JWST TECHNICAL REPORT

Title: A PSF library for coronagraphy with JWST	Doc #: JWST-STScI-004774, SM-12 Date: 21 March 2017 Rev: -
Authors: C. Stark, L Pueyo and The Coronagraph Working Group — Phone: 410-338-4895	Release Date: 12 May 2017

1 Abstract

To image faint objects hiding in the glare of their host stars, JWST coronagraphs will be used to block out a large fraction of light from the star. Remaining starlight (stellar PSF) will be removed through a process called PSF subtraction, which requires a measurement of a reference star PSF. The precision with which astronomers can measure reference PSFs and perform this subtraction determines the faintest detectable objects. Because of the wavefront errors associated with the JWST segmented aperture, the precision of this PSF subtraction will be critical to JWST coronagraph science (Krist et al. 2007, Boccaletti et al. 2015). A library of reference PSFs that grows over the lifetime of JWST can greatly improve this precision, increasing the scientific productivity of the JWST coronagraphs. In this document, we describe the concept for such a library. We describe the functionality of a searchable coronagraph PSF library and recommend a list of data products and metadata to be included in the library.

2 Introduction

To image faint objects, such as planets, hiding in the glare of their host stars, JWST coronagraphs will be used to block out a large fraction of light from the star. However, the coronagraph is not perfect and some starlight will continue to leak through, obscuring otherwise detectable objects. We refer to this leaked starlight as the stellar PSF.

To remove the unwanted remaining starlight, astronomers literally subtract the stellar PSF from the science image. To do this, astronomers must first measure the stellar PSF, which is typically done by observing another reference star assumed not to have similar faint objects nearby. The accuracy and precision with which astronomers can perform this subtraction determines the dimmest object that can be detected and will ultimately determine the scientific productivity of the JWST coronagraphs.

Operated by the Association of Universities for Research in Astronomy, Inc., for the National Aeronautics and Space Administration under Contract NAS5-03127

Much effort is being devoted to measuring the PSFs of reference stars as precisely as possible (e.g., the CWG is studying how multiple PSF measurements over a small grid of dithers can improve the coronagraph performance). Recent advances in PSF subtraction and data analysis techniques have revealed a method that requires no additional observation time and improves the science output of all coronagraph observations: a communal PSF library. In this scenario, an observer has access to all publicly available PSFs measured by all other observers, not just those conducted directly by the observer. This allows one to significantly improve PSF subtraction by creating an optimized reference PSF, formed from a combination of those available in the PSF library.

Figure 1 illustrates the PSF subtraction process using a reference PSF library. The science target observation is shown in the upper-left. In the lower left, a collection of reference PSFs, drawn from the library, provide a good match to the science observation. By decomposing each reference PSF into its fundamental modes, one can form an optimized reference PSF, shown in the top middle, using codes like KLIP (Choquet et al. 2014). By subtracting the optimized reference PSF from the science observation, one can obtain the reduced image at top-right, showing the detection of 3 faint point sources next to the star.

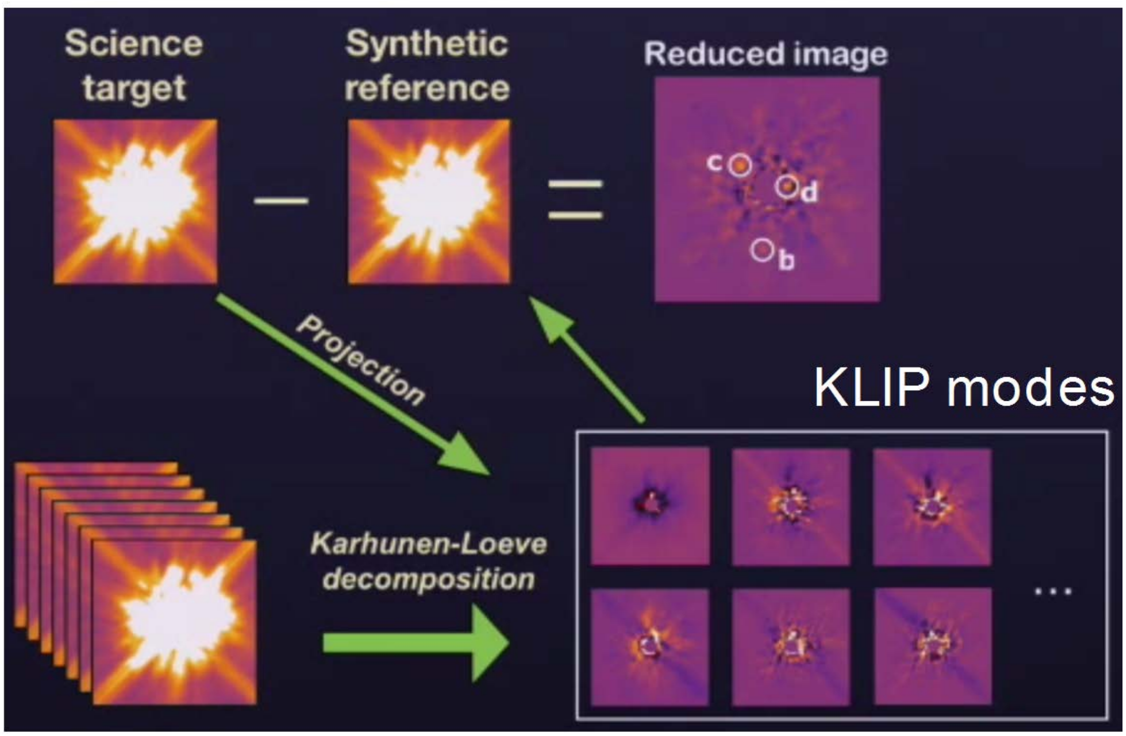

Figure 1 Illustration of the PSF subtraction process using an optimized reference PSF formed from a PSF library. A subset of reference PSFs is drawn from the PSF library (lower-left) and decomposed into their basic modes (lower-right) to form an optimized PSF (top-middle). Subtracting this optimized PSF from the science image (top-left) enables the detection of faint sources (top-right)

This PSF optimization process, and the utility of such a PSF library, has already been proven with HST NICMOS observations; exoplanets and debris disks that went undetected at the time of their NICMOS observations can be retrieved a posteriori by assembling the entire history of the NICMOS coronagraph and using it as a library of references (Lafrenière et al. 2009, Soummer et al. 2011, 2014). Figure 2 shows three example new companion candidates that went undetected for years after their initial NICMOS observations, but were recently detected using a PSF library (Choquet et al. 2014).

However, building such a library is a time-consuming, tedious process for any single user and the astronomical community only benefits if all are allowed access. Additionally, such a library needs to be constructed and updated on a time scale of months, given the limited lifetime of JWST. Thus, the Coronagraph Working Group recommends the implementation of a publicly accessible coronagraphic PSF library in order to support both the calibration pipeline and the observation planning tools.

Having a PSF library available for JWST coronagraph observations would:

- Maximize the science return of executed coronagraph programs by providing the observer and the pipeline with the all the relevant instrument knowledge to calibrate systematic noise in scientific exposures.
- Speed up JWST coronagraphic discoveries, enabling timely follow-up with the limited-lifetime observatory.
- Streamline as efficiently as possible the scientific flow from observations to publications.
- Provide an empirical baseline to assess the instrument performance predicted by observation planning tools, allowing us to improve JWST modeling and observation planning in the future.

This technical report outlines the high-level needs associated with developing such a library of coronagraphic PSFs for JWST. In Section 3 we provide an overview of the envisioned PSF library and several use case scenarios. In Sections 4 and 5 we provide a list of the data and metadata required and describe the associations between them.

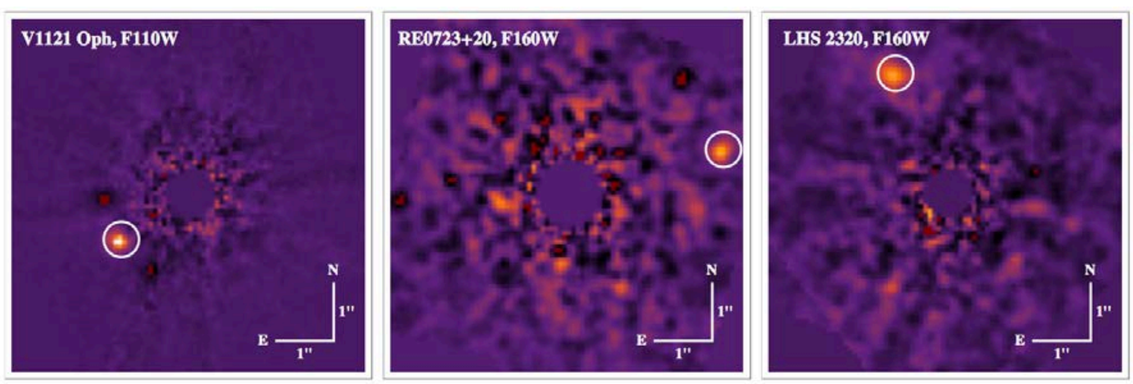

Figure 2 Three example companion candidates that went undiscovered in their initial NICMOS observations, but were recently revealed using a large PSF library (Choquet et al. 2014).

3 Overview of PSF library

3.1 Recommendations regarding concept, not implementation

By "PSF library," the Coronagraph Working Group (CWG) envisions a dynamic collection of coronagraph data that expands in time as observations are made and is:

1. Periodically used by the coronagraph pipeline to reprocess and improve existing coronagraph data publicly available on the MAST archive.
2. Accessible to the public via a PSF library query to deliver customized compiled data packages.

The CWG recognizes that all the data and information discussed in this document will be stored at the Science and Operations Center as part of the archival and engineering tracking tools currently being developed for JWST; the CWG does not expect an independent archive. The CWG envisions that the "collection of coronagraph data" discussed above can be reference files stored in the Calibration Reference Data System (CRDS). Although these files will be available via MAST, storing them in CRDS would allow more flexibility and also the ability to store only PSFs that have been certified by the Coronagraph WG. Having them in CRDS will also guarantee the coronagraph data are recalibrated whenever a new PSF quality dataset becomes available and is delivered to CRDS. The Calibration Pipeline currently assumes the PSF Library will be in CRDS.

3.2 Periodic updates by the pipeline

New high contrast PSF subtraction methods developed in the literature for HST coronagraph observations have demonstrated the power of processing data using an optimized reference PSF generated from a large set of available reference PSFs. For HST observations, this large set of reference PSFs already exists. For JWST, the set of reference PSFs will start at zero and grow with time. Thus, the pipeline should take advantage of the expanding set of reference PSFs and periodically reprocess the final data products using an updated, re-optimized reference PSF. The PSF library will serve as the pipeline's source of reference PSFs, with metadata links (discussed below) to efficiently optimize a reference PSF for every science observation.

3.3 User queries of the PSF library

The CWG expects three dominant use cases of the PSF library, one of which will require direct queries of the PSF library itself:

1. *Novice user*

   Novice users will likely only require the most up-to-date final data products of the coronagraph pipeline. This includes the pipeline-processed science targets and any associated metadata. The CWG expects that this data could be accessible by a standard MAST archive download.

2. *Intermediate user*

   Intermediate users will want the final data products like novice users, but will also likely want to check the pipeline's processing methods. Intermediate users will therefore desire the final data products, plus all of the exposures of the reference PSFs used by the pipeline and intermediate data products generated by the pipeline. The CWG expects that this data could be accessible by a MAST archive download, with trace-back to the intermediate data products and reference PSFs.

3. *Expert user*
   Expert users may not want to rely on the pipeline reduction, preferring to reduce the data on their own using a handpicked collection of reference PSFs. In this case, users would need to query the PSF library itself based on a variety of qualifying criteria. For example, the user may query for all observations astrophysically similar to a given target and with RMS wavefront error less than a certain amount. Expert users will therefore use the PSF library in CRDS to obtain a larger, filtered set of reference PSFs that match the user's detailed query.

The CWG envisions that, when the PSF library is queried via the CRDS web portal or command line tool, CRDS will include the relevant exposures and make them available for download. Figure 3 illustrates a possible structure for the compiled package. Within the compiled package are a series of data sets, each of which correspond to a given exposure. Each data set will be comprised of metadata relevant to the given exposure, likely contained in the header, along with data products. The data products can include the exposure itself as well as intermediate data products produced/used by the pipeline. For instance the data products could include the background-subtracted coronagraph images, as well as images aligned and reduced by the automated coronagraph pipeline—this gives the user flexibility to "take over" the image reduction at any stage desired.

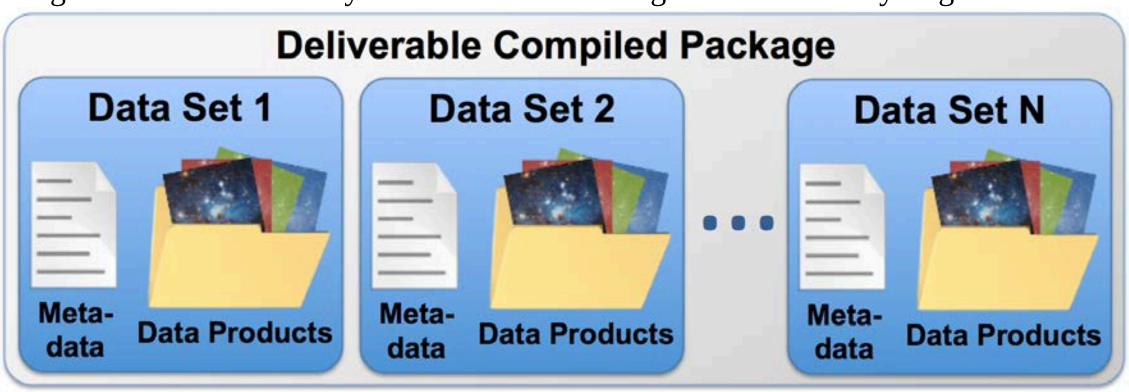

Figure 3 Compiled package delivered to user, generated from a single query of the coronagraph PSF library. For each of N exposures that match the query criteria, the compiled package contains a corresponding data set. Each data set contains metadata and a bundle of data products, which contains the exposure itself and any additional pipeline data products requested.

As an example, let's consider an expert user querying the PSF library database for a set of reference PSFs that all match some detailed criteria. The PSF library would search the metadata of each exposure in the library to check if the exposures meet the search criteria, then compile the data package for all relevant exposures. If the expert user requested all intermediate data products, then for each exposure the PSF library would provide access to a broad array of data products (discussed below).

It is important to note that targets labeled as "science targets" and those labeled as "reference targets" can both serve as reference PSFs returned by a given query. Some targets initially tagged as science targets may not be astrophysically contaminated (no planet or disk), and thus can function as good reference PSFs. Therefore any science target exposures that are no longer proprietary could be included in a query for reference PSFs.

4 Data to be delivered by coronagraph PSF library

To be useful to both novice and expert users, the coronagraph PSF library reference file in CRDS should include access to a wide variety of data products, including standard exposure data, intermediate pipeline data products, and engineering data. This data could likely originate from a broad range of sources, including DMS, the coronagraph pipeline, and WebbPSF. All data contained in the PSF reference library must be publicly available and should be restricted to those data sets that have passed their proprietary period.

Additional data products that may be useful will be made easily accessible to the user by passing along standard associations in the metadata of all Level 3 data products. The metadata will also contain critical information describing the observations, which we detail in Section 4. This metadata could also originate from a broad range of sources, including APT, engineering databases, the coronagraph pipeline, and observers themselves.

The CWG recommends that the information required by the coronagraph PSF library be automatically populated during the execution of a given coronagraph program, from observation planning all the way to data reduction with the pipeline operating in its nominal mode. This is expected to be done as the DMS pipeline queries CRDS. In early cycles, this process might have to be manually checked and curated, with a gradual evolution towards automation.

Here we define the bundle of standard data products that should be accessible via a query to the PSF library. For both science targets and reference targets (and for both NIRCam and MIRI), the following should be delivered to the user for each exposure:

Data 1: the reduced fits image after applying flat field, dark subtraction, cosmic ray rejection, and distortion correction. We note that the use of a contemporaneous sky (background) calibration image is not critical for NIRCam, but may be required for MIRI. This corresponds to the output of the CALIMAGE2 step of the pipeline.

Data 2: the sky/background-subtracted data (for MIRI observations only). These files are produced in the CALCORON3 step of the pipeline.

Data 3: the pre-processed, aligned fits image in which the PSF has been shifted such that the star is located precisely on a reference pixel. These files are produced in the CALCORON3 step of the pipeline. There are only four coronagraph options for MIRI, so a reference pixel will be chosen for each one of the MIRI coronagraphs depending on commissioning data. Since there are multiple NIRCam coronagraph configurations, the reference pixel might have to be adjusted depending on the mode.

For science targets, the following should also be delivered for each exposure:

Data 4: the reduced fits image(s) that corresponds to the nominal result of the pipeline obtained using one or several of the PSF subtraction algorithms (using default reduction parameters) and the associated reference PSFs. By default, the associated reference PSFs will be defined by the programmatic-level coronagraph observing sequence designed in APT. As time passes and the PSF library expands, the number of possible reference PSFs for a given exposure may increase. Thus, this reduced data product (intended for non-expert users) should be periodically updated, on a cycle-to-cycle basis.

Data 5: any associated science calibration exposures (for photometry and astrometry, if they exist).

Data 6: the reduced fits image(s) that correspond to the nominal result of the pipeline calibrated using the science calibration exposures (i.e., counts have been converted to units of contrast, if possible, and the frame origin is positioned at the stellar location of the coronagraph image).

4.1 Simulation data

The PSF library will serve essentially as a repository of empirical coronagraph performance. Efficient, accurate ETC calculations require such a library of instrument responses (Pueyo et al., Pontoppidan et al.) and will generate a library of simulated coronagraph responses. Thus, to simplify operations and avoid two separate coronagraph PSF libraries, we also recommend that the coronagraph PSF library serve as a repository of simulated coronagraph images. This simulated PSF Library reference file can also be accessible via CRDS, which in turn will make them accessible to ETC and any other SOC observation planning tools. Putting them in CRDS will guarantee that the same files are used by all front end and back end systems. A series of representative cases would be stored, each with:

- a simulation of the pre-processed aligned fits image.
- a simulation of the reduced fits images that corresponds to the nominal result of the pipeline obtained using one or several of the PSF subtraction algorithms implemented by default in the automatic SOC pipeline (using default reduction parameters).

4.2 Physical location of the data

The CWG does not have strong recommendations regarding the physical infrastructure associated with storage of this data provided that the functionalities in this document are available.

5 Metadata in the coronagraph PSF library

Users will need some basic information about the exposures, and some users may want additional details to re-reduce the data by creating their own optimized reference PSFs. Thus, the CWG recommends that metadata be maintained and included with each exposure to retain basic exposure data and important associations between data sets. This metadata will also likely serve as the data to be searched during a user query of the PSF Library. In this section, we detail the metadata required.

The CWG envisions two classes of metadata:

1. Telemetry & engineering metadata
2. Science & astrophysical metadata

Below we list these two classes, the metadata required for each, and the sources of the metadata. We have organized them into subcategories, largely defined by their sources. Some of this metadata consists of associations between exposures/other data. For each subcategory of metadata, we provide notes on how the listed associations should be formed and updated.

5.1 Telemetry & engineering (T&E) metadata

Metadata A: Information about OTE/instrument state

Proposed Sources: Engineering database, etc.

Content:

- Last wavefront sensing estimate prior to exposure (overall wavefront RMS and wavefront maps if archived on MAST)
- Association with NIRCam images from last wavefront sensing before exposure
- Wavefront sensing immediately after exposure (overall wavefront RMS and wavefront maps)
- Association with NIRCam images from wavefront sensing immediately after exposure
- Mirror commands sent between sensing and exposures (if necessary)

Metadata B: Information about exposures

Proposed Sources: Data header files, Common Archive Object Model (CAOM) catalog, etc.

Content:

- Exposure time
- Target acquisition telemetry (all info pertaining to location of star on detector)
- Filter selected
- Requested coronagraph configuration, observatory status regarding the coronagraph configurations (e.g., actual location of star on the NIRCam bar).
- Association with detector calibration files (dark, flats, sky background)

5.2 Science & astrophysical (S&A) metadata

Metadata C: Information about the target

Proposed Sources: APT, Simbad, etc.

Content:

- Target coordinates
- Spectral type
- Brightness in the JWST filter of interest
- Proper motion
- Parallax
- Association with astrophysical calibration files (astrometry, photometry)

Metadata D: Information about the observation plan

Proposed Sources: PPS-derived data in DMS, APT, Users, etc.

Content:

- Program ID
- Associations between all science and reference exposures for a given coronagraphic sequence as defined at the programmatic level
- Associations with sky background exposures
- Associations with additional reference exposures in the library outside of the initial coronagraphic sequence

Note: Most associations above flow down from the program design in APT to each exposure. The initial associations between exposures are formed from the program design in APT. However, additional associations can be made as the library expands (last association listed above). Expert users will suggest to the SOC curator that new associations be made between exposures with different sequences/programs to form a larger set of associated reference PSFs.

Metadata E: Information about the astrophysical scene

Sources: Users (initially ingested via curation by the CWG)

Content:

- Information about unexpected astrophysical contaminants: is the source an unexpected binary? Is there a background galaxy in the field of view? Unexpectedly high sky background?
- Information about faint astrophysical structures around the source: presence of a sub-stellar companions or of a disk. This information will be necessary to weed out bad reference PSFs, but will need to be user-generated after the priority period has ended and vetted by the CWG.

5.3 Physical location of metadata

The CWG does not have strong recommendations regarding the physical infrastructure for metadata storage, provided that the functionalities in this document are available. However, it is important to note that the metadata information will come from multiple sources listed above, with varying ease of accessibility. Some of the data may have to be compiled separately from its original source to ensure quick search capability (e.g., metadata contained in headers).

6 Conclusions

We have introduced and discussed the need for a coronagraphic PSF library. The PSF library will be a query-able tool allowing users to obtain a customized data package containing data sets of reference stars and non-proprietary science targets that can be used as PSF references. This data package will deliver the default pipeline products, as well as enable users at all experience levels to better understand how to improve their data reduction and directly deliver the data necessary to do so. As outlined in this document, the data set will contain a range of data products produced at various steps of the pipeline and will include metadata describing the astrophysical scene, program information, and telescope/instrument state—this metadata will come from sources ranging from APT to the pipeline to the end user. We note that the CWG envisions that as more observations are made, the PSF library will grow over time, and data products produced by the coronagraph pipeline should be periodically updated.

7 References

Boccaletti, A., et al. 2015, PASP, 127, 633

Choquet, E., et al. 2014, IAUS, 299, 30

Krist, J. E., et al. 2007, Proc. SPIE, 6693

Lafrenière, D., et al. 2009, ApJ, 694, 148

Soummer, R., et al. 2011, ApJ, 741, 55

Soummer, R., et al. 2014, ApJ, 786, 23