Starshade Exoplanet Data Challenge Reference Documentation

Prepared by Sergi R. Hildebrandt, Renyu Hu, and Mario Damiano, JPL/Caltech

The Starshade Exoplanet Data Challenge

Link to SEDC @ SIP website.

The objectives of analyzing the synthetic images are to detect the embedded planets, measure their locations (astrometry) and brightness and planet-star flux ratio (photometry), and determine the S/N of the planet detection. The objective includes extraction of the planet's spectra when spectroscopic data are provided. The objectives also include to measure the brightness, orientation, and (when possible) structures of the exo-zodiacal light. The list of instrumental and astrophysical components are described in each release notes.

The simulations provided by the SEDC were generated with the public software package **SISTER**[1].

Starshade Systems

The SEDC considers the following starshade systems:

1. Starshade Rendezvous Probe: "blue" 425-552 nm, and "green-red" 615-850 nm passbands. SRP, or Rendezvous, in the sequel.
2. Habitable Exoplanet Observatory: "visible" 450-975 nm, and "NIR" 975-1800 nm passbands. HabEx in the sequel.

Initial data will be broadband imaging and future data will include spectroscopic data, following the expected observing sequence during each mission.

The starshade designs are realistic models with a design contrast of 10^{-10} reflecting imperfections in the petal's shape that are representative of mission requirements. The "contrast" is defined as the ratio between the intensity of starlight averaged in a resolution element at the inner working angle to the peak intensity of the star without the starshade. We expect this ratio at wider angle separation to decrease rapidly and this information is provided in the calibration data ("occulter transmission").

The starshade simulations provided here include the effects of the residual formation flying motion of the starshade. The starshade will be maintained within ~1m with respect to the line of sight in the lateral plane, and this offset is changing in direction and magnitude every few minutes.

The rest of the instrumental parameters for the optical and detector elements of the systems follow the same specifications as the current best estimates of each mission. Visit IPAC website for details about the Nancy Grace Roman Space Telescope, and for the HabEx observatory, the website cited above.

Exoplanetary Systems

The data challenge considers a system formed by a host star with some specific number and types of exoplanets with a particular inclination one 'exoplanetary system'. Each exoplanetary system has a set of different scenarios, corresponding to different epochs and exozodiacal cloud models and intensity.

For the SRP mission, the SEDC considers the following 10 exoplanetary systems associated with 4 stars:

---

[1] http://sister.caltech.edu. S.R. Hildebrandt, S.B. Shaklan, E.J. Cady, M.C. Turnbull (2020) "SISTER: Starshade Imaging Simulation Toolkit for Exoplanet Reconnaissance". JATIS. Accepted for publication.

1. tau Ceti: 1 and 2.
2. epsilon Indi A: 3 and 4 .
3. sigma Draconis: 5, 6, 7, and 8.
4. beta Canum Venaticorum: 9 and 10.
5.

For the HabEx mission, the SEDC considers the following 8 exoplanetary systems associated with the same 4 stars:

6. tau Ceti: 1 and 2.
7. epsilon Indi: 3 and 4 .
8. sigma Draconis: 5, and 6.
9. beta Canum Venaticorum: 7 and 8.

Each system will have a set of scenarios that correspond to 2 independent epochs (‘visits’), 2 models of exozodiacal dust, 3 (undisclosed) intensity levels of exozodiacal dust (lower, intermediate and higher), and 3 (undisclosed) levels of SNR (lower, intermediate and higher). The simulations consider the effects of current best estimates of residual starshade motion due to formation flying and the solar glint. That is, a total of 36 particular scenarios for each exoplanetary system. The choice is made to highlight the relevance of the instrumental effects of the starshade system on the scientific results of the analysis. One the most pressing issue that SEDC may help to resolve is how the SNR of planet detections from processing the retrieval of the synthetic images compare with the idealized, input SNR.

Exozodiacal dust disk: All data in Release 1 assume a smooth exozodiacal dust (meaning that the dust density decreases with semi-major axis smoothly) and do not have resonant structures. We do include the orientation of the exozodiacal dust disk, and the inclination ranges from 0 to 90 degrees, except that for tau Ceti we assume an inclination of 35 degrees consistent with that of the observed outer disk.

In order to fulfill the goals of the SEDC, we may include some specific examples with lower performance than current best estimates in future releases.

Data Format

The simulations and calibration data (PSF and starshade transmission curves) are provided in FITS format. The files contain headers with Keywords containing all relevant disclosed information. Some of the information, such as planet location, flux ratio, albedo, or exozodiacal dust intensity are not provided, as they are part of the blind data analysis goals.

A generic filename has the following syntax:

sister_(R/H)(Scenario #)**_v**(visit #)**_sez**(exozodi intensity level #)**_sn**r(SNR level #)**_passband** (in nanometers)**_r**(release #)**.fits**

An example would be:

`sister_R02_v1_sez2_snr1_0425_0552_nm_r1.fits`

This file would correspond to the second scenario of the Rendezvous mission, during its first visit to this system (tau Ceti) with an intermediate level of smooth exozodiacal dust and the lowest level of SNR. The passband is 425-552 nm, aka ‘blue’ and the data file belongs to the first release. FInally, we note that the exozodiacal dust with resonant structures is labeled as ‘rez’.

We stress that the indices next to **sez/rez** and **snr** are **not** the intensity of the exozodiacal dust in units of solar system intensity, or the actual SNR of the reference planet, but simple labels that refer to the three (undisclosed) levels considered: lower, intermediate and higher values.

Release 1

The first release consists of one observation of the **10** different astrophysical scenarios for **Rendezvous** and its blue passband (**425-552 nm**).

The simulations correspond to **1** epoch (visit). The location of the planet(s) is undisclosed.

The simulations consider a **smooth** model of exozodiacal light with an (undisclosed) intermediate intensity.

We provide **3** levels of SNR. That is, three different integration times. Thus, the total number of files is **30** (recall that the total number of particular scenarios for each exoplanetary system is 36, so release 1 is only a fraction of the whole simulated data that will be released in broad imaging).

The FOV of the simulated data is 67x67 pixels. The angular size per pixel is 21.85 mas, so that the FOV is approximately 1.464 arcsec x 1.464 arcsec. The corresponding physical dimension (in e.g., AU) depends on the star observed. This FOV constitutes a fraction of the entire camera’s frame, which is approximately 22.4 arcsec x 22.4 arcsec. The 67x67 central area contains the more relevant data for the starshade analysis.

We also provide some calibration data, see its corresponding section, that may be helpful to estimate the astrometry and photometry of the simulated data.

The FITS Header of the scenarios provides other information, such as the pixel scale of the images, the passband in consideration, properties of the host star, and other relevant data. A sample of the Keywords can be seen in the Appendix A section.

Some of the limitations of the simulations of this release that will be revisited in forthcoming releases are:

Planets do not move during the observation. For most cases, this is a good approximation. In some cases with long integration times, the planet(s) could move over a very few pixels at most. However, this means that the planet(s) are faint and close to the system capability for detection (broad imaging), not mentioning characterization (spectroscopy). Therefore, it is not a good candidate for a future program. We will include planetary motion when we deal with long integration times, e.g., spectroscopy, although we may share the planet's location within some margin based on expected RV ancillary data.

The detector model is an EMCCD in analogue mode. The starshade throughput and the type of planets considered in this first release allow one to just use the EMCCD detector in Roman in analogue mode instead of photon counting mode. This simplifies the noise simulations that you may also need to perform. In the next releases, we expect to include a photon counting mode that will allow us to compare the results.

The synthetic images in release #1 do not include straylight reflected on the starshade surface from the Earth, the Milky way or other potential straylight sources due to the starshade degradation after micrometeorite collisions. These terms however are estimated to be smaller than the solar glint, which has been included in the synthetic images.

Calibration data

In order to perform the **broad imaging** data analysis, we provide the following system information:

1. **PSF response**: an averaged PSF response of the starshade-telescope system, obtained using SISTER and it is an accurate representation of the expected optical response. The average is a direct mean value of the PSF response at different wavelengths across the passband under consideration[2]. The PSF response is provided at different angular distances from the starshade center. It is normalized to the PSF response far away from the starshade center, which is equivalent to the telescope's response itself. Optical losses, as the one associated with the secondary and the struts in the case of the Roman telescope (Rendezvous) are included in the simulated data and in the value of the star flux (see later).[3]

   The PSF spatial extent is very large in order to provide 99.5% of the energy contained in the full PSF response. When using the PSF, one may select a portion of the central

[2] If the astrophysical signal has a strong dependence on wavelength, the effective PSF would be different. For the case of broad imaging, it may be good enough in order to estimate average values, but we can certainly provide a wavelength dependent PSF if your analysis requires such level of accuracy.

[3] In the case of Rendezvous, if one would like to retrieve the true source counts, there's still a factor 1/0.824664=1.2126 due to the presence of the secondary and the struts supporting it that are obstructing the telescope's aperture. HabEx has an unobstructed telescope design, and there's no additional factor to take into account. We do not expect to use this correction (see the Calibration data section).

response and know how much energy is encircled. Similarly, one may use the file to derive the average FWHM far away from the starshade center.

For instance, or Rendezvous and the 425-552 nm passband, `psf_averaged_0425_0552_nm.fits`.The pixel size, angular distances from the starshade center, and other parameters are found in its Header, see also Appendix B for more information and a figure.

2. **Occulter transmission**: an average transmission for the effect due to the starshade diffraction. We compare the flux that would arrive at the telescope with and without the presence of the starshade. The average is a direct mean value of the transmission at different wavelengths across the passband under consideration, which does not vary much across wavelengths (see footnote). This information can be used to retrieve the actual counts from the sources, removing the 'blocking' effect of the starshade. Only, in regions close to or inside the IWA does its value become significantly different to 1.

   For instance, for Rendezvous and the 425-552 nm passband, the filename is `starshade_averaged_transmission_0425_0552_nm.fits`. See its Header and Appendix C for more information and a figure.

3. **Star flux**: we provide the normalized count rate of the star's flux, integrated over the instrumental passband in consideration, that would be detected in case the starshade were not present (and the detector's parameters set to not saturate the exposures).

   If you divide the estimated counts on the planets, corrected both for the fraction of the energy encircled in your photometric estimation and the starshade transmission, by this value of the star counts, you would derive an estimation of the planet's flux ratio. There is no need to correct for the telescope's blocked area in the case of Rendezvous since this (constant) effect is both included in the star counts and when you estimate the planet counts on the data, so that it cancels out when deriving the ratio. This estimation will be an estimated average of the actual flux ratio that depends on the wavelength dependence of the planet's albedo.

4. **Integration time**: it is found in the FITS header of each simulated file. See Appendix A.

5. **Detector parameters**: they are defined in the FITS header of each simulated file. See Appendix A.

Other calibration data will be included when we release the simulated data with spectroscopy. For instance, the detector's QE, telescope transmission, or the starshade transmission as a function of wavelength. If they were necessary in the study of broad band imaging, we can provide them earlier. Similarly with other system parameters that you may find convenient for your analysis.

FAQ

Q1: Where can I get the simulated data?
A1: You can download them from the SIP website: **FIXME** add link.

Q2: Why does the PSF data have the first/second column swapped?
A2: The FITS files were created with CFITSIO in Matlab. It seems that other programming languages with inverted column/row order may swap the first two indices. The PSF is a square array for each of the angular distances to the starshade center.

Resources for the Participants

FIXME: Slack channel and Mario's contact email.

Acknowledgments

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

If you publish results based partly on the simulations provided here, please acknowledge it in the following form:

1. An acknowledgment statement – "Some of the results in this paper have been derived using starshade simulations from the SISTER package"
2. At the first use of the SISTER acronym, a footnote placed in the main body of the paper referring to the SISTER website – currently http://sister.caltech.edu
3. A reference to S.R. Hildebrandt, S.B. Shaklan, E.J. Cady, M.C. Turnbull (2020) "SISTER: Starshade Imaging Simulation Toolkit for Exoplanet Reconnaissance". JATIS. Accepted for publication.

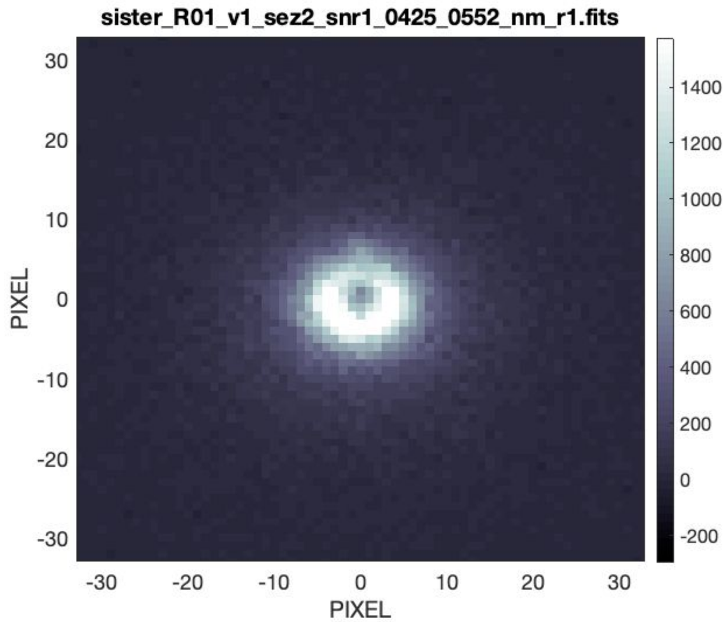
Acronyms

EMCCD: Electron multiplier CCD
FOV: Field of View
HabEx: Habitable Exoplanet Observatory
IWA: Inner Working angle
SEDC: Starshade Exoplanet Data Challenge
SISTER: Starshade Imaging Simulation Toolkit for Exoplanet Reconnaissance

SNR: signal-to-noise ratio
SRP: Starshade Rendezvous Probe

Appendix A: FITS Header for the data simulations

This is an image for the file `sister_R01_v1_sez2_snr1_0425_0552_nm_r1.fits`.



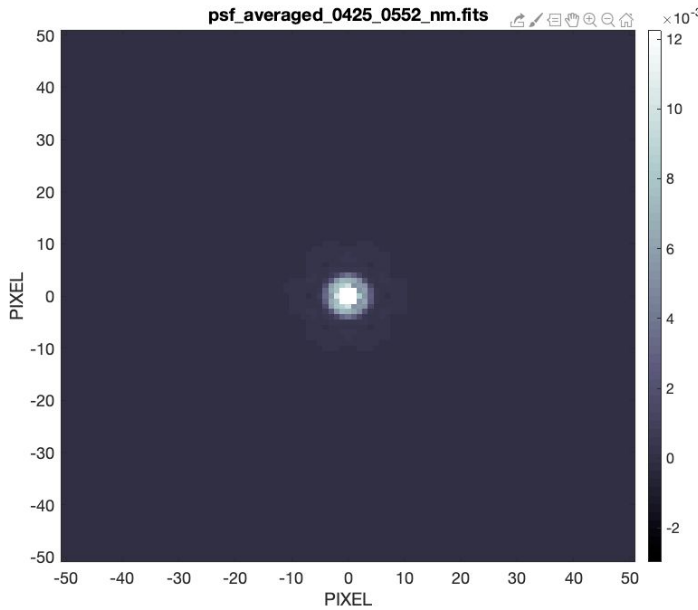
This is the FITS header for the same file `sister_R01_v1_sez2_snr1_0425_0552_nm_r1.fits`. Notice that the total integration time is 1686 seconds (`INTTIME`).

```
{'SIMPLE'  }    {'T'                   }    {' file does conform to FITS standard             '                       }
{'BITPIX'  }    {[                 -64]}    {' number of bits per data pixel                  '                       }
{'NAXIS'   }    {[                   2]}    {' number of data axes                            '                       }
{'NAXIS1'  }    {[                  67]}    {' length of data axis 1                          '                       }
{'NAXIS2'  }    {[                  67]}    {' length of data axis 2                          '                       }
{'EXTEND'  }    {'T'                   }    {' FITS dataset may contain extensions            '                       }
{'COMMENT' }    {0x0 char              }    {'  FITS (Flexible Image Transport System) format is defined in 'Astronomy'}
{'COMMENT' }    {0x0 char              }    {'  and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H
'}
{'SIM'     }    {'R01'                 }    {' starshade system and scenario                  '                       }
{'VISIT'   }    {[                   1]}    {' visit number, aka epoch                        '                       }
{'XCENTER' }    {[                  33]}    {' pixel x-coordinates in 00LL convention         '                       }
{'UNITS'   }    {'normalized counts'   }    {' total number of e- divided by the gain         '                       }
{'FF'      }    {[                   1]}    {' formation flying. 0: not included. 1: included '                       }
{'MINLAM'  }    {[                 425]}    {' minimum simulated wavelength (nanometers)      '                       }
{'MAXLAM'  }    {[                 552]}    {' maximum simulated wavelength (nanometers)      '                       }
{'DELTALAM'}    {[                   5]}    {' PSF basis. Wavelength spacing (nanometers)     '                       }
{'PIXSCALE'}    {[             21.8500]}    {' (mas=milli-arcsec)                             '                       }
```

{'DIAMTEL' } {[2.3600]} {' (meters) ' }
{'YCENTER' } {[33]} {' pixel y-coordinates in 00LL convention ' }
{'OBSCURED'} {[0.4187]} {' effective ratio of secondary&struts to DIAMTEL ' }
{'JITTER' } {[14]} {' pointing jitter RMS (mas) ' }
{'DESIGN' } {'NI2_test_case_1em10'} {' SS design (NI2: Rendezvous, TV3: HabEx) ' }
{'TELDIST' } {[37242]} {' SS-Telescope range (kilometers) ' }
{'STARNAME'} {'tau Ceti' } {' star's common name ' }
{'VSTAR' } {[3.5000]} {' star's V magnitude ' }
{'STARDIST'} {[3.6500]} {' star's distance (parsec, 1 AU=1e3/STARDIST mas)' }
{'STARMASS'} {[0.7830]} {' star's mass divided by solar mass ' }
{'STARRAD' } {[0.7930]} {' star's radius divided by solar radius ' }
{'STARLUM' } {[0.7830]} {' star's luminosity divided by solar luminosity ' }
{'STARTEMP'} {[5780]} {' star's temperature (K) ' }
{'STARFLX' } {[5780]} {' star's total detected flux (normalized counts) ' }
{'EXOZODI' } {'SMOOTH CLOUD' } {' exozodi spatial distribution ' }
{'EXOLVL' } {[2]} {' exozodi intensity *label* ' }
{'SNRLVL' } {[1]} {' SNR level *label* ' }
{'INTTIME' } {[1686]} {' total integration time (sec) ' }
{'FRMTIME' } {[60]} {' integration time of each individual frame (sec)' }
{'NFRAMES' } {[31]} {' actual number of individual frames (integer) ' }
{'DETTYPE' } {'emccd am' } {' detector type: CCD, EMCCD AM or PC ' }
{'DETGAIN' } {[200]} {' detector's gain (dimensionless) ' }
{'ENF2' } {[2]} {' excess noise factor squared (dimensionless) ' }
{'READOUT' } {[100]} {' read-out noise (e/pix/frame) ' }
{'CIC' } {[0.0200]} {' clock induced charges (e/pix/frame) ' }
{'DARKCURR'} {[4.2222e-04]} {' dark current (e/pix/sec) ' }
{'POISSONL'} {[10]} {' Poisson -> Normal (with continuity correction) ' }
{'END' } {0x0 char } {0x0 char }

Appendix B: FITS Header for the PSF response

Here is an image of the PSF at the geometric IWA of the Rendezvous, 425-552 nm passband, `psf_averaged_0425_0552_nm.fits`. One may notice a slight left/right asymmetry due to the starshade diffraction. Close to the starshade center, the PSF response gets more affected by the starshade diffraction. The PSF spatial extent is very large in order to provide 99.5% of the energy contained in the full PSF response. When using the PSF, one may select a portion of the central response and know how much energy is encircled. Similarly, one may use the file to derive the average FWHM far away from the starshade center.



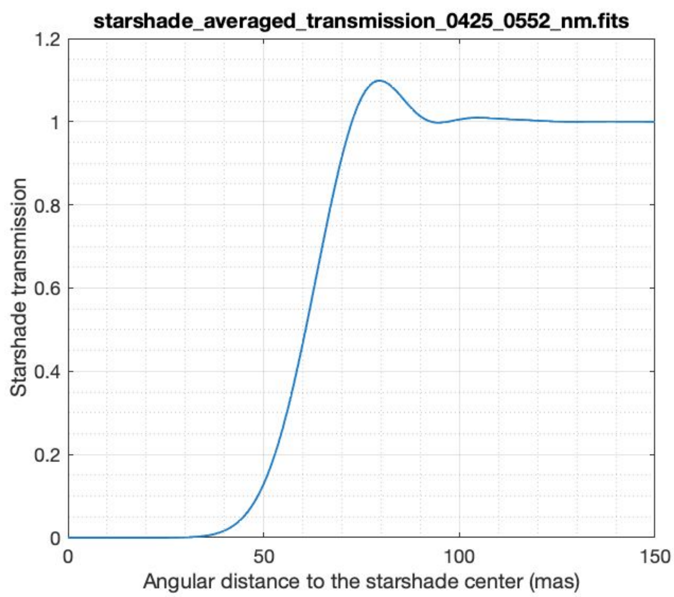
Beware that the first and second indices of the PSF array may be swapped (See FAQ). In the example below: , the PSF array should be 151x103x103.

```
{'SIMPLE'  }    {'T'                 }    {' file does conform to FITS standard             '                        }
   {'BITPIX'  }    {[               -64]}    {' number of bits per data pixel                  '                        }
   {'NAXIS'   }    {[                 3]}    {' number of data axes                            '                        }
   {'NAXIS1'  }    {[               103]}    {' length of data axis 1                          '                        }
   {'NAXIS2'  }    {[               151]}    {' length of data axis 2                          '                        }
   {'NAXIS3'  }    {[               103]}    {' length of data axis 3                          '                        }
   {'EXTEND'  }    {'T'                 }    {' FITS dataset may contain extensions            '                        }
   {'COMMENT' }    {0x0 char            }    {'  FITS (Flexible Image Transport System) format is defined in 
'Astronomy'}
   {'COMMENT' }    {0x0 char            }    {'  and Astrophysics', volume 376, page 359; bibcode: 
2001A&A...376..359H '}
   {'UNITS'   }    {'Dimensionless'     }    {' normalized <PSF> at different angles from the S'                        }
   {'PSFSIZE' }    {'103x103'           }    {' pixels                                         '                        }
   {'NDIST'   }    {[               151]}    {' number of radial positions                     '                        }
   {'MINDIST' }    {[                 0]}    {' minimum angular distance (mas=milli-arcsec)    '                        }
   {'MAXDIST' }    {[               150]}    {' maximum angular distance (mas=milli-arcsec)    '                        }
   {'DDIST'   }    {[                 1]}    {' step between consecutive PSF (mas=milli-arcsec) '                        }
   {'MINLAM'  }    {[               425]}    {' minimum PSF wavelength (nanometers)            '                        }
   {'MAXLAM'  }    {[               552]}    {' maximum PSF wavelength (nanometers)            '                        }
   {'DELTALAM'}    {[                 5]}    {' PSF basis. Wavelength spacing (nanometers)     '                        }
   {'PIXSCALE'}    {[           21.8500]}    {' (mas=milli-arcsec)                             '                        }
   {'DIAMTEL' }    {[            2.3600]}    {' (meters)                                       '                        }
   {'XCENTER' }    {[                51]}    {' pixel x-coordinates in 00LL convention         '                        }
   {'YCENTER' }    {[                51]}    {' pixel y-coordinates in 00LL convention         '                        }
   {'OBSCURED'}    {[            0.4187]}    {' effective ratio of secondary&struts to DIAMTEL '                        }
   {'JITTER'  }    {[                14]}    {' pointing jitter RMS (mas)                      '                        }
```

{'DESIGN' } {'NI2_test_case_1em10'} {' SS design (NI2: Rendezvous, TV3: HabEx) '
}
{'TELDIST' } {[37242]} {' SS-Telescope range (kilometers) ' }
{'END' } {0x0 char } {0x0 char

Appendix C: FITS Header for the starshade transmission

Here's a figure for the occulter (aka starshade) transmission for Rendezvous and the 425-552 nm passband, `starshade_averaged_transmission_0425_0552_nm.fits`:



starshade_averaged_transmission_0425_0552_nm.fits:

24x3 cell array

{'SIMPLE' } {'T' } {' file does conform to FITS standard ' }
{'BITPIX' } {[-64]} {' number of bits per data pixel ' }
{'NAXIS' } {[2]} {' number of data axes ' }
{'NAXIS1' } {[151]} {' length of data axis 1 ' }
{'NAXIS2' } {[2]} {' length of data axis 2 ' }
{'EXTEND' } {'T' } {' FITS dataset may contain extensions ' }
{'COMMENT' } {0x0 char } {' FITS (Flexible Image Transport System) format is defined in 'Astronomy'}
{'COMMENT' } {0x0 char } {' and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H '}
{'UNITS1' } {'angular distance from SS center'} {' (mas=milli-arsec) ' }

```
  {'UNITS2'  }    {'dimensionless'                }    {' normalized SS transmission at different angles '                    }
  {'NDIST'   }    {[                          151]}    {' number of radial positions                     '                    }
  {'MINDIST' }    {[                            0]}    {' minimum distance (mas=milli-arcsec)            '                    }
  {'MAXDIST' }    {[                          150]}    {' maximum distance (mas=milli-arcsec)            '                    }
  {'DDIST'   }    {[                            1]}    {' step between consecutive PSF (mas=milli-arcsec) '                   }
  {'MINLAM'  }    {[                          425]}    {' minimum PSF wavelength (nanometers)            '                    }
  {'MAXLAM'  }    {[                          552]}    {' maximum PSF wavelength (nanometers)            '                    }
  {'DELTALAM'}    {[                            5]}    {' PSF basis. Wavelength spacing (nanometers)     '                    }
  {'PIXSCALE'}    {[                      21.8500]}    {' (mas=milli-arcsec)                             '                    }
  {'DIAMTEL' }    {[                       2.3600]}    {' (meters)                                       '                    }
  {'OBSCURED'}    {[                       0.4187]}    {' effective ratio of secondary&struts to DIAMTEL '                    }
  {'JITTER'  }    {[                           14]}    {' pointing jitter RMS (mas)                      '                    }
  {'DESIGN'  }    {'NI2_test_case_1em10'          }    {' SS design (NI2: Rendezvous, TV3: HabEx)        '
}
  {'TELDIST' }    {[                        37242]}    {' SS-Telescope range (kilometers)                '                    }
  {'END'     }    {0x0 char                       }    {0x0 char                                                              }
```