopy



Ygouf et al. 2022



Polarimetry



Anche et al. 2023

Pipeline Requirements – Data Levels

Packetized data	Raw imaging and spectroscopic images	Calibrated image		Single image, normalized by exposure time, with WCS provided	Final combined image for entire observing sequence. PSF-subtracted, if applicable.	Analysis results used to verify TTR5
Level 0 (L0)	**Level 1 (L1)**	**Level 2a (L2a)**	**Level 2b (L2b)**	**Level 3 (L3)**	**Level 4 (L4)**	**Tech Demo Analysis (TDA)**
•Collection of packetized data. Data products include the instrument and observatory telemetry and the raw packetized stellar and calibrated target data received at the ground stations.	•The L0 CGI data are combined with the formatted engineering telemetry and appropriate metadata to generate L1 data. L1 data (imaging and spectroscopic) are formatted into 2D FITS (or equivalent) images. •The L1 data generated by the SSC are stored in the DAE and transmitted to the SOC for ingestion into the Roman archive.	•The L1 data files undergo a first round of image calibration to produce L2a data. •A modified version of the L2a II&T pipeline for on-sky data processing will be delivered by the Project Science (PS) + Community Participation Program (CPP) teams to the CTC •These data, including the relevant calibration reference files are generated by the CTC in the Data Analysis Environment (DAE). Data are validated for format, naming convention and data quality by the CDMS, and delivered to the SOC for ingestion into the Roman archive.	•The L2a data files undergo a second round of image calibration to produce L2b data. Photon-counted data are coadded to build up enough flux to allow application of the additional calibration corrections (eg: flatfield). •A modified version of the L2b II&T pipeline for on-sky data processing will be delivered by the Project Science (PS) + Community Participation Program (CPP) teams to the CTC. •The L2b data are generated by the CTC in the DAE. L2b data are validated by the CDMS and delivered to the SOC for ingestion into the Roman archive.	•L2b data are normalized by exposure time. Polarimetry images are refactored into FITS files with one polarization state in the primary image and the orthogonal state in the first image extension. World Coordinate System solutions are added to imaging and polarimetry data. Single object spectra will be wavelength calibrated, and metadata specific to the slit observations will be added to the headers (type of slit aperture, position of the slit, etc). •The pipeline for on-sky data processing will be delivered by PS + CPP to the CTC. •The L3 data are generated by the CTC in the DAE. L3 data are validated by the CDMS and delivered to the SOC for ingestion into the Roman archive.	•L4 data are distortion-corrected (if applicable), PSF-subtracted (if applicable), aligned, and combined. •The pipeline for on-sky data processing will be delivered by PS + CPP to the CTC. •The L4 data are generated by the CTC in the DAE. L4 data are validated by the CDMS and delivered to the SOC for ingestion into the Roman archive.	•These are products of observation analysis (e.g.: photometry, spectra, or detection limits) in physical units (e.g., magnitudes, nm, arcsec, and flux ratio noise). These products are used to verify TTR5 and to interpret astrophysical object observations. •The pipeline for data processing will be delivered by PS + CPP to the CTC. •Tech Demo Analysis Products are generated by the CTC in the DAE. Tech Demo Analysis Products are not validated by CDMS nor archived.
	L1 data produced by SSC	DRP requirements span those data levels				Not a requirement and no formal archiving at SSC

Pipeline Requirements

TTR5 on-sky data processing (narrow FOV imaging data only)

- R3.0.1 – Fall 2024
  - ❑ 1189255 L1 -> L2a processing (common for all EXCAM data)
  - ❑ 1115535 Analog L2a -> L2b processing
- R3.0.2 – Spring 2025
  - ❑ 1077735 Photon-counted L2a -> L2b processing
  - ❑ 1077736 L2b -> L3 processing
  - ❑ 1077737 PSF-subtracted L3 -> L4 processing
  - ❑ 1115536 PSF-subtracted L4 processing to analysis products
  - ❑ 1115537 Per-object L3 -> L4 processing
  - ❑ 1115538 Per-object L4 processing to analysis products

Goal Modes On-Sky Data Processing

- R3.0.1 – Fall 2024
  - ❑ 1090895 Prism dispersion scale and orientation
  - ❑ 1090896 Wavelength solution
  - ❑ 1090897 Slit calibration
  - ❑ 1090898 Line spread function
- R3.1 – Fall 2025
  - ❑ 1090891 Polarization flat-field
  - ❑ 1090892 Polarization absolute flux calibration
  - ❑ 1090893 Stokes vectors
  - ❑ 1090894 Mueller matrix

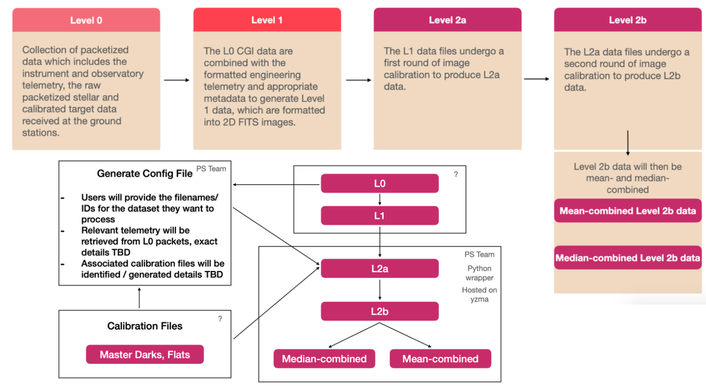
Calibration Reference Data

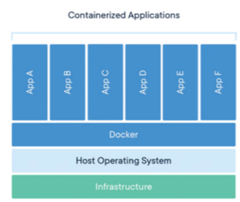
- R3.0.1 – Fall 2024
  - ❑ 1050914 Location for Boresight Calibration
  - ❑ 1050915 Orientation for Boresight Calibration
  - ❑ 1090871 Flat field
  - ❑ 1090880 Charge transfer inefficiency correction
  - ❑ 1050922 EXCAM fixed bad pixel maps
  - ❑ 1090879 Compute trap parameters
  - ❑ 1090867 per-pixel fixed pattern noise map
  - ❑ 1090868 per-pixel dark current map
  - ❑ 1090869 per-pixel clock induced charge CIC map
  - ❑ 1090870 median fixed-pattern and CIC residual
  - ❑ 1050921 Compute synthetic master dark
  - ❑ 1090872 K-gain
  - ❑ 1090873 Photon transfer curve
  - ❑ 1159970 Nonlinearity map
- R3.0.2 – Spring 2025
  - ❑ 1090874 Absolute flux calibration
  - ❑ 1090876 ND filter sweet-spot
  - ❑ 1090877 Computation of localized optical density
  - ❑ 1090878 Distortion map
  - ❑ 1090881 Individual core throughput PSFs
  - ❑ 1090882 Postprocessing for mask center location
  - ❑ 1090883 Core throughput map
  - ❑ 1090884 Core throughput dataset
  - ❑ 1159971 Astrophysical plate scale

Pipeline Architecture - II&T Legacy

- II&T version of Data Reduction Pipeline
  - Scope: Produce a tool so that Verification Activity (VA) owners can produce their own analog L2a and optionally L2b EXCAM data for II&T (Higher-level data products–level 3 and above–are neither needed nor have an application during this II&T)
  - Encompasses L1 to L2b wrapper and associated functions
  - L0 to L1 conversion is SSC responsibility
  - Github repo
  - Use of docker images/containers
  - Testing and validation framework (VIs/PBATs)
    - *see Tim's presentation on V&V*

II&T Data Processing Overview



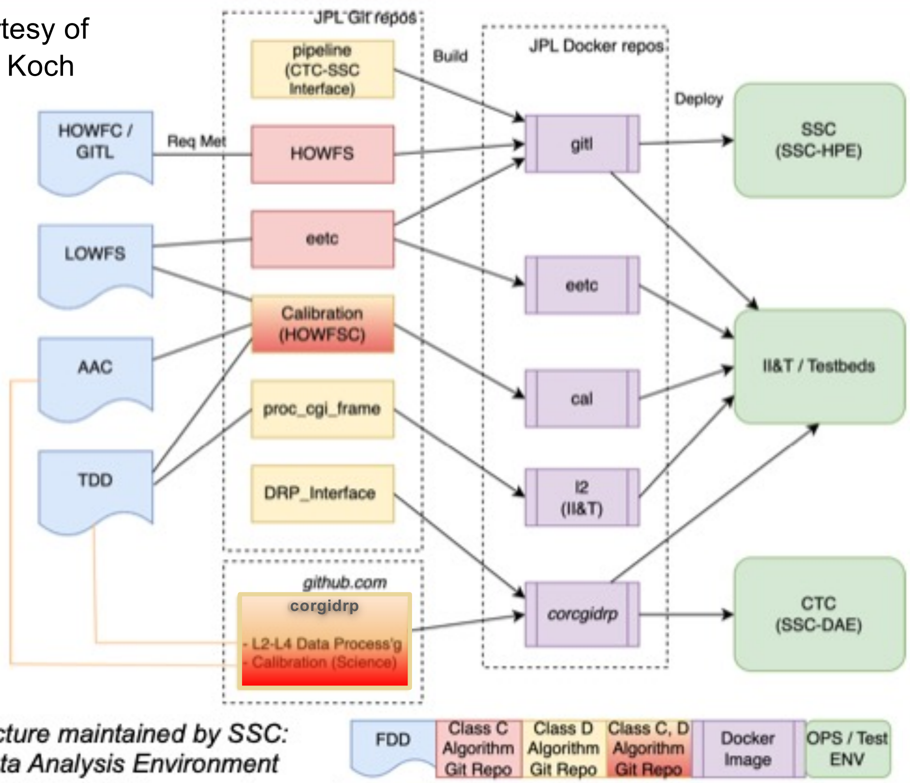


- On-sky DRP (nicknamed **CorGI**) built on this legacy but presents some architecture differences
  - Work is needed to:
    - Port the existing functions into the new architecture
    - Build associated docker image
    - Develop:
      - L2 to L4 functions and wrapper, calibration functions, and associated testing and validation framework

Pipeline Architecture – DRP into context

- **Production vs standalone software (CorGI)**

Courtesy of
Tim Koch

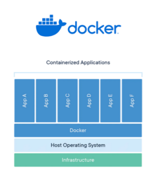


Infrastructure maintained by SSC:
DAE=Data Analysis Environment
HPE=HOWFSC GITL processing environment

Class C software critical to the operations of the instrument. "If this ain't working, we can't use CGI" (some L1-L3 calibration functions)

Class D software is important to the success of the instrument but is not required for operations (L3+ and associated calibrations)

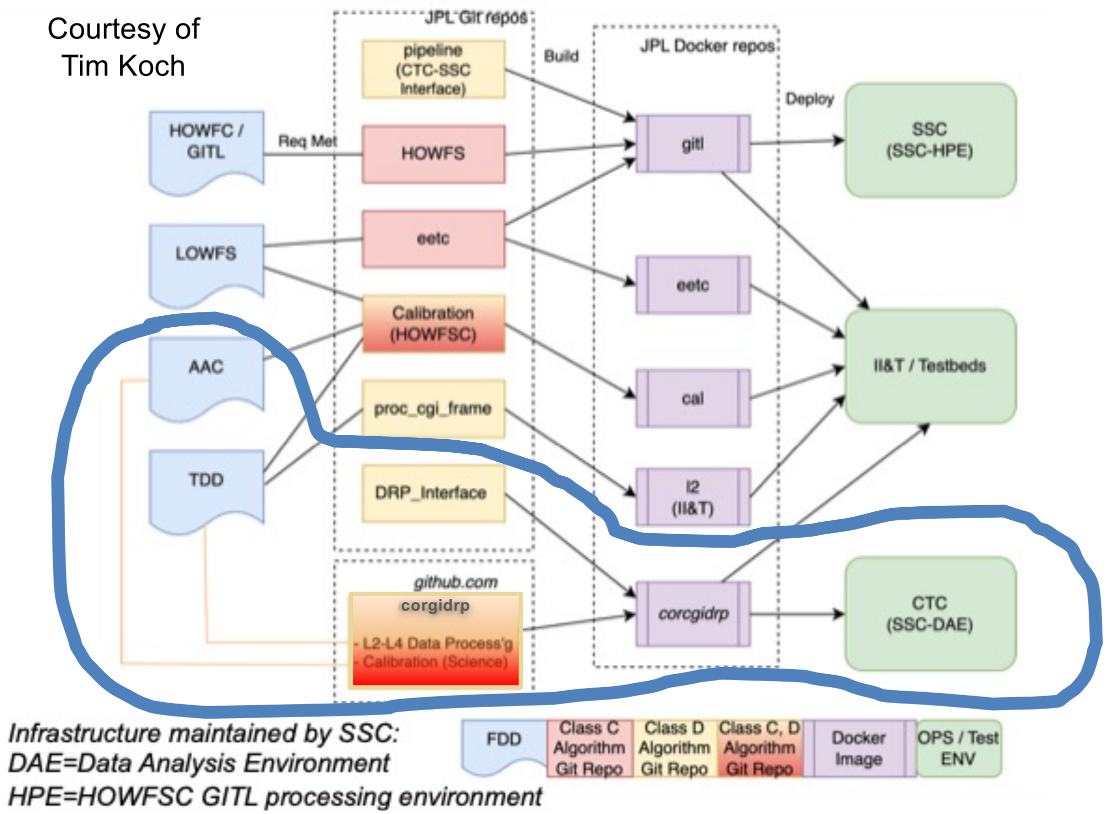
Docker Framework



- CTC/pipeline
  - L2: level-2 PGE (Docker image)
- Key benefits of using Docker:
  - **Isolation and Portability, Dependency Management, Reproducibility, Version Control and Rollback**
- Main Drawback:
  - Might introduce some overhead, especially for smaller projects or when dealing with extremely resource-constrained environments.

Pipeline Architecture – DRP into context

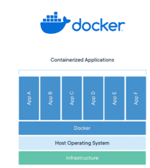
- **Production vs standalone software (CorGI)**



Class C software critical to the operations of the instrument. "If this ain't working, we can't use CGI" (some L1-L3 calibration functions)

Class D software is important to the success of the instrument but is not required for operations (L3+ and associated calibrations)

Docker Framework



- CTC/pipeline
  - L2: level-2 PGE (Docker image)
- Key benefits of using Docker:
  - **Isolation and Portability, Dependency Management, Reproducibility, Version Control and Rollback**
- Main Drawback:
  - Might introduce some overhead, especially for smaller projects or when dealing with extremely resource-constrained environments.

Pipeline Architecture

- **DRP standalone software architecture (CorGI)**
  - **Hosted on public github:**
    - **https://github.com/roman-corgi/corgidrp**

Terminology

CORGI-DRP: The overall data reduction pipeline infrastructure being developed by the CPP. Consists of: ...

The CORGI walker: The system that, given a new dataset, populates recipe template to generate recipe. Includes the functions to execute a recipe file.

BARC: The table of calibration file metadata. Stored as a .csv file.

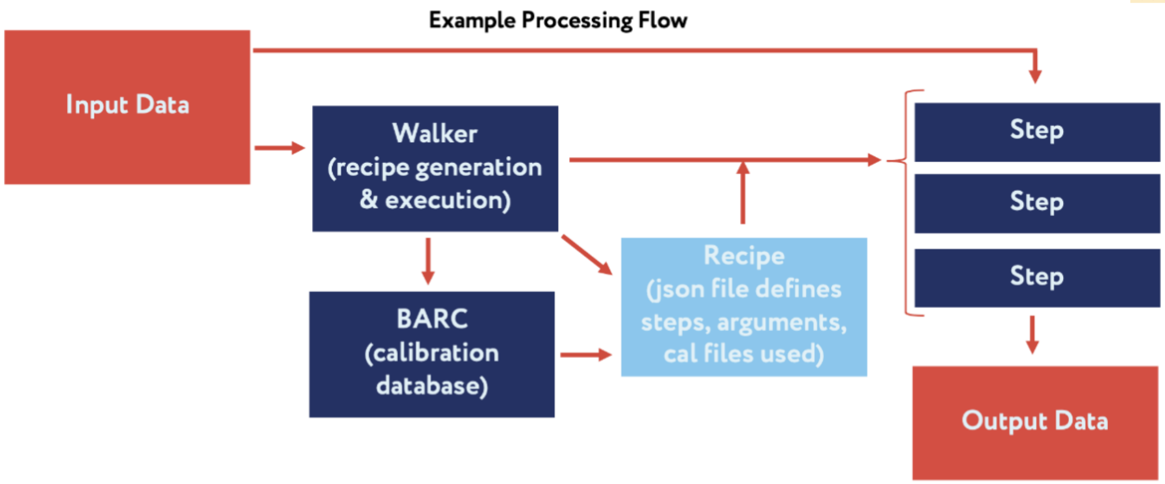
Step Function: A single python function that executes an algorithm to a set of data. Also known as "step" or "function".

Recipe: A list of step functions and function arguments to be run in sequence on a *specific* dataset. *Equivalent to a* ***run configuration file.***

Recipe Template: A template for a recipe file that will be applied to a general class of observing scenarios. A template will be used as a basis for generating a recipe upon the arrival of a new observational dataset. The templates will have specific fields (e.g. filenames, function arguments) that will be filled in by the walker.

Dataset: Dataset objects are instances of a specific DRP python class that contain the data and headers for multiple related data products.

Watchdog: The system that triggers DRP execution on receipt of a new L1 data.



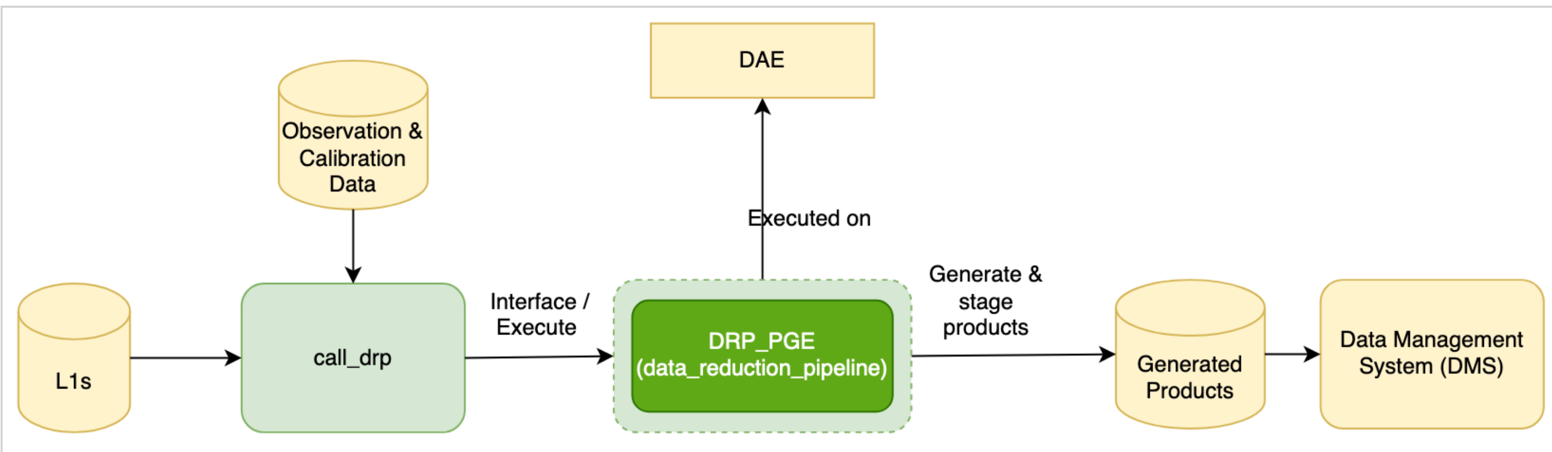
Millar-Blanchaer et al. 2024

Pipeline Architecture

- **Interfaces for DRP production software**

Interface with Community Participation Program (CPP) - Data Reduction Pipeline (DRP)

DRP are developed by Project Scientist (PS) / Community Participation Program (CPP) and delivered to CTC via PS. The open source method is adopted and CPP's source code resides at https://github.com/roman-corgi/corgidrp.



Pipeline Testing and Validation

Test Type	Description	Developed by	Executed by
Unit Tests	pytest that calls ut_* and are traditional unit tests	PS/CPP	GSW
Self Diagnostics Tests	Unit tests plus larger function like nulltest	PS/CPP	GSW
Simulated Tests	Demonstrates and simulates the test runs with simulation test data, pipeline, end-to-end testing (eg L2-L4 generation)	PS/CPP/ GSW	GSW
Acceptance Test	Includes all unit / self-diagnostics / simulated tests. Reviews the tests' status, logs messages, reports PASS/FAIL results	GSW	GSW
VAP (Verification Activity Plan) Regression Tests	CLW: *Description*; GSW gathers the test results and summary, and requirement metrics from V&V.	V&V team	V&V team

Releases Schedule - Timeline

Releases Schedule - Details

R3.0.1

TTR5 L1 to L2b processing

+ Spectroscopy functions (Goal mode)

+ Subset of calibration functions

- 2024 Sept 23: GFSC & CPP to PS
- 2024 Oct 01: PS to CTC
- 2024 Dec 18: CTC to SSC

Fall 2024

Spring 2025

R3.0.2

TTR5 Photon-counted L2a to L2b processing

+ TTR5 L2b to L4 processing

+ Remaining calibration functions

- 2025 Apr 1: CPP to PS
- 2025 Apr 14: PS to CTC
- 2025 May 26: CTC to SSC
- note: L1 headers are expected to be finalized by June 2025 (SSC R3.0)

Fall 2025

R3.1

Polarimetry functions (Goal mode)

- 2025 Sept 25: CPP to PS
- 2025 Oct 9: PS to CTC
- 2025 Nov 20: CTC to SSC

Spring 2026

R3.2

Launch version

- 2026 Mar 3: CPP to PS
- 2026 Mar 17: PS to CTC
- 2026 Apr 28: CTC to SSC

Release R3.0.1 - Progress Update

R3.0.1

TTR5 L1 to L2b processing

+ Spectroscopy functions (Goal mode)

+ Subset of calibration functions

- 2024 Sept 23: GFSC & CPP to PS
- 2024 Oct 01: PS to CTC
- 2024 Dec 18: CTC to SSC

Fall 2024

- July 2023 - CPP selection announcement - COMPLETED
- 10/19/2023 - First CPP Full team meeting - COMPLETED
- 1/16/2024 - First CPP DRP working group meeting - COMPLETED
- 2/7-9/2024 - First CPP in-person meeting - COMPLETED
- 6/30/2024 - CPP Internal delivery of L1 to L2b processing step functions - COMPLETED
- 7/22/2024 - CPP Internal delivery of calibration functions - COMPLETED
- 7/25/2024 - Initial delivery of L1 to L2a processing wrapper to CTC - COMPLETED
- 7/31/2024 - Successful porting of II&T L1 to L2b to DRP architecture - COMPLETED

Release R3.0.1 - Progress Update

- Successful porting of II&T L1 to L2b to DRP architecture:

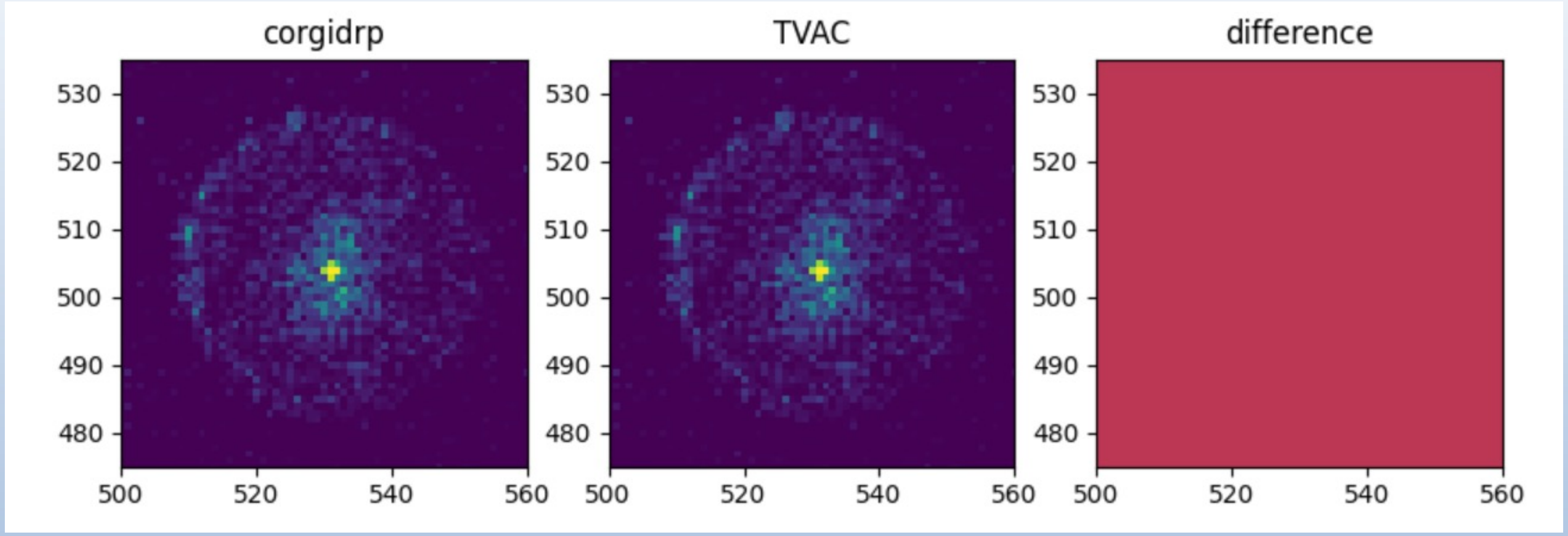


Figure produced by Jason Wang on 7/31:
corgidrp produces the **exact** same output as the TVAC pipeline

Release R3.0.1 - Next steps

R3.0.1

TTR5 L1 to L2b processing

+ Spectroscopy functions (Goal mode)

+ Subset of calibration functions

- 2024 Sept 23: GFSC & CPP to PS
- 2024 Oct 01: PS to CTC
- 2024 Dec 18: CTC to SSC

Fall 2024

- July 2023 - CPP selection announcement - COMPLETED
- 10/19/2023 - First CPP Full team meeting - COMPLETED
- 1/16/2024 - First CPP DRP working group meeting - COMPLETED
- 2/7-9/2024 - First CPP in-person meeting - COMPLETED
- 6/30/2024 - CPP Internal delivery of L1 to L2b processing step functions - COMPLETED
- 7/22/2024 - CPP Internal delivery of calibration functions - COMPLETED
- 7/25/2024 - Initial delivery of L1 to L2a processing wrapper to CTC - COMPLETED
- 7/31/2024 - Successful porting of II&T L1 to L2b to DRP architecture - COMPLETED
- 9/9/2024 – Complete integration and end-to-end testing of L1 to L2b processing and calibration functions
- 9/23/2024 – CPP delivery of R3.0.1 to PS
- 9/30/2024 – GSFC delivery of spectroscopy functions to CPP
- 10/01/2024 – PS delivery of R3.0.1 to CTC
- 11/12/2024 – CTC release of docker image
- 11/13/2024 – 12/4/2024 – VnV Test campaign
- 12/4/2024 – 12/14/2024 – Data Review
- 12/18/2024 – CTC delivery of R3.0.1 to SSC

Summary

- Roman Coronagraph DRP built on II&T Legacy
- Progress so far:
  - CPP ramped up earlier this year and provided their first interim delivery to CTC in July and second in August
  - Successful preliminary result from porting pipeline
- Huge thanks to developers and stakeholders for their work so far!
- Great deal of work ahead until launch and very good start

DRP CPP working group co-leads: Max-Millar Blanchaer (UC Santa Barbara) & Jason Wang (Northwestern University)

Developers and Consultants: Kevin Ludwick (UAH), Sergi Hildebrandt (JPL), Julia Milton (JPL), Ell Bogat (GSFC), Juergen Schreiber (MPIA), Ramya M. Anche (University of Arizona), Amanda Chavez (Northwestern University), Matthias Samland (MPIA), Lisa Altinier (LAM), Taïchi Uyama (CSUN), Rob Zellem (GSFC), Vanessa Bailey (JPL), Cynthia Wong (JPL), Al Niessner (JPL)

Back-up charts

Pipeline requirements

- L4 parents

1026525 - The CTC shall process Level 1 CGI data provided by SSC into Level 2 data.

1026526 - The CTC shall process Level 2 CGI data into Level 3 data.

1026527 - The CTC shall process Level 3 CGI data into Level 4 data.

1026528 - The CTC shall create Calibration Reference Data products required for Level 2 data processing.

1026529 - The CTC shall create Calibration Reference Data products required for Level 3 data processing.

1026530 - The CTC shall create Calibration Reference Data products required for Level 4 data processing.

1026531 - The CTC shall deliver Level 2 - Level 4 data, along with the Calibration Reference Data used to compute it, to SSC in accordance with the Roman SSC-CTC OIA.

1026532 - The CTC shall produce the final report on the technology demonstration verifying performance against PLRA requirements within 18 months after the completion of commissioning.

1026545 - The CTC shall maintain a technology demonstration target catalog for internal use.

1026547 - The CTC shall generate Observation Specifications using the Command Product Generation Software (CPGS).

Pipeline requirements

- TTR5 on-sky data processing (narrow FOV imaging data only)

Note: All on-sky data will also use the II&T L1 → L2a" pipeline, denoted by 'Same as L1 → L2a pipeline used for "Analog Data for II&T"' in the table above.

1115535 - Given a level L2a analog EXCAM data frame and associated cosmic ray mask, CTC GSW shall apply the "L2a → L2b" pipeline for "TTR5 Analog PSF-subtracted on-sky data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

1077735 - Given one or more level L2a photon-counted EXCAM data frames and associated cosmic ray masks, CTC GSW shall apply the "L2a → L2b" pipeline for "TTR5 Photon-counted PSF-subtracted on-sky data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

1077736 - Given a level L2b EXCAM data frame and associated bad pixel map, CTC GSW shall apply the "L2b → L3" pipeline for "TTR5 Analog PSF-subtracted on-sky data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

Note: Any data analysis marked above as "Same as L2b → L3 pipeline used for "TTR5 Analog PSF-subtracted on-sky data"" will also use this pipeline.

1077737 - Given one or more level L3 EXCAM data frames of target and/or reference stars at one or more roll angles intended for PSF-subtraction, CTC GSW shall apply the "L3 → L4" pipeline for "TTR5 Analog PSF-subtracted on-sky data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

Note: Any data analysis marked above as "Same as L3 → L4 pipeline used for TTR5 Analog PSF-subtracted on-sky data"" will also use this pipeline.

1115536 - Given 1) a PSF-subtracted level L4 EXCAM data frame, 2) the measured star apparent magnitude, 3) an absolute flux calibration scale factor, and 4) a core throughput calibration scale factor, CTC GSW shall apply the "L4 → TDA Products" pipeline for "TTR5 Analog PSF-subtracted on-sky data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

Note: Any data analysis marked above as '"Same as L4 → TDA Products pipeline used for "TTR5 Analog PSF-subtracted on-sky observation data"' will also use this pipeline

1115537 - Given one or more level L3 EXCAM data frames not intended for PSF-subtraction, CTC GSW shall apply the "L3 → L4" pipeline for "TTR5 Analog Non-PSF-subtracted on-sky observation data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

Note: Any data analysis marked above as 'Same as L3 → L4 pipeline used for "TTR5 Analog Non-PSF-subtracted on-sky data"' will also use this pipeline.

1115538 - Given 1) a level L4 EXCAM data frame without PSF subtraction, 2) an absolute flux calibration scale factor, and 3) a core throughput calibration scale factor CTC GSW shall apply the "L4 → TDA Products" pipeline for "TTR5 Analog Non-PSF-subtracted on-sky observation data" as described the table "Imaging Data Processing Pipelines" in D-105748 TDD FDD.

Note: Any data analysis marked above as 'Same as L4 → TDA Products pipeline used for "TTR5 Analog Non-PSF-subtracted on-sky observation data"' will also use this pipeline.

Pipeline requirements

- Calibration of EXCAM on-sky pointing

> 050914 - Given 1) a clean CGI image of an astrophysical field with one or more point sources, and 2) a representation of the star type, magnitude, and ICRF location of each source in the vicinity of targeted field, the CTC GSW shall compute the location of a point on the EXCAM detector in the ICRF frame to within 3mas RMSE.
>
> 1050915 - Given 1) a clean CGI image of an astrophysical field with two or more point sources, and 2) a representation of the star type, magnitude, and ICRF location of each source in the vicinity of targeted field, the CTC GSW shall compute the orientation of the EXCAM axes with respect to celestial north to within 0.05 degrees RMSE.
>
> NOTE: The values in requirements 1050914 and 1050915 represent the suballocations of the required calibration performance (30mas 3-sigma, 0.3 degrees 3-sigma per CTC L4s 1026523 and 1026524) to systematic algorithmic residuals. Remaining contributions from terms such as shot noise and PSF smearing will be handled operationally by observation planning choices. Anyone using this data set for absolute pointing measurement should keep in mind that while any one frame may be measured to high precision, RWA jitter, ACS residuals, and ACS drift mean that frame-to-frame measurements are *significantly* less repeatable than 30mas 3-sigma without significant averaging. Differential pointing measurements with WFI are expected to require less averaging, as ACS residuals will be largely common-mode between CGI and WFI.

- Astrometric calibration

> 1159971 - Given 1) one or more sets of clean CGI images of an astrophysical field with several stellar point sources, taken with different pointings, and 2) a representation of the star type, magnitude, and ICRF location of each source in the vicinity of targeted field, the CTC GSW shall compute the average pixel scale along both perpendicular axes of EXCAM's field of view in mas/pixel.
>
> NOTE: This is part of Calibration Reference Data used for the L2b -> L3 pipeline.
>
> 1090878 - Given 1) one or more sets of clean CGI images of an astrophysical field with several stellar point sources, taken with different pointings, and 2) a representation of the star type, magnitude, and ICRF location of each source in the vicinity of targeted field, the CTC GSW shall compute a distortion map for the EXCAM focal plane.
>
> NOTE: This is part of Calibration Reference Data used for the L2b -> L3 pipeline.

Pipeline requirements

- Flat fields calibration

1090871 - Given 1 or more sets of clean focal-plane images collected with "collect flat" at a particular on-sky pointing, the CTC GSW shall construct a single flat field calibration reference image.

- Calibration of K-gain/Nonlinearity

1090872 - Given 1) Level-1 pupil-plane images at a particular on-sky pointing with a constant EM gain but different exposure times, 2) the EM gain, and 3) a list of the exposure times associated with the frames, the CTC GSW shall compute k-gain (DN-to-electron conversion factor).

1090873 - Given 1) Level-1 pupil-plane images at a particular on-sky pointing with a constant EM gain but different exposure times, 2) the EM gain, and 3) a list of the exposure times associated with the frames, the CTC GSW shall compute a photon transfer curve.

1159970 - Given 1) one or more sets of Level-1 pupil-plane images at a particular on-sky pointing, each set with a different constant EM gain and a range of exposure times, 2) a list of EM gains associated with each set, and 3) a list of the exposure times associated with each set, the CTC GSW shall compute the nonlinearity curve for each EM gain.

> NOTE: the nonlinearity curve is a lookup table of correction factors in the vicinity of 1.0 as a function of applied gain and measured counts for the given EM gain, which may be interpolated on for any specific gain/count combination to correct for nonlinearity. It is a piece of Calibration Reference Data which is used in the L1 → L2a ground processing pipeline. See the Tech Demo Data FDD (D-105748) for further details.

Pipeline requirements

- Absolute flux calibrations

1090872 - Given 1) Level-1 pupil-plane images at a particular on-sky pointing with a constant EM gain but different exposure times, 2) the EM gain, and 3) a list of the exposure times associated with the frames, the CTC GSW shall compute k-gain (DN-to-electron conversion factor).

1090873 - Given 1) Level-1 pupil-plane images at a particular on-sky pointing with a constant EM gain but different exposure times, 2) the EM gain, and 3) a list of the exposure times associated with the frames, the CTC GSW shall compute a photon transfer curve.

1159970 - Given 1) one or more sets of Level-1 pupil-plane images at a particular on-sky pointing, each set with a different constant EM gain and a range of exposure times, 2) a list of EM gains associated with each set, and 3) a list of the exposure times associated with each set, the CTC GSW shall compute the nonlinearity curve for each EM gain.

> NOTE: the nonlinearity curve is a lookup table of correction factors in the vicinity of 1.0 as a function of applied gain and measured counts for the given EM gain, which may be interpolated on for any specific gain/count combination to correct for nonlinearity. It is a piece of Calibration Reference Data which is used in the L1 → L2a ground processing pipeline. See the Tech Demo Data FDD (D-105748) for further details.

Pipeline requirements

- Absolute flux calibrations

1090874 - Given 1) 1 or more sets of clean focal-plane images collected on an unocculted spectrophotometric standard star, and 2) the ND filter and CFAM filter configuration in use during that data collection, and 3) transmission curves for the ND and color filters, the CTC GSW shall compute a single absolute flux calibration value to convert photoelectrons to physical flux units for a given spectral type.

1090876 - Given 1) M sets of clean focal-plane images collected at different FSM positions on an unocculted spectrophotometric standard star with a neutral density (ND) filter in place in FPAM, and 2) a single absolute flux calibration value to convert photoelectrons to physical flux units, computed for the same CFAM filter, the CTC GSW shall compute a "sweet-spot dataset", an Mx3 matrix with the OD for each dither and the EXCAM (x,y) position for the star center.

> NOTE: this data set captures small-scale variation of focal-plane attenuation, including areas where transmission deviates significantly from expected. The FPAM encoder position associated with a sweet-spot dataset should be stored with it, because both will be needed when this data is used.

1090877 - Given 1) a clean unocculted focal-plane image on EXCAM with an ND filter, 2) the FPAM encoder position for that clean image, 3) a "sweet-spot dataset" taken with the same ND filter and CFAM filter, 4) the FPAM encoder position for the sweet-spot dataset, and 5) a transformation matrix from motions along the EXCAM pixel grid to motions of the FPAM, the CTC GSW shall compute the expected optical density applied to the clean image by the ND filter at that focal-plane location.

> NOTE: this tool uses the FPAM data to account for mechanism nonrepeatability and merges that with the measured OD variation.

Pipeline requirements

- Core throughput calibrations

1090881 - Given a core throughput dataset consisting of M clean frames (nominally 1024x1024) taken at different FSM positions, the CTC GSW shall estimate the pixel location and core throughput of each PSF.

NOTE: the list of M clean frames may be a subset of the frames collected during core throughput data collection, to allow for the removal of outliers.

1090882 - Given 1) the location of the center of the FPM coronagraphic mask in EXCAM pixels during the coronagraphic observing sequence and 2) the FPAM and FSAM encoder positions during both the coronagraphic and core throughput observing sequences, the CTC GSW shall compute the center of the FPM coronagraphic mask during the core throughput observing sequence.

1090883 - Given 1) an array of PSF pixel locations and 2) the location of the center of the FPAM coronagraphic mask in EXCAM pixels during core throughput calibrations, and 3) corresponding core throughputs for each PSF, the CTC GSW shall compute a 2D floating-point interpolated core throughput map.

1090884 - Given 1) a core throughput dataset consisting of a set of clean frames (nominally 1024x1024) taken at different FSM positions, and 2) a list of N (x, y) coordinates, in units of EXCAM pixels, which fall within the area covered by the core throughput dataset, the CTC GSW shall produce a 1024x1024xN cube of PSF images best centered at each set of coordinates.

Pipeline requirements

- Dark frames

1090867 - Given one or more raw dark frames collected during a dark frame collection calibration activity, the CTC GSW shall compute a per-pixel map of fixed-pattern noise in electrons.

1090868 - Given one or more raw dark frames collected during a dark frame collection calibration activity, the CTC GSW shall compute a per-pixel map of dark current in electrons per second.

1090869 - Given one or more raw dark frames collected during a dark frame collection calibration activity, the CTC GSW shall compute a per-pixel map of EXCAM clock-induced charge in electrons.

1090870 - Given one or more raw dark frames collected during a dark frame collection calibration activity, the CTC GSW shall compute a value of the median fixed-pattern and CIC residual in the raw frame prescan region.

NOTE: this will be used to correct per-frame bias estimation for the part that is not bias (and so does not appear in the active region of the detector).

1050921 - Given 1) a per-pixel map of fixed-pattern noise in electrons, 2) a per-pixel map of dark-current in electrons per second, 3) a per-pixel map of EXCAM clock-induced charge in electrons, 4) an EM gain value, and 5) an exposure time in seconds, the CTC GSW shall compute a master dark frame associated with that gain and exposure time.

NOTE: This is part of the Calibration Reference Data.

Pipeline requirements

- Trap pumping data

> 1090879 - Given 1) a set of clean focal plane images collected by a trap pumping sequence, and 2) the detector parameters (temperature, phase time, number of lines, number of cycles, and clocking scheme) associated with each set, the CTC GSW shall compute the following values for each trap in the region of interest: location (pixel, electrode), capture cross-section, and release time constant at the EXCAM observation operating temperature.
>
> 1090880 - Given 1) a set of charge trap parameters consisting of trap location (pixel, electrode), capture cross-section, and release time constant at the EXCAM observation operating temperature and 2) clean focal plane image(s) collected during an observation sequence, the CTC GSW shall compute and apply a charge transfer inefficiency (CTI) correction to the supplied images.
>
> NOTE: CTI correction will only be used for correcting image in ground-based pipelines, not in the "Process N Frames" activity. It is a piece of Calibration Reference Data which is used in the L2a → L2b ground processing pipeline. See the Tech Demo Data FDD (D-105748) for further details.

- Bad pixel maps

> 1050922 - Given 1) an EXCAM master dark frame, 2) a threshold for the number of standard deviations above the mean to flag hot pixels in the dark frame, 3) an EXCAM flat-field frame, and 4) a threshold for the fraction of the local mean level below which to flag low-response pixels, the CTC GSW shall compute an EXCAM fixed bad pixel map.
>
> NOTE: this is part of the Calibration Reference Data.

Pipeline requirements

- Polarimetric Calibrations

1090891 - Given 1 or more sets of clean focal-plane images collected with "collect flat" at a particular on-sky pointing and a Wollaston prism inserted, the CTC GSW should construct a single flat field calibration reference image for that pair of polarization states.

1090892 - Given 1) 1 or more sets of clean focal-plane images collected on an unocculted spectrophotometric standard star, 2) Wollaston and CFAM filter configuration in use during that data collection, and 3) a single integrated-light OD value corresponding to the target's observed position on the ND filter, if applicable, the CTC GSW should compute a single absolute flux calibration value to convert photoelectrons to physical flux units for the integrated light measurement from a given Wollaston prism for a given spectral type.

1090893 - Given 1) 2 clean focal-plane images of a polarimetric standard star, one with each Wollaston prism, and 2) Wollaston, CFAM filter, and ND filter configuration in use during that data collection, the CTC GSW should compute the Stokes Vector (3 coordinates I, Q, U) associated with that observation.

NOTE: this assumes that ND filters contribute negligibly to instrumental polarization, and so this measurement can be applied to observations with any/no ND.

1090894 - Given 1) a set of Stokes vectors computed from the polarimetric observations on a minimum of 3 polarimetric standards (1 unpolarized and 2 polarized) which fall on the same EXCAM pixel, 2) Wollaston, CFAM filter, and ND filter configuration in use during that data collection, and 3) the theoretical Stokes vector for each of those standard targets, the CTC GSW should compute an estimate of the optical system Mueller matrix, which is a single (3x3 floating point) matrix, spatially invariant in EXCAM.

NOTE: this assumes that instrumental polarization effects (depicted by the Mueller matrix) are field-invariant, and that none of the targets are circularly polarized (which is reasonable for relevant astrophysical polarization sources).

Dither during data collection is used to collect extra data to ensure that there are calibrator observations at the same pixel even across different stars. Not all polarimetric calibrator frames will end up being used to compute the Mueller matrix.

Pipeline requirements

- Spectroscopic Calibrations

1090895 - Given (1) a set of cleaned, prism-dispersed sub-band EXCAM images of an unocculted star, (2) the relative image shifts between sub-bands due to filter wedge angles, and 3) 2D templates for each sub-band image, the CTC GSW should estimate (1) a third-order polynomial fit to the spectral dispersion profile (displacement on EXCAM detector array versus wavelength, "prism dispersion scale"); and (2) the orientation angle of the spectral dispersion axis on EXCAM.

NOTE: templates should account for sub-band filter transmission.

1090896 - Given (1) a cleaned image of a prism-dispersed DM satellite spot observed through the FSAM slit mask and the narrowband calibration filter (Filter 3D for Band 3), (2) the relative image shifts between bands due to filter wedge angles, (3) the FPAM, FSAM, and DPAM positions associated with that cleaned image, and (4) a prism dispersion scale and orientation calibration, CTC GSW should compute a wavelength-to-pixel map ("wavelength solution").

1090897 - Given (1) a series of cleaned images of a prism-dispersed unocculted star observed through the FSAM slit mask, observed with the same CFAM filter, and acquired over a grid of FSM offsets and (2) an estimate of the spectroscopic target source position on EXCAM and its alignment error from the FSAM slit, the CTC GSW should identify the dispersed star image whose PSF-to-FSAM slit alignment most closely matches that of the target source.

1090898 - Given a cleaned EXCAM image of an unocculted, prism-dispersed star observed through the narrowband filter (Filter 3D) and FSAM slit, the CTC GSW should compute the spectroscopic line spread function by integrating the signal along the cross-dispersion axis and then extracting the result to a 1-dimensional array in which each element is mapped to a wavelength along the dispersion axis.

"Unofficial" Pipeline Requirements

Pages / Roman Coronagraph CPP Home / Data Reduction and Simulation Working Group

Edit · View inline comments · Save for later · Watching · Share · ···

Corgi-DRP Implementation Document

Created by Max Millar-Blanchaer, last modified by Julia Milton on Aug 21, 2024

Max Millar-Blanchaer, Jason Wang, Marie Ygouf, Vanessa Bailey, Cynthia Wong, Alex Greenbaum

Apr 9, 2024

Revision Draft 0.1

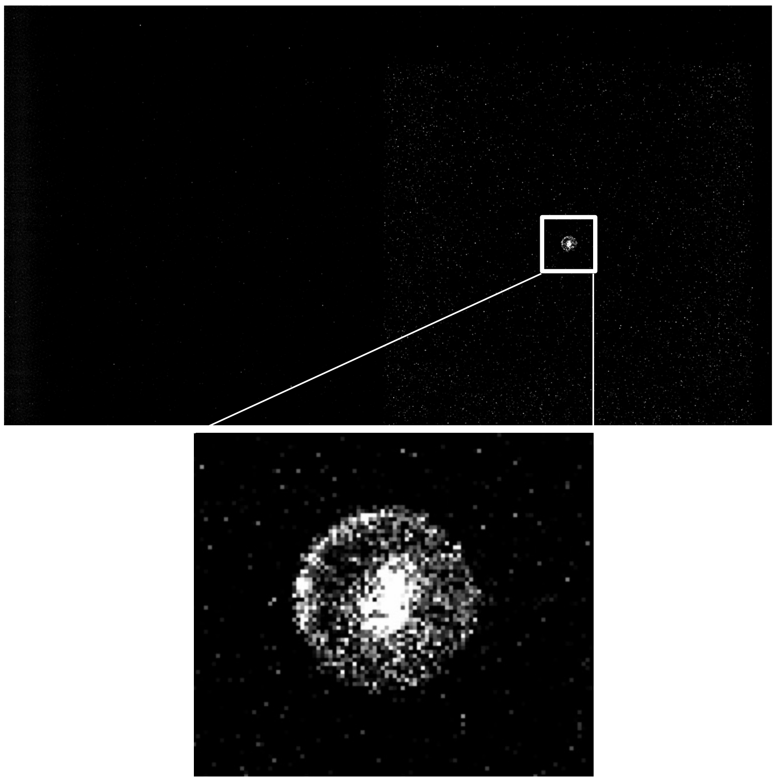
Document Scope

The purpose of this document is to provide written documentation of the CORGI-DRP as-built implementation. It is intended to be used internally by the PS, CTC, SSC and CPP teams in order to coordinate the various teams efforts. It is a live document and will evolve as the DRP evolves. It does not serve to define any formal requirements for anyone.

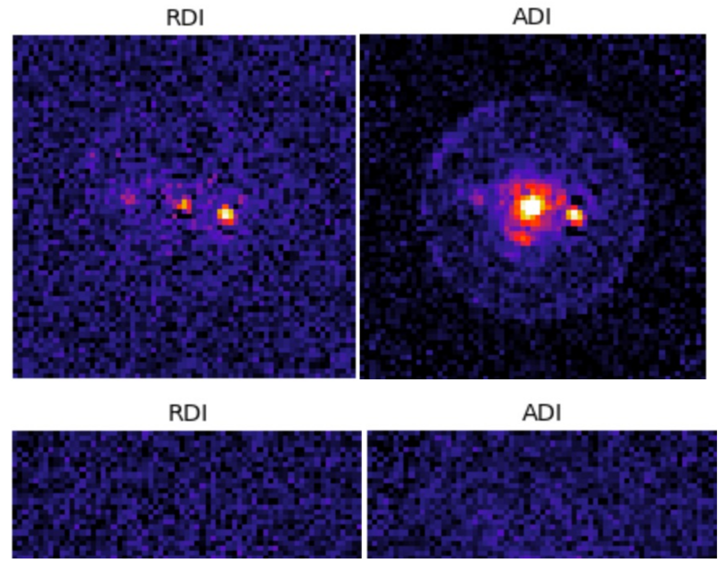
1. Given a list of L1 fits files and the associated CPGS xml, the DRP must identify, populate, and execute the appropriate recipe from the list of recipes in Table 1.
   1. Given a list of locally-available calibration files, the DRP must automatically identify which calibration files should be used to populate the recipe file or identify that the appropriate calibration files are not present locally
   2. The DRP must produce calibration files from L1 FITS files from calibration sequences.
2. The DRP must store the history of data reduction steps in the header of any output files.
   1. The DRP must record the calibration files used in the header of processed files
3. Other users (including those beyond the CTC) should be able to create their own recipe files for other use cases beyond those listed in Table 1.

Introduction - Pipeline Objectives

Going from



To



No
MUF

F

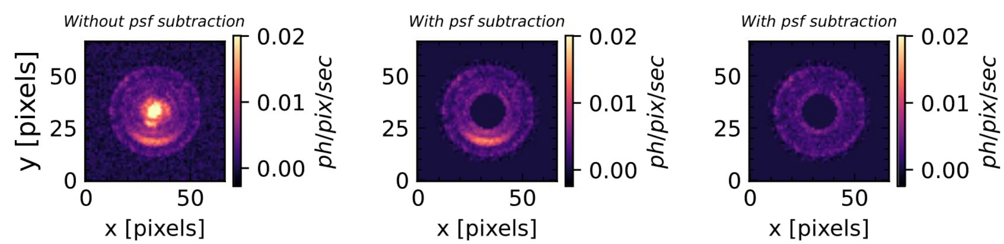


Figure 13. Left: the I_0 component of ϵ Eridani is scaled a hundred times fainter and added to the speckle field noise. Center: the I_0 component is processed with the speckle field images. Right: speckle field image without the I_0 injected.

Outline

- **Introduction**
  - Pipeline objectives
  - Stakeholders PS/CTC/SSC/CPP and project organization
- **Overall Implementation Strategy of the DRP**
  - Pipeline requirements
    - Data Levels
    - TTR5 vs Goals vs Calibration Products Requirements
  - Architecture
    - II&T Legacy
    - Putting DRP into context - Production vs standalone software
    - Standalone software architecture
    - Interfaces for production software
  - Testing and Validation
- **Roadmap and Next Steps**
  - Releases schedule
  - Details of current and future releases
  - Progress update
- **Summary**