



SSDC-DP-017

SPHEREx Explanatory Supplement

Version: 1.0

Authors:

SPHEREx Science Data Center Team: (Rachel Akeson, Gregory Dubois-Felsmann, Andreas Faisst, Tamim Fatahi, Tatiana Goldina, Yuna Kwon, Dan Masters, Christina Nelson, Roberta Paladini, Gaby Torrini, Phani Velicheti)

SPHEREx Science Team Contributors: (Matthew Ashby, Yoonsoo Bach, James Bock, Sean Bryan, Brendan Crill, Olivier Doré, C. Darren Dowell, Candice Fazar, Howard Hui, Phil Korngut, Chi Nguyen, Michael Werner, Michael Zemcov)

Contents

- 1 Mission Overview** **4**
 - 1.1 Mission Objectives 4
 - 1.2 Spacecraft 4
 - 1.3 Payload 5
 - 1.4 Survey Strategy and Spectral Accumulation 7
 - 1.5 Mission and Science Operations Summary 9
 - 1.6 Acknowledgments 10
- 2 The SPHEREx Data Products** **11**
 - 2.1 Spectral Images 11
 - 2.2 Image Anomalies and Features 13
 - 2.3 Known Limitations 17
- 3 The SPHEREx Data Processing** **20**
 - 3.1 Data Ingestion and Raw Image Processing 20
 - 3.2 Processing Steps 21
 - 3.2.1 Nonlinearity Correction 21
 - 3.2.2 Variance Map 22
 - 3.2.3 Persistence Masks 23
 - 3.2.4 Bad Pixel Flagging 25
 - 3.2.5 Dark Subtraction 27
 - 3.2.6 Absolute Gain Correction 27
 - 3.2.7 Astrometry 27
 - 3.2.8 Source Masks 29
 - 3.2.9 Outlier Pixel Detection 31
 - 3.2.10 Zodi Prediction 31
 - 3.2.11 Level 2 Data Quality Assurance (DQA) 32
 - 3.3 Photometric Measurements 32
- 4 Calibrations** **35**
 - 4.1 Calibration Products 35
 - 4.1.1 Absolute Gain Matrix 35
 - 4.1.2 Exposure-Averaged Point Spread Functions (PSFs) 35
 - 4.1.3 Dark Current 35



4.2	Calibration Procedures	36
4.2.1	Photometric/Absolute Gain Calibration Derivation	36
4.2.2	Astrometric Calibration Derivation	38
4.2.3	PSF Measurement	40
4.3	Additional Products and Inputs	43
4.3.1	Spectral WCS	43
4.3.2	Additional Products	45
A	Spectral Image Header Example	49
B	Absolute Gain Matrix Header Example	54
C	Average Point Spread Function (PSF) Header Example	55
D	List of Acronyms	57



Change Record

Version	Date	Description	Owner
0.1	2025-04-03	Initial draft.	Yuna Kwon
1.0	2025-07-02	Quick-Release Version Release.	Yuna Kwon

(Document generated on July 1, 2025 from commit df85955-dirty of 2025-07-02.)

1 Mission Overview

1.1 Mission Objectives

The Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) is a NASA Astrophysics Medium-class Explorer (MIDEX) mission, designed to perform the first all-sky spectral survey in the optical to Near-infrared (NIR) range, covering wavelengths from 0.75 to 5.0 μm . SPHEREx launched on 12 March 2025 at 03:10:12 UTC aboard a SpaceX Falcon 9 rocket from Vandenberg Space Force Base.

On May 1, 2025, the spacecraft began regular science operations following an In-Orbit Checkout (IOC) period during which the flight and ground systems were commissioned, and key survey parameters were optimized. Over its nominal 25-month science mission, SPHEREx will conduct four complete sky surveys, yielding 1.4 trillion spectral-spatial elements (voxels) per all-sky pass. The mission will construct a comprehensive three-dimensional cosmos map, enabling transformative insights into the origin and evolution of the universe. The mission addresses three principal science themes aligned with NASA's strategic astrophysics priorities [Doré et al., 2014]:

- **Probing Inflation and the Early Universe:** SPHEREx will acquire low-resolution spectra for over 450 million galaxies, enabling a three-dimensional reconstruction of the large-scale structure of the Universe across an unprecedented cosmological volume. These data will constrain primordial non-Gaussianity via the f_{NL} parameter, with a target sensitivity of $\sigma(f_{\text{NL}}) \leq 1$, providing critical tests to distinguish between single- and multi-field inflationary models.
- **Tracing the Origin of Planetary Systems and Biogenic Ices:** SPHEREx will perform the largest spectroscopic survey of ice absorption features in circumstellar environments, observing young stellar objects, protoplanetary disks, and molecular clouds throughout the Milky Way. The mission will collect over 20,000 high-quality ice spectra, enabling statistically robust correlations between ice composition and environmental conditions.
- **Exploring Galaxy Evolution via the Extragalactic Background Light (EBL):** Through deep spectral imaging of the ecliptic poles, SPHEREx will measure spatial power spectra of EBL fluctuations. These data will disentangle the contributions of intra-halo light, faint dwarf galaxies, and sources from the Epoch of Reionization, providing new constraints on early galaxy populations and the cosmic star formation history.

In addition to its primary science objectives, SPHEREx will generate an archival dataset capturing a diverse range of astronomical phenomena [Crill et al., 2020, Ivezić et al., 2022, Alibay et al., 2023], including:

- $\gtrsim 1$ billion galaxy spectra
- $\gtrsim 450$ million galaxy redshifts
- $\gtrsim 1$ million quasar spectra
- $\gtrsim 100$ million stellar spectra
- $\gtrsim 600,000$ exoplanet target stars
- $\gtrsim 700,000$ ice absorption spectra
- $\gtrsim 100,000$ asteroid and comet spectra

1.2 Spacecraft

The SPHEREx spacecraft is based on the Ball Configurable Platform (BCP-100) bus, previously utilized in missions such as Imaging X-ray Polarimetry Explorer (IXPE) [Weisskopf et al., 2022]. Designed for operations in low Earth orbit (LEO), the platform delivers high pointing accuracy, thermal stability, and efficient data throughput.

Throughout the SPHEREx mission, the spacecraft operates in a sun-synchronous polar orbit. SPHEREx's operations are governed by geometric pointing constraints to mitigate Sun, Earth, and Moon contamination. Key spacecraft orbital parameters are summarized in Table 1.

It achieves high-precision three-axis stabilization through a combination of star trackers, a gyro-based Scalable Space Inertial Reference Unit (SSIRU), GPS, four reaction wheels, and three magnetic torque rods. Telemetry and command operations utilize the S-band for uplink (2 kbps) and low- or high-rate telemetry (32 kbps and 2 Mbps, respectively), while science data are transmitted via Ka-band downlink at 600 Mbps. Power is supplied by a body-fixed solar array paired with a 22.5 Ah lithium-ion battery, providing uninterrupted operation even during eclipse seasons. An onboard recorder stores up to eight days of science data (approximately 1074 Gbits), ensuring robust data capture between ground contacts.

Table 1: Orbital parameters of the SPHEREx spacecraft

Property	Value
Mean semi-major axis	7,037.41 km
Eccentricity	0.00034
Inclination	97.951°
Longitude of ascending node	27.06°
Argument of periapsis	121.8°
Time to periapsis	1,987 sec
Mean Local Time of Ascending Node (MLTAN)	6:00 AM
Orbital period	98 min

Key spacecraft functionalities are provided by the following subsystems: the Attitude Determination and Control Subsystem (ADCS), which maintains precise orientation; the Command and Data Handling (CDH) subsystem, including spacecraft avionics; Flight Software (FSW), responsible for receiving and executing survey sequences, as well as spacecraft control; the Electrical Power Distribution Subsystem (EPDS), which ensures continuous power delivery; the Telecommunications subsystem, supporting uplink and downlink operations; the Thermal Control subsystem, maintaining operating temperatures for critical electronics; and the Spacecraft Structure, which supports both the spacecraft bus and payload.

The spacecraft is not equipped with onboard propulsion. Consequently, orbital decay follows the natural evolution of the orbit, predominantly influenced by atmospheric drag. End-of-mission de-orbiting is achieved passively, with atmospheric re-entry expected within 25 years following the conclusion of science operations.

Detailed descriptions of each subsystem and its associated software architectures are provided in Alibay et al. [2023].

1.3 Payload

The SPHEREx payload includes a wide-field cryogenically cooled telescope and spectrophotometric imaging system optimized for spectral mapping. It features a 20-cm effective aperture, all-aluminum free-form Three Mirror Anastigmat (TMA) telescope with a broad Field of View (FoV) of $3.5^\circ \times 11.3^\circ$. The optical path is split by a dichroic beam splitter, allowing simultaneous observations in short- and mid-wave infrared bands. The entire assembly is thermally isolated from the spacecraft to maintain radiometric stability. With no moving parts beyond the one-time deployment of an aperture cover during IOC, the design emphasizes robustness, maximizing spectral throughput and operational reliability. A schematic illustration of the optical light path and an example of Linear Variable Filter (LVF) spectroscopy toward the Andromeda Galaxy are shown in Figure 1. Animation illustrating collecting signals through LVF is available on the SPHEREx/California Institute of Technology (Caltech) official website¹.

The SPHEREx instrument is organized into three primary subsystems: the optical subsystem (telescope, beam splitter, and Focal Plane Assemblies (FPA)s), the thermal subsystem (passive radiative cooling structure), and the Instrument Control Electronics (ICE) electronics (detector readout, onboard processing, and storage management).

Cryogenic telescope and detector temperatures are achieved entirely through passive means. A three-stage V-groove radiator system, comprised of nested conical photon shields, supplemented by a dedicated focal plane radiator, effectively dissipates heat to space, reducing telescope thermal emission, suppressing

¹<https://spherex.caltech.edu/page/survey>

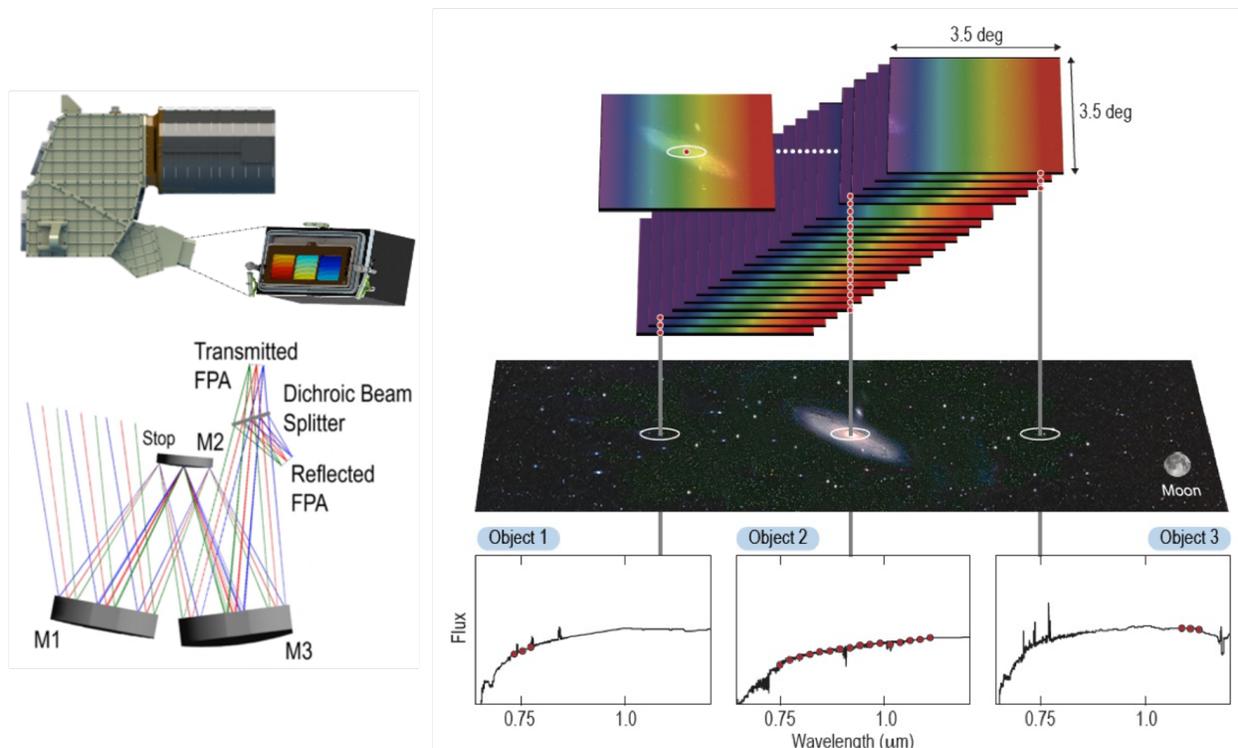


Figure 1: Light path through the SPHEREx optical system and an example illustration of LVF spectroscopy targeting the Andromeda Galaxy. (Credit: Image courtesy of SPHEREx-JPL/Caltech)

detector dark current and facilitating thermal stability throughout the mission. Key instrument parameters are summarized in Table 2 [Crill et al., 2020] and the normalized spectral resolution is shown in Figure 2. Figure 3 presents the preliminary expected all-sky point source sensitivity after four full-survey passes, assuming a median sky background photocurrent.

Table 2: SPHEREx Instrument Parameters

Parameter	Value
Effective aperture	20 cm
Focal ratio	f/3
Field of view	$2 \times (3.5^\circ \times 11.3^\circ)$; dichroic split
Wavelength coverage and design resolution	0.744–5.002 μm
<ul style="list-style-type: none"> • Band 1: $\lambda = 0.744\text{--}1.116 \mu\text{m}$; $R = 39$ • Band 2: $\lambda = 1.099\text{--}1.651 \mu\text{m}$; $R = 41$ • Band 3: $\lambda = 1.636\text{--}2.421 \mu\text{m}$; $R = 41$ • Band 4: $\lambda = 2.423\text{--}3.822 \mu\text{m}$; $R = 35$ • Band 5: $\lambda = 3.809\text{--}4.420 \mu\text{m}$; $R = 112$ • Band 6: $\lambda = 4.412\text{--}5.002 \mu\text{m}$; $R = 128$ 	
Detectors	3 \times Hawaii-2RG 2.6 μm cutoff 3 \times Hawaii-2RG 5.3 μm cutoff
Read mode	Sample-Up-the-Ramp (SUR)
Frame time	1.5349 s
Exposure time	113.58 s
Pixel scale	6.15" \times 6.15"

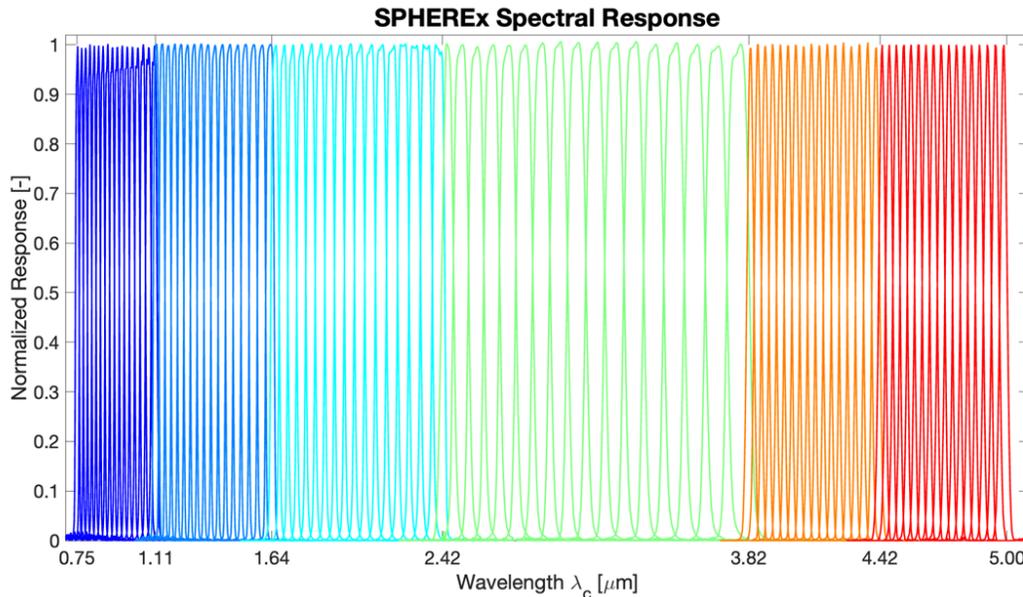


Figure 2: The measured normalized spectral response across the SPHEREx wavelength range.

The focal plane is divided into two FPAs, each consisting of three Teledyne H2RG HgCdTe $2k \times 2k$ detector arrays. A dichroic beam splitter separates the short-wavelength and long-wavelength channels: the Short-wave Infrared (SWIR) FPA covers $0.75\text{--}2.44 \mu\text{m}$ (in reflection) and operates at 62 K, while the Mid-wave Infrared (MWIR) FPA spans $2.40\text{--}5.01 \mu\text{m}$ (in transmission) and operates at 45 K. Each detector is paired with an LVF, positioned above the array, which allows each pixel to sample a narrow spectral bandpass. As the sky drifts across the field of regard given Earth-, Sun-, and Moon- avoidance constraints, successive exposures at offset positions reconstruct full low-resolution spectra for each celestial source.

The detectors employ SUR readout mode to optimize Signal-to-Noise (S/N) performance. Custom-built Video8 Application-Specific Integrated Circuit (ASIC) boards, developed by SPHEREx/Caltech, perform low-noise readout, real-time photocurrent estimation, and onboard data compression. The ICE system manages digitized sample processing and compression before transferring the data to dedicated onboard storage for subsequent downlink. Detailed specifications of the detectors are provided in Crill et al. [2020], and further descriptions of the ICE hardware and firmware can be found in Alibay et al. [2023].

1.4 Survey Strategy and Spectral Accumulation

SPHEREx takes observations that deliver the required all-sky and deep field coverage. We plan these observations using the Survey Planning Software (SPS) developed for SPHEREx. Unlike other surveys with repeating geometric scan strategies, the unique requirements of the SPHEREx mission led us to develop the SPS to plan observations on-the-fly that deliver high time on science and high sky coverage. A paper describing this software in detail is currently in preparation. Among other capabilities, the SPS plans science observations, pauses observations for downlinks, mitigates outages due to passage through the elevated radiation environment of the South Atlantic Anomaly, plans slews to avoid bringing the telescope close to the Moon while slewing, and selects a safe place to point in rare cases when no science targets are observable.

The SPS only plans observations that are in the allowable pointing zone. Operating from low-earth orbit, the SPS plans attitudes for SPHEREx that avoid pointing near the Sun, Moon, and Earth. The SPS also keeps the solar panel pointed close enough to the Sun to meet power requirements. To mitigate the shuttle glow observed in flight during experiments conducted in IOC, the SPS also avoids pointing the telescope near the direction of flight. Taken together, these constraints define the time-varying allowable pointing zone.

SPHEREx has a single main target list that includes targets from the all-sky survey as well as targets in both deep fields. All targets are organized into *target groups*. To move to a desired target group,

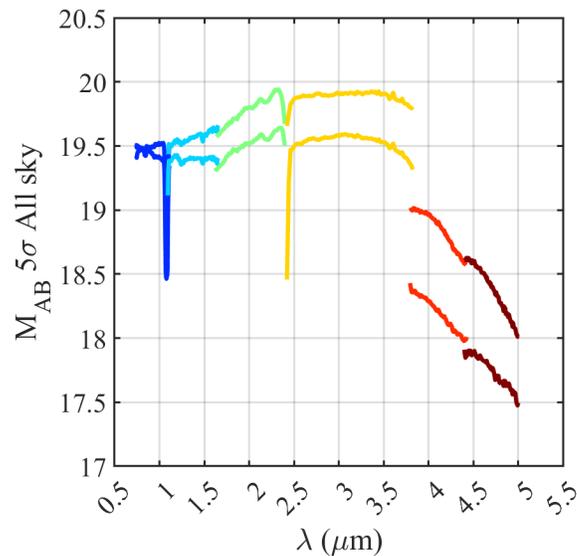


Figure 3: Preliminary expected all-sky point source sensitivity after four surveys assuming a median sky background photocurrent. Each color represents an individual SPHEREx band. For each, two sets of curves are displayed, which encompass the current range of uncertainty. These are dominated by measured quantities, including the absolute spectrophotometric gain, the total noise per exposure, and the effective Point Spread Function. Updates to this curve will be provided as the survey and analysis products progress. The two narrow band dips at 1.08 and 2.42 μm are caused by the increase in photocurrent from HeI emission in Earth’s atmosphere and the attenuation at the R/T transition of the dichroic beamsplitter, respectively.

SPHEREx executes a *large slew* maneuver. After the observatory settles at this attitude, SPHEREx takes a single exposure, then executes a *small slew* maneuver to the next target in the group, and continues taking exposures followed by small slews until the target group is about to move out of our allowable pointing zone. The SPS then plans a large slew to the next target group, and observing continues in this way throughout our science operations. To plan these slews efficiently, the SPS selects the sequence of target groups to minimize the number of time-expensive large slews and maximize the number of observations made within a target group before it moves out of the allowable pointing zone.

A single target group consists of 17 pointings at an identical ecliptic longitude and spaced by one channel width in ecliptic latitude. For any single spectral channel, completing a target group yields a square region of full coverage in that single spectral channel. To achieve full coverage in all spectral channels across the entire sky, we distribute these target groups in a great circle around the equator at all longitudes to form a single latitude ring (Fig. 4). We construct similar latitude rings at great circles at all latitudes, naturally with fewer target groups in the latitude rings at higher latitudes. To mitigate pointing error as well as distortion in the optical system, we overlap the latitude rings, and also overlap the target groups within each latitude ring. This yields an All-Sky target list that achieves full spectral and spatial coverage, efficiently distributes observing time evenly in solid angle across the sky, and contains overlap as mitigation and redundancy.

To efficiently align all observations in the All-Sky Survey, the observatory is always rotated about the telescope boresight such that the wide axis of the field of view aligns with lines of constant ecliptic latitude. This rotation angle is the ecliptic Position Angle (PA) of the observatory, and is measured relative to the line of constant ecliptic latitude passing through the telescope boresight. To meet the solar panel illumination requirements, during ascending halves of the orbit we align the observatory with a PA of 0° , and during descending halves of the orbit the observatory is aligned with a PA of 180° . This also plays a key role in the definition of the surveys. Survey 1 is composed of data taken during the first 12.5 months of science operations with PA = 0° , Survey 2 is composed of data taken during the first 12.5 months of science operations with PA = 180° , Survey 3 is composed of data taken during the second 12.5 months of science operations with PA = 0° , and Survey 4 is composed of data taken during the second 12.5 months of science operations with

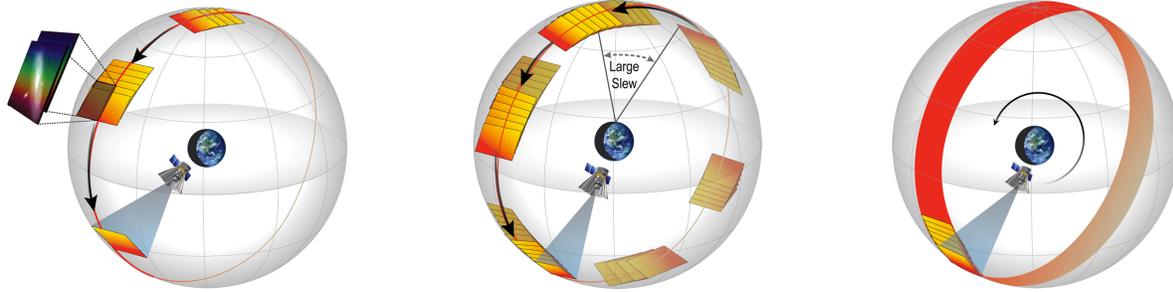


Figure 4: SPHEREx completes full spectral sampling near the terminator, governed by multiple pointing constraints. As the orbit progresses, the observable arc of sky gradually rotates, enabling full-sky spectral coverage twice every year. Throughout the mission, SPHEREx will accumulate approximately 5×10^{12} spatial-spectral elements. (Credit: Image courtesy of SPHEREx-JPL/Caltech)

PA = 180° .

Each target in the All-Sky Survey is spaced by an angle on the sky that corresponds to one spectral channel in the LVF. To enable internal data consistency checks and to ensure full spectral coverage even of narrow spectral features, Surveys 3 and 4 are shifted on the sky relative to Surveys 1 and 2 by an angle on the sky that corresponds to half of a spectral channel. This shift Nyquist-samples the spectra taken across the entire sky, twice, when data is combined across the full mission.

SPHEREx conducts a Deep Survey to collect data to achieve our galaxy formation science goals. One field is located at the north ecliptic pole, and the other is located near the south ecliptic pole. The southern deep field is offset slightly from the pole to avoid foregrounds and bright stars. The Deep Survey target list was developed to meet the power spectral sensitivity required by the extragalactic background light investigation. The strategy was to design a repeatable tiling pattern that, together with the sky rotation observed at the poles across the year from our sun-synchronous orbit, would achieve the required coverage. The pointing targets that make up this pattern at each pole are generated by a separate tool from the core survey software. The pattern is constructed by stepping the boresight along the spectral direction to create target groups similar to those in the All-Sky Survey. Unlike the All-Sky Survey, in the Deep Survey target list, these target groups are highly overlapping to yield deep repeated coverage.

1.5 Mission and Science Operations Summary

SPHEREx is a collaborative mission involving multiple institutions, with mission operations led by Jet Propulsion Laboratory (JPL) in Pasadena, California. The mission is led by Principal Investigator (PI) Dr. Jamie Bock based at Caltech, which was also responsible for the payload development in partnership with JPL. The spacecraft and telescope were designed, constructed, and tested by BAE Space and Mission Systems (BAE), which also led observatory-level integration and testing. The payload, including the telescope, optics, and focal plane assemblies, as well as the passive thermal control subsystem, was designed, integrated, and tested by Caltech and JPL. The Korea Astronomy and Space Science Institute (KASI) contributes as an instrument and science partner, and provides a specialized cryogenic testing chamber for instrument testing. Science data processing is managed by the SPHEREx Science Data Center (SSDC) at Caltech/IPAC, with long-term data archiving handled by the NASA/IPAC Infrared Science Archive (IRSA). The SPHEREx Science Team (SST), composed of researchers from Caltech, JPL, KASI, and numerous partner institutions, is responsible for scientific analysis across SPHEREx's three primary themes (Sect. 1.1).

The Mission Operations Team manages mission operations, including command sequencing and flight system performance monitoring. It also oversees the operation of the SPS, which was provided by Arizona State University. Spacecraft communications are executed via NASA's Near Space Network (NSN) ground stations using S-band (low rate uplink and downlink) and Ka-band (high rate downlink). Science data and select telemetry are routed to the SSDC. There, data are depacketized, calibrated, and processed for



scientific use. The SSDC also implements algorithms developed in collaboration with the PI and SST to produce the data processing pipeline and data products described in Sections 2 and 3.

The Mission Operations Team at JPL is responsible for running the SPS and Planning Generation and Simulation Operations (PGSO) and uplinking the resulting command sequences to the spacecraft. In general, two survey plans are loaded per week, and these plans are designated by the week of the year (e.g., Week 20 in 2025 2025W20), whether a given plan is the first or second half of the week (1 or 2) and a letter designating the version (A, B, etc.), resulting in a final name of, e.g. 2025W20_2B.

1.6 Acknowledgments

If you use data from SPHEREx in a publication, please include the following acknowledgment:

This publication makes use of data products from the Spectro-Photometer for the History of the Universe, Epoch of Reionization and Ices Explorer (SPHEREx), which is a joint project of the Jet Propulsion Laboratory and the California Institute of Technology, and is funded by the National Aeronautics and Space Administration.

2 The SPHEREx Data Products

This document provides a comprehensive overview of the SPHEREx Quick Release (QR) data products. The data processing pipeline and calibrations used for the QR products are initial versions, and both will be updated for later data releases. To access SPHEREx data, see <https://irsa.ipac.caltech.edu/Missions/spherex.html>, which includes links to the available data search, visualization, and download options.

Figure 5 illustrates the data processing flow and the associated data products. Main data products, shown in green, will be publicly distributed via IRSA and are described in detail in their respective sections, as noted by reference numbers shown in parentheses. Ancillary or intermediate products, shown in orange, are generated during internal processing and either contribute to the creation of the main products or are embedded within them (e.g., as layers in Multi-Extension FITS (MEF) files). Data pre-processing and product generation steps, shown in pink, are elaborated in Section 3.2. Arrows in the diagram indicate the directional flow of data, emphasizing the transformation from raw telemetry to science-ready outputs.

2.1 Spectral Images

Spectral image files are produced by the SSDC Level 2 pipeline for each sky pointing and detector. While no zodiacal light subtraction is applied, each file includes an estimated background level (using the model described in Crill et al. [2025]) to support downstream analysis. These products are released to the public as part of the QR process within 60 days of on-orbit acquisition. In addition, all spectral images collected to date will be re-processed and re-released during the annual Data Releases after the completion of Years 1 and 2.

Filename Convention Each spectral image file is named according to a standardized convention that encodes the observation ID (including survey planning period, and large and small slew counters), detector number, and Header and Data Unit (HDU) processing time. Table 3 provides a representative example of the filename structure and its components.

Table 3: Example of filename convention

ID Type	String form
Full image name	level2_2025W22_2B_0001_1D3_spx_l2b-v4-2025-152.fits
Observation ID	2025W22_2B_0001_1
• Survey Planning Period	2025W22_2B
• Large Slew Counter [†]	0001
• Small Slew Counter [‡]	1
Detector #	3
Processing Time	2025-152 (152 nd day of the year 2025)

Notes: [†]: Starts from 0, [‡]: Starts from 1.

Data Format Each spectral image is stored as an MEF file comprising six HDUs: five image extensions and one binary table. The contents of each extension are as follows:

- **IMAGE** Calibrated surface brightness flux in scientific units (MJy/sr), stored as a $2,040 \times 2,040$ pixel array.
- **FLAGS** Bitmap of per-pixel status and processing flags, also $2,040 \times 2,040$ in size. (Sects. 3.2.4, 3.2.8, 3.2.9)
- **VARIANCE** Per-pixel variance estimate in units of $(\text{MJy/sr})^2$, provided in a $2,040 \times 2,040$ array. (Sect. 3.2.2)

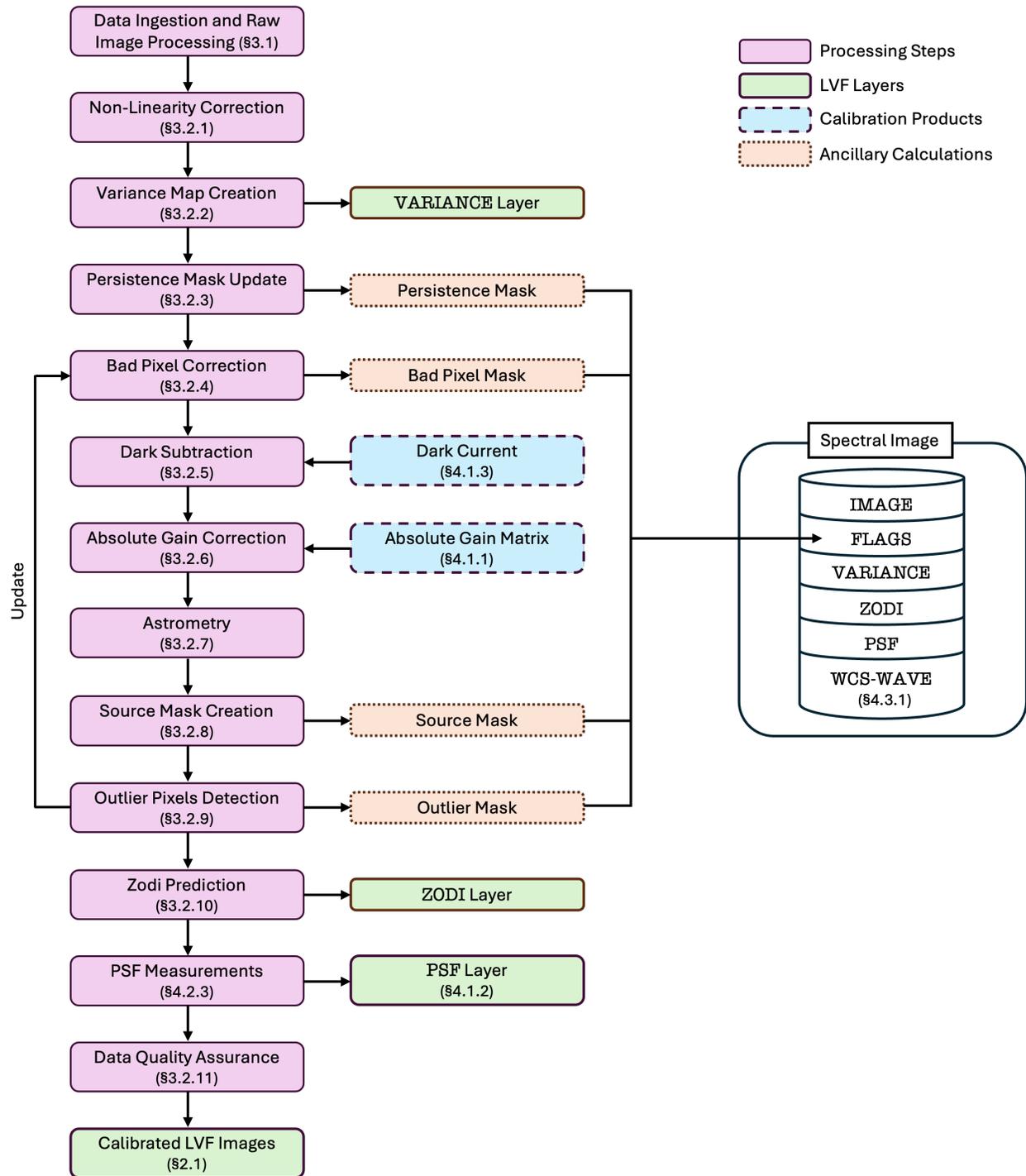


Figure 5: SPHEREx data processing flowchart and associated products. Pink boxes represent individual data processing steps. Green boxes with bold outlines denote data products that include layers of spectral images. Orange boxes with dotted outlines indicate ancillary calculations used in or derived from the main processing pipeline. Blue boxes with dashed outlines correspond to calibration products. Arrows illustrate the flow of inputs and outputs between components. Section numbers in parentheses refer to detailed descriptions in the text. Layer information for a multi-extension spectral image is shown.

- **ZODI** Model-estimated zodiacal light background in units of MJy/sr, given as a $2,040 \times 2,040$ array. (Sect. 3.2.10)
- **PSF** Image cube of over-sampled PSFs, represented as a $101 \times 101 \times 121$ array, corresponding to 121 PSF samples each sized 101×101 pixels. PSF is normalized to its integral and dimensionless. (Sect. 4.2.3).
- **WCS-WAVE** Spectral World Coordinate System (WCS) lookup table mapping pixel coordinates to central wavelengths and bandwidths, formatted as a binary table with 3 columns and 1 row. (Sect. 4.3.1)

Image Header Description The header of a spectral image file encodes key metadata used for identification, calibration tracking, and scientific analysis. Key fields include the observation ID, detector number, product creation time, exposure start time (in UTC), and total exposure duration. Astrometric metadata, such as coordinate system definitions and optical distortion parameters, are also recorded to support accurate spatial registration. Each file header contains unique identifiers for the calibration products applied during processing, along with an approximate mapping between pixel coordinates and their corresponding central wavelengths. Additional metadata relevant to downstream analysis includes spacecraft and instrument parameters – for example, mean detector array and telescope mirror temperatures during the exposure.

The header also includes information on the planned pointing for the telescope boresight (`SPS_LON_ECL`, `SPS_LAT_ECL`, and `SPS_PA_ECL`) in ecliptic coordinates as specified in the survey plan. These values are the same for all 6 bands in each observation. The `SPS_PA_ECL` keyword can be used to determine which survey a given observation belongs to:

- From $t = 0$ to 12.5 months, all-sky data taken with $PA = 0$ is Survey 1
- From $t = 0$ to 12.5 months, all-sky data taken with $PA = 180^\circ$ is Survey 2
- From $t = 12.5$ to 25 months, all-sky data taken with $PA = 0$ is Survey 3
- From $t = 12.5$ to 25 months, all-sky data taken with $PA = 180^\circ$ is Survey 4

The pointing targets for Surveys 3 and 4 are shifted by $6'$ with respect to Surveys 1 and 2 to provide finer spectral sampling.

The RA and DEC calculated per band as part of the astrometric calibration (Sect. 3.2.7) are available in the standard `CRVAL1` and `CRVAL2` keywords. The header also records the full list of data processing modules executed to generate the file, including associated software versioning. A representative header structure is shown in Table A.20, with all entries specific to the illustrated example. While real data headers include `COMMENTS` and `HISTORY` fields that document processing history, these are omitted from the table for brevity.

In spectral image files, the long-keyword convention (`HIERARCH`)², originally developed by the European Southern Observatory (ESO), is used to support header keywords that exceed the standard 8-character Flexible Image Transport System (FITS) limit. This convention ensures compatibility with automated metadata parsing during file ingestion and analysis. The `IMAGE` extension contains a WCS header that provides an approximate pixel-to-wavelength mapping.

Data Product Size Each spectral image file is approximately 70 MB in size.

2.2 Image Anomalies and Features

A small percentage of the SPHEREx on-sky data show a range of image artifacts, atmospheric features, and anomalies. Some of these were predicted before launch, while others are stronger or more frequent. A list of these issues is given below, along with example images. Later versions of the SPHEREx data pipeline will address some of these issues. A list of images where the substantial portion of the image is contaminated is maintained at <https://spherex.caltech.edu/page/anomalyimages>.

²https://fits.gsfc.nasa.gov/registry/hierarch_keyword.html

Anomalies Defined as signals that originate from the behavior of the optics or detectors, or from interactions with cosmic rays or other particles.

- **Frame edge ghost (Fig. 6 left):** Sources which are focused into a narrow ‘frame edge’ region just outside the detector FoV can scatter light into the detector active area. This ghost has a roughly triangular shape and can be produced by any of the four edges of each detector. Due to the precise positioning during sequences of small slews, ghosts on the left or right edge can appear in multiple exposures.
- **Linear bright source ghost (Fig. 6 center):** Very bright sources off the top or bottom edge of the Bands 4–6 detectors can produce linear ghosts as shown in the figure. This ghost does not depend on the precise positioning of the source.
- **Beam splitter ghost (Fig. 6 right):** Very bright sources can show a reflection from the Dichroic Beam Splitter (DBS) consisting of a $\sim 10'$ diameter disk. The ghost is most prominent in the beam splitter transition region near the wavelength boundary of Bands 3 and 4, but the ghost is also present at lower levels elsewhere in Bands 3–6.

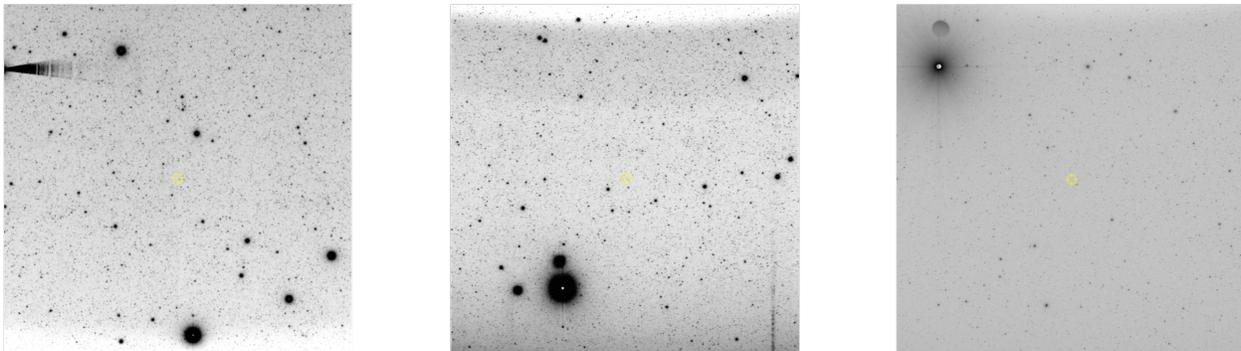


Figure 6: Left: Example of a frame edge ghost (top left of image). Center: Example of a linear bright source ghost (lower right corner). Right: Example of a beam splitter ghost (upper left corner).

- **8° ghost (Fig. 7 left):** A stray light artifact produced by very bright stars approximately 8° from the centers of Bands 2, 3, 5, and 6 has been found in a fraction of exposures in those bands.
- **Moon stray light feature at detector edge (Fig. 7 right):** A fraction of exposures in Bands 1, 4, 5, and 6 show excess emission at the left or right edges, believed to be attributable to stray light when the Moon illuminates the telescope assembly.
- **Extended cosmic ray signatures:** Every SPHEREx image includes detector pixels impacted by cosmic rays. The vast majority of these events directly impact only one pixel, and above a threshold of ~ 100 photoelectrons, the cosmic ray signal is removed by the onboard transient detection. However, as with other missions such as JWST, many of the cosmic rays also affect neighboring pixels, including ones that do not trigger the transient detection. Electrical crosstalk to nearest neighbors is one common path to an extended cosmic ray signature (Fig. 8 bottom left). Less frequently, events with paths near the plane of the detector array, with especially large deposited energy, or clustered in a shower, will affect multiple pixels directly. These events include ‘snowballs’³, which produce a core of pixels flagged with OVERFLOW and TRANSIENT, along with an extended halo of unflagged pixels (Fig. 8 bottom right). For many of these extended cosmic ray signatures, it may be necessary to discard pixels beyond those that are flagged in science analyses.
- **Cosmic-ray-induced glow:** Cosmic rays also induce a faint, diffuse glow in the SPHEREx detectors, which is more visible when the Transient fraction is large. The glow is strongest in Bands 1 and 2 and produces a spectral gradient with more emission on the long-wavelength side. The glow response in Bands 4–6 is weak overall, with a brighter rim along the four edges of the detector array.

³<https://jwst-docs.stsci.edu/known-issues-with-jwst-data/shower-and-snowball-artifacts#gsc.tab=0>
Caltech/IPAC

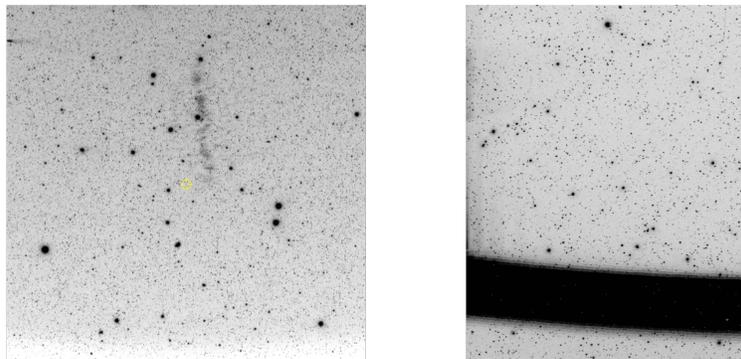


Figure 7: Left: In the upper middle of exposure, a complex and diffuse stray light artifact produced in Band 4 by a very bright star approximately 8° away. Right: Excess emission along the lower left edge of a Band 1 exposure, caused by Moon stray light. (The brighter, horizontal band is the $1.083 \mu\text{m}$ atmospheric He line.)

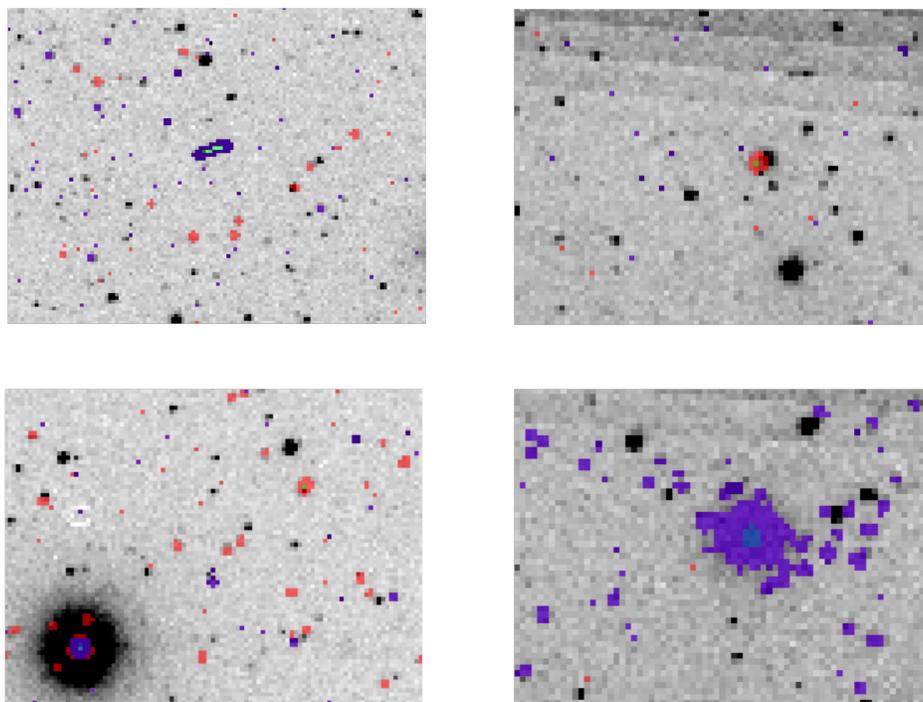


Figure 8: Examples of anomalous cosmic ray flagging. The gray scale is the spectral image flux, and the displayed flags are TRANSIENT (red), OVERFLOW (green), and SUR_ERROR (blue). Note that red and blue flags on the same pixel will appear as purple. Larger cosmic rays will have OVERFLOW flags in the center of the transient group. Snowballs have very large centers and are often accompanied by a number of nearby smaller transients. In all cases, pixels surrounding the transients show excess signal, possibly due to crosstalk or blooming. Top Left: Very early cosmic ray (center exhibits a transient prior to transient flagging and is flagged for OVERFLOW only), Top right: Late cosmic ray, Bottom left: Late cosmic rays with blooming source in bottom left of image, Bottom right: Early snowball.

- **Crosstalk (Fig. 9):** In addition to the aforementioned crosstalk among neighboring detectors, occurring at the percent level, crosstalk also produces weaker artifacts at larger distances. These have been observed to take three forms in SPHEREx: 1) a negative image of a bright source displaced vertically by the readout row skipping spacing of ~ 32 rows; 2) a smear resulting from a slowly decaying response

to a bright source in the last read pixel of the prior row; 3) crosstalk among detector readout channels spaced by a multiple of 64 pixels horizontally.

Features Defined as signatures from atmospheric or non-astrophysical objects.

- **Atmospheric He Emission line (Fig. 10 upper row):** Bands 1 and 2 include emission from the atmospheric He line at $1.083 \mu\text{m}$ [Brammer et al., 2014], varying from a factor few to a factor of 30 greater surface brightness than the Zodiacal, or Zodi Light (ZL) background. The line amplitude varies over the orbit and with time, but astrophysical sources are visible behind the emission, and the variance layer of the LVF includes the impact of this emission line. The line is strong enough to produce weak but detectable spectral leak features in Band 1 (in pixels with bandpass center near $0.84 \mu\text{m}$) and Band 2 (near 1.49 and $1.25 \mu\text{m}$).
- **Auroral emission lines (Fig. 10 bottom row):** Near-infrared emission lines are a prominent feature of the aurora, appearing particularly in Bands 1 and 2, with occasional much fainter lines in Bands 3, 4, and 5. These lines are generated by excited atoms and molecules in the upper atmosphere, primarily nitrogen (N_2). The lines vary with solar activity, are more frequently seen when SPHEREx is above the polar regions, and are generally much fainter than the $1.08 \mu\text{m}$ He line. These spectral features can be a factor of a few brighter than the ZL background.
- **Other atmospheric lines:** Other emission lines from Earth's atmosphere with variable amplitude can be present at 0.78 , 0.84 , 0.93 , 1.13 , 1.81 , 2.63 , 2.75 , 4.03 , and $4.64 \mu\text{m}$.
- **Shuttle glow (Fig. 11):** Shuttle glow is an emission that arises when spacecraft in low Earth orbit interact with the residual atmosphere, particularly atomic oxygen [Viereck et al., 1991]. As the spacecraft travels through this environment, high-speed atomic oxygen impacts surfaces in the ram direction, exciting surface materials and producing a faint glow, including in the near-infrared. To mitigate this effect, the SPHEREx survey plan limits the allowable ram angle during observations. In addition, a one-dimensional shuttle glow spectral template was constructed during commissioning calibration.
- **Diffuse Optical Transients (Fig. 12):** Roughly 1 % of exposures show diffuse and ephemeral features believed to be foreground objects or particles transiting quickly through or near the FoV.
- **Earth orbiting satellites (Fig. 13):** Streaks due to Earth-orbiting satellites can be seen in many images. The brightest examples can be several pixels wide and are flagged on board as transients. Fainter satellite streaks are flagged by the outlier module or a combination of transient and outlier flags. Satellite streaks near the noise level may not be flagged.

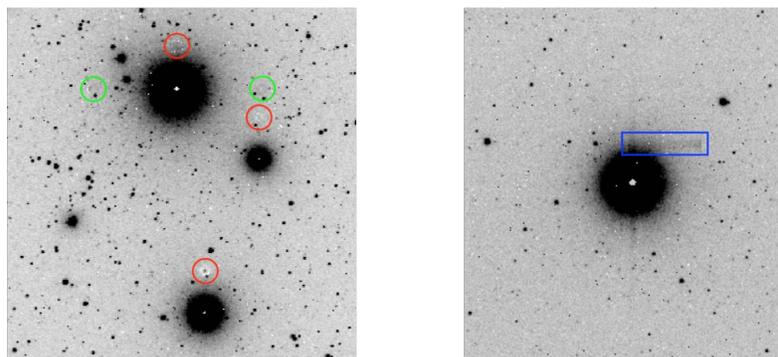


Figure 9: Left: Examples of (in red) negative crosstalk displaced by ~ 32 pixels vertically and (in green) readout channel crosstalk displaced by a multiple of 64 pixels horizontally from bright sources. Right: Example of (in blue) crosstalk smear displaced by 64 pixels horizontally and ~ 32 pixels vertically from a bright source.

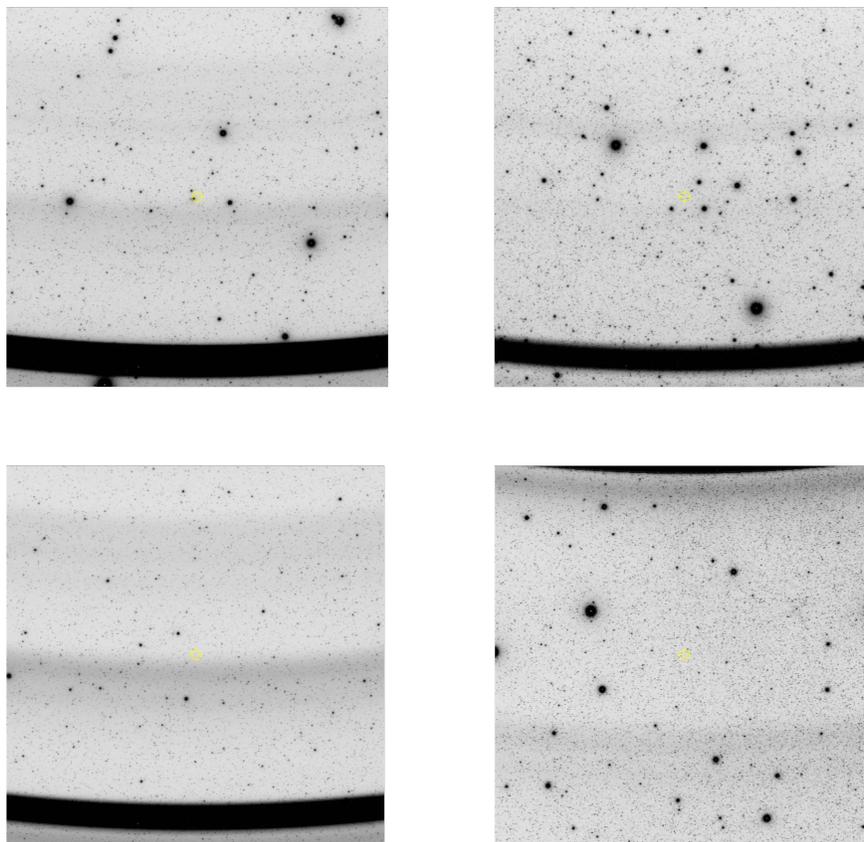


Figure 10: Features from atmospheric or non-astrophysical objects. Top Left: Strong He lines in Band 1, Top right: Weak He lines in Band 1, Bottom left: He and auroral lines in Band 2.

2.3 Known Limitations

Incorrect source masking As described in Section 3.2.8, the spectral images include flags for pixels which have significant flux contributions from known sources. The size of the source mask is set by the magnitude from the Reference Catalog. If the estimated magnitude for the relevant pixel is incorrect, the mask will be too large or too small (Fig. 14).

Pixels at the dichroic edge (Fig. 14) The long-wavelength side of Band 3 and the short-wavelength side of Band 4 each have rows of pixels where the sensitivity and spectral responses are impacted by the DBS. Pixels where the throughput is below 50 % have the **DICHOIC** flag set (see Table 9). Users should exercise caution when working with sources in these areas of the detectors, as the noise will be increased and the photometric calibration will be more uncertain.

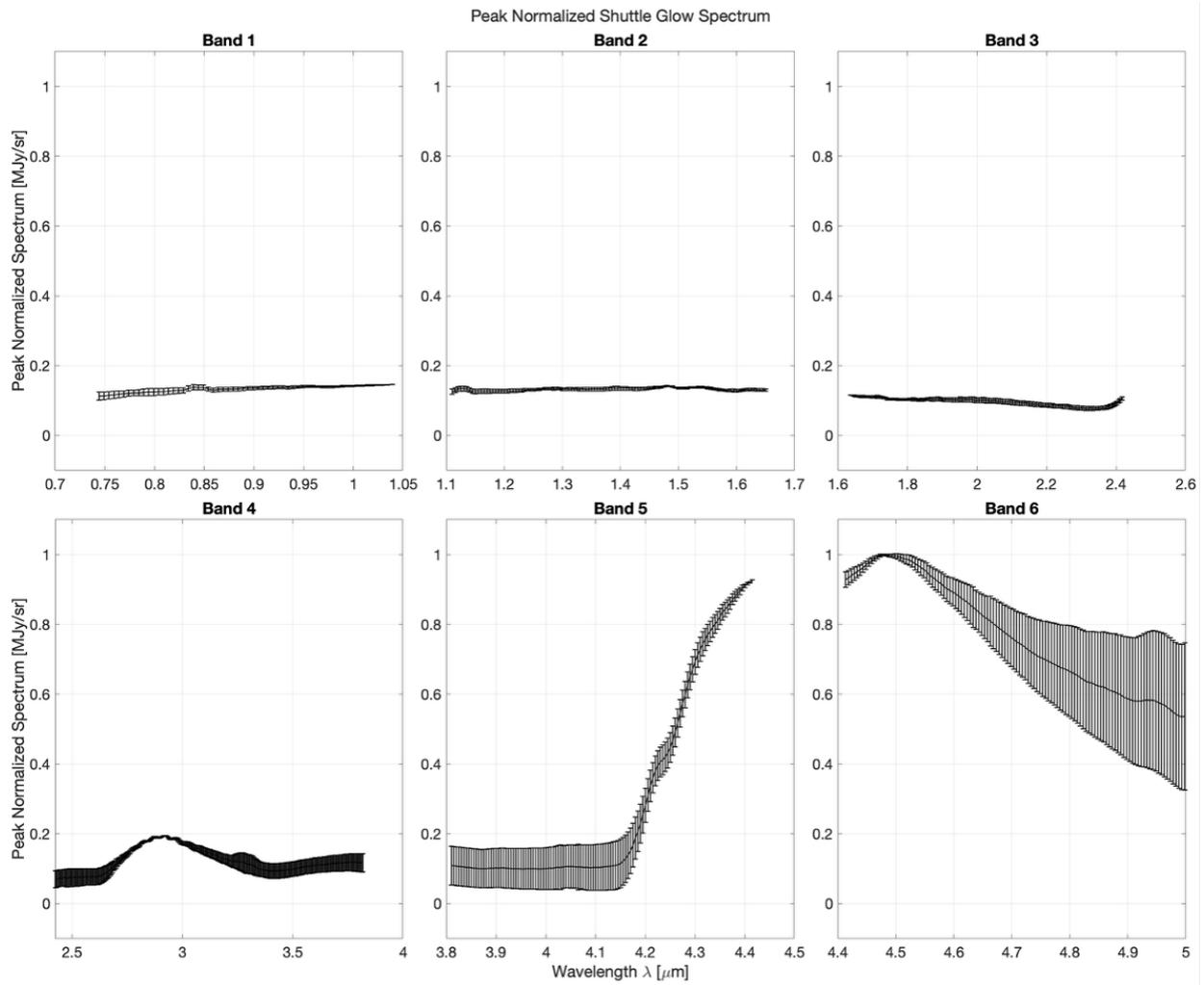


Figure 11: Shuttle glow emission spectrum. The template is flux-calibrated in MJy/sr across detectors and peak-normalized, showing that most of the contamination originates at $\sim 2.9 \mu\text{m}$ in Band 4 and at $\sim 4.5 \mu\text{m}$ in Bands 5 and 6.

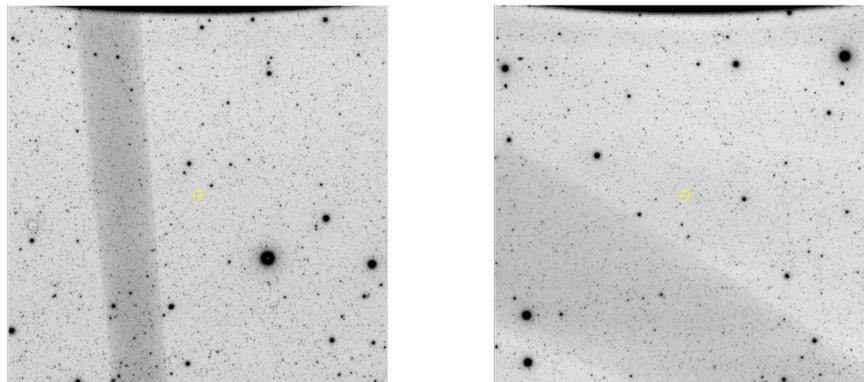


Figure 12: Examples of diffuse optical transients.

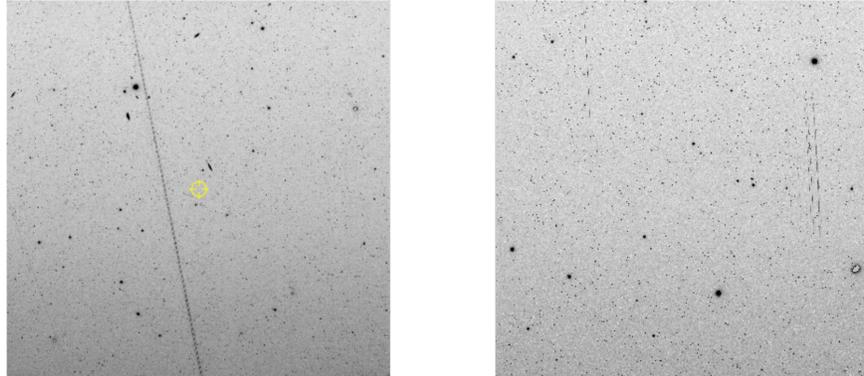


Figure 13: Examples of spectral images with satellite streaks.

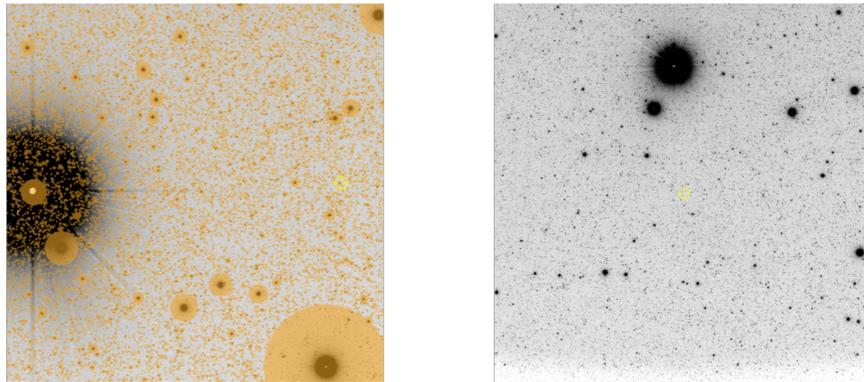


Figure 14: Left: example of a source mask which is too small, the source mask flag is shown in orange. Right: a Band 3 image showing the dichroic-impacted pixels at the bottom of the frame.

3 The SPHEREx Data Processing

Data Processing Operation The SPHEREx data pipeline follows a tiered structure consisting of five distinct processing levels:

- **Level 0:** Raw telemetry processing, including on-board data handling, ingestion, decompression, and basic quality assessment.
- **Level 1:** Conversion of raw measurements into engineering units, such as photocurrent in electrons per second.
- **Level 2:** Photometrically and astrometrically calibrated spectral images and data cubes; calibrated spectral images form the QR data products.
- **Level 3:** Forced photometry performed at predefined reference positions, yielding catalogs
- **Level 4:** High-level science data products, including derived parameters such as galaxy redshifts, ice column densities, and deep image mosaics.

The pipeline described in this document includes Levels 1 to 3, although not all products are produced as part of the QR data production. The pipeline also derives the relevant calibrations. Section 4 describes the calibration products, the algorithms used to derive them, and the current performance. For SPHEREx, the traditional flat field and photometric gain products are combined into the Absolute Gain Matrix (Sect. 4.2.1) due to the wavelength variations across the band.

The SSDC is responsible for the development, implementation, and operation of processing pipelines from Level 1 through Level 3, in collaboration with the SST. Level 4 science products and advanced analysis algorithms are developed by the SST, tailored to the mission's three scientific pillars. All public data products will be distributed through IRSA, which hosts an interactive suite of tools for data searching, visualization, and retrieval.

The SPHEREx spectral images and associated calibration products released as part of the QR data set (Sect. 2) are processed by the SSDC. This section outlines the data processing procedures and algorithms, where applicable, that are applied to produce these QR products. Each subsection corresponds to the numbered pink boxes in Figure 5, which trace the data generation workflow.

3.1 Data Ingestion and Raw Image Processing

The tasks described in this section correspond to Level 0 and Level 1 of the SPHEREx data processing pipeline.

Onboard the observatory, the instrument's warm electronics compute pixel-level slopes from the detector readouts, compress and packetize the data, and transfer them to the spacecraft for temporary storage. The spacecraft then telemeters these science packets to the ground via Ka-band multiple times per day (typically ~ 5). Upon receipt, the raw data are downlinked through the NASA NSN. At Caltech/IPAC, the SSDC ingests science data packets from Data Acquisition Processing & Handling Network Environment (DAPHNE) via Ka-band and engineering telemetry and navigation data from the Ground Data System (GDS) via S-band. These data streams are combined to generate Level 0 raw images. Each ingestion event includes the generation of a File Receipt Report, which is sent to GDS, and a push notification from DAPHNE confirming the availability of new data. At this stage, raw images and associated metadata are staged for processing by downstream pipelines.

After each downlink, the data packets are decompressed, and any missing pixel blocks are flagged. Coarse astrometry is determined based on the survey plan and geometric offsets. This initial pointing solution is accurate to $2' - 10'$ (and a rotation of a few degrees) and is to be refined later by the fine astrometry module (Section 3.2.7). To validate the SUR readout mode, an 8×34 pixel subsection of the detector is extracted and compared against the full SUR sequence to monitor noise performance and temporal stability. The raw pixel data are then converted from Analog Digital Units (ADU) to engineering units (electrons per second). For a small number of pixels, the Level 1 processing sets the pixel value to NaN. This is done if the on-board error bit and either the early overflow or SUR_ERROR flags are set (see Table 8). Corrections are applied for amplifier chain drifts using reference pixels and interleaved phantom pixels. These corrections are flagged and recorded alongside the image data. The resulting calibrated files are written in a standardized

format and archived. Before advancing to Level 2 processing, each Level 1 image file undergoes a series of automated quality checks. A subset of these validation results is communicated back to the Mission Operations System (MOS) for operational monitoring.

3.2 Processing Steps

Following Level 0 and Level 1 processing, Level 2 data processing is performed. At this stage, Level 1 images undergo astrometric and photometric calibration using a combination of ancillary catalogs and internally generated calibration products from the pipeline.

3.2.1 Nonlinearity Correction

This processing unit corrects for detector Non-Linearity (NL), which arises from declining gain as charge accumulates within each H2RG pixel (e.g., Zengilowski et al. 2020). As a prerequisite to correction, observed (uncorrected) photocurrent measurements are converted into (uncorrected) electron counts by incorporating the pixel-specific integration time. This transformation enables precise application of the NL correction. The corresponding inputs and outputs of this processing step are summarized in Table 4.

Table 4: Inputs and Outputs of the NL Correction Module

Input	Output
Uncorrected photocurrent images, NL correction parameters, Charge thresholds for algorithm applicability, SUR parameter settings	NL-corrected photocurrent images, Binary flag

Input and Output Description The NL correction module is applied exclusively to the $2,040 \times 2,040$ pixel photo-sensitive region of each detector, excluding both phantom pixels and reference rows (see Sect. 3.1). The input NL calibration parameters are derived from ground-based testing in which the detector array is illuminated with diffuse, constant intensity and the signal is recorded for each frame of an exposure.

The output files retain the same FITS image format, units, and size as the input, except that the detector signals have been linearized. For instance, signal levels near the upper end of the dynamic range are boosted to approximate the response of an ideal linear detector.

Description of Algorithm The SUR parameters are set on board such that slope fits use only samples below the Overflow threshold, which is set at approximately half of the full-well capacity of the detectors. In this range, the nonlinearity is described to sufficient accuracy using a quadratic:

$$Q_{\text{observed}} = Q_{\text{collected}}(1 - Q_{\text{collected}}/Q_{\text{nl}}), \quad (1)$$

where $Q_{\text{collected}}$ is the true charge deposited in the pixel, Q_{observed} is the observed signal in the pixel (in the same units and using the calibration applicable at low charge), and Q_{nl} is the nonlinearity parameter. The on-board electronics measures slope fits to Q_{observed} vs. sample index during the exposure, and it is the task of the nonlinearity module to estimate $Q_{\text{collected}}$ and therefore the correction factor $(1 - Q_{\text{collected}}/Q_{\text{nl}})^{-1}$. Typical values for Q_{nl} are $1.4\text{E}6 \text{ e}^-$ for Bands 1–3 and $0.8\text{E}6 \text{ e}^-$ for Bands 4–6. The correction factor is 1.05–1.08 for full-length exposures that reach the Overflow threshold at the end.

Assuming the model in Equation 1, combined with some SUR parameters and some algebra, the corrected photocurrent for each pixel is found to be:

$$F = \frac{Q_{\text{nl}} \left[1 - \sqrt{1 - 4T_{\text{fr}}(N + \text{SURDLY})F'/Q_{\text{nl}}} \right]}{2T_{\text{fr}}(N + \text{SURDLY})}, \quad (2)$$

where F' is the observed photocurrent, $T_{\text{fr}} \approx 1.53$ seconds is the frame duration, $\text{SURDLY} = 3$ is the index of the first frame used in the slope fit, and N is the index of the last frame used in the slope fit. For

uninterrupted SURs, $N = \text{SURLIM} = 77$. For an Overflow-interrupted SUR, N is derived through iterated solution of Equation 2 and:

$$N = \text{minimum} \left(\text{SURLIM}, \text{floor} \left[\frac{Q_{nl}(1 - \sqrt{1 - 4Q_{over}/Q_{nl}})}{2T_{fr}F} \right] \right), \quad (3)$$

where Q_{over} is the Overflow threshold in charge units. If the `SUR_ERROR` flag is set (see Sect. 3.2.4 and Table 9), then no correction is applied, i.e., $F = F'$. If the `TRANSIENT` flag is set (see Sect. 3.2.4 and Table 9), then an estimate is made for the value of N since it is not known exactly.

For pixels with low confidence in the nonlinearity correction due to poor intrinsic linearity or inadequate calibration data, the `NONLINEAR` flag is set by this module (see Sect. 3.2.4 and Table 9).

3.2.2 Variance Map

This processing unit generates per-pixel variance maps for each detector array, producing an estimate of the variance based on the downlinked ramp or flux images processed on board. The method is based on the derivations present in Robberto [2007], which describe the statistical foundations for variance estimation under various non-destructive readout schemes used in infrared detectors. Variance estimates are derived from the raw image data and updated at each step in SSDC Level 2 pipelines where the flux image is processed, ensuring the propagated variance remains accurate. The corresponding inputs and outputs of this processing step are summarized in Table 5.

Table 5: Inputs and Outputs of the Variance Map Module

Input	Output
Level 1 detector images, Per-pixel electronic noise map	Variance map, Binary flag

Input and Output Description The inputs to this module include the Level 1 FITS image ($2,072 \times 2,176$), which contains both phantom pixels and repeated samplings of the reference rows, expressed in engineering units (in e^-/s). Additional inputs include a per-pixel electronic noise map for each array, with dimensions of $2,048 \times 2,048$ pixels (in e^-), and an accompanying flag map. Reference pixels are retained in the variance calculation to support trending and analysis. The electronic noise model accounts for both Correlated Double Sampling (CDS) noise and excess readout noise, characterized through pre-flight laboratory measurements. However, shot noise due to dark current is excluded from this estimate, as it is incorporated separately from the image itself – dark current is not subtracted at this processing stage.

The output consists of a variance map with the same dimensions ($2,072 \times 2,176$, in $(e^-/s)^2$), representing the maximum likelihood estimate of combined Poisson and electronic noise per pixel. Variance values for phantom pixels may be set to `NaN`, while those for reference pixels (including repeated samplings) are directly sourced from the $2,048 \times 2,048$ electronic noise map. A corresponding flag is also generated to indicate uncertainty in the variance estimate, triggered in cases of overflow or when transient events disrupt integration.

Note: While the variance module initially produces a variance layer with dimensions of $2,072 \times 2,176$ in units of $(e^-/s)^2$, subsequent calibration to physical units reduces the array size to $2,040 \times 2,040$ and converts the units to $(\text{MJy}/\text{sr})^2$. This final calibrated version is the only variance product exposed to users and is delivered as the `VARIANCE` layer of the spectral image (Sect. 2.1).

Description of Algorithm The underlying mathematics of this module follows the framework detailed in Robberto [2007]. For SPHEREx, the detector readout scheme involves uniformly sampling the accumulating signal, consistent with the method described in Section 2.2 of that reference. Accounting for both readout and Poisson noise contributions, the maximum likelihood estimator for the total signal variance is derived

by combining Equations 1.39 (readout noise) and 1.49 (Poisson noise) in Robberto [2007]:

$$\text{Var} = \left[\sigma_{\text{CDS}}^2 \frac{6(N-1)}{N(N+1)} + \frac{6}{5} \frac{N^2+1}{N^2-1} F \cdot T_{\text{int}}(N) \right], \quad (4)$$

where σ_{CDS} is the CDS readout noise for a given pixel, N is the number of reads, F is the measured slope of the accumulating signal, and $T_{\text{int}}(N)$ is the integration time for N reads. This expression estimates the variance of the total integrated signal. To compute the variance of the measured slope, the result is divided by $T_{\text{int}}(N)^2$.

However, laboratory testing revealed an additional excess readout noise component that effectively increases the CDS noise. Since pre-flight testing directly measures per-pixel electronic noise, these values are used in place of modeled noise estimates. Incorporating this lab-based variance, Equation 4 becomes:

$$\text{Var} = \left[\sigma_{\text{det}}^2(ij, N) + \frac{6}{5} \frac{N^2+1}{N^2-1} F \cdot T_{\text{int}}(N) \right], \quad (5)$$

where $\sigma_{\text{det}}^2(ij, N)$ represents measured per-pixel electronic noise variance for detector pixel (i, j) and for a total of N reads.

There are two specific cases in which the commanded number of reads can differ from the actual number used in slope estimation:

- **Overflow:** For bright sources that saturate a pixel mid-ramp, the effective number of reads N is inferred from the measured flux and the overflow threshold at which sampling ceased. In such cases, the estimated N is substituted into the Poisson term on the right-hand side of Equation 5 to evaluate variance. Given the dominance of photon noise in this regime, the electronic noise term can be neglected.
- **Transient interruption (e.g., cosmic ray hit):** If the integration is disrupted by a transient event, the actual number of usable reads is unknown. In this scenario, the module defaults to using the per-pixel electronic variance for the minimum possible number of reads to yield a conservative estimate of total variance.

If a pixel is flagged as both **OVERFLOW** and **TRANSIENT**, the variance is estimated using the overflow prescription above (photon noise will dominate electronic noise).

3.2.3 Persistence Masks

This processing unit applies a parameterized analytic model to estimate the level of persistent flux in each pixel based on the cumulative exposure history of the survey. Image persistence is a well-documented artifact in H2RG detectors, characterized by a slowly decaying residual signal from previous exposures – typically arising where bright sources previously illuminated the array. This residual flux can contaminate subsequent science images. According to the prevailing model described by Smith et al. [2008], image persistence originates from charge traps within the detector diode depletion region. These traps accumulate charge during standard photodiode operation and release it gradually following a detector reset. Based on the persistence estimates, downstream modules generate masks to flag and exclude pixels likely to be affected, thereby mitigating contamination in the science data. The corresponding inputs and outputs of this processing step are summarized in Table 6.

Input and Output Description The input to this module consists of $2,040 \times 2,040$ slope-fit images from SUR (in e^-/s) for each of the six detectors, along with associated Level 1 binary flag maps that include SUR transient, overflow, and pixel error indicators (corresponding to the **TRANSIENT**, **OVERFLOW** and **SUR_ERROR** flags, see Sect. 3.2.4 and Table 9). Additional inputs include the elapsed time between the end of the previous exposure and the start of the current one, as well as exposure time and the coefficients of the persistence current model from prior exposures. This information is then used to derive persistence-state model coefficients for the current observation.

The output comprises a series of FITS files containing $2,040 \times 2,040$ arrays with the predicted persistence image for each detector and the resulting mask image for the current exposure. Pixels in the persistence image exceeding the predefined threshold are flagged as contaminated and recorded in the mask to be excluded from downstream analysis.

Table 6: Inputs and Outputs of the Persistence Mask Module

Input	Output
Slope-fit images from SUR, SUR flags from level 1 data, Elapsed time between consecutive exposures, Elapsed time of each exposure (TELAPSE), Persistence model parameters from previous exposure (s)	Binary array of Persistence flags

Description of Algorithm The persistence current $I_p(t)$, in units of e^-/s , is modeled as a sum of exponentially decaying terms:

$$I_p(t) = \sum_{n=1}^N A_n C_n e^{-t/\tau_n}, \quad (6)$$

where A_n is an exposure-dependent coefficient based on the value of the coefficient from previous exposures, the elapsed time t since the end of the most recent exposure and the prior exposure's fluence F , given by

$$F = f \times t_{int}, \quad (7)$$

where f is the exposure's slope-fit and t_{int} is the exposure's elapsed time from start to stop, TELAPSE. The coefficients C_n are unitless scaling factors, and τ_n is the decay constant for term n in seconds. The coefficients C_n are selected such that the summation, when $A_n = 1$, approximates the form $1/(t+1)$, with $t = 0$ defined at the detector reset following an illuminating exposure. Each array is assigned fixed values for C_n and τ_n , typically spanning 6 to 8 terms.

For exposure m , the coefficient A_{n_m} accounts for persistence from the previous exposures by exponentially decaying their contribution over the time elapsed since that exposure to the end of the most recent exposure. This is recursively encompassed in the second term of Equation 8:

$$A_{n_m} = B_n(f_{m-1}, t_{int}) + B'_n(f_{m-2}, t_{int}) e^{-(t_{slew_{2 \rightarrow 1}} + t_{int})/\tau_n}, \quad (8)$$

where B_n and B'_n are retrieved from a lookup table based on the slope fit from exposures $m-1$ and $m-2$, respectively and $t_{slew_{2 \rightarrow 1}}$ is the time between the end of exposure $m-2$ and the beginning of $m-1$ in seconds. This formulation assumes persistence is additive. If the the elapsed times differ between exposures, the notation generalizes as $t_{int} \rightarrow t_{int_{m-1}}$ and $t_{int_{m-2}}$ in the first and second coefficients respectively, and the average elapsed time replaces t_{int} in the exponential term.

In the absence of previous exposures, both B_n and B'_n are zero, yielding $A_{n_m} = 0$. When only a single prior exposure exists, $B'_n(f_{m-2}) = 0$. For subsequent exposures ($L = m+1$), the contribution from exposure m is preserved as:

$$B'_n(f_{L-2}, t_{int}) = A_{n_m}. \quad (9)$$

The persistence current predicted for exposure m is evaluated at the temporal midpoint of exposure m :

$$I_p(t_m) = \sum_{n=1}^N A_{n_m} C_n e^{-t_m/\tau_n}, \quad (10)$$

where A_{n_m} is from Equation 8, and

$$t_m = t_{slew_{1 \rightarrow 0}} + \frac{1}{2} t_{int}, \quad (11)$$

with $t_{slew_{1 \rightarrow 0}}$ denoting the time between the end of exposure $m-1$ and the start of exposure m .

The look-up tables producing B_n and B'_n show three regions: zero, linear, and saturation. These result in a piecewise function generically described in Equation 12, where F represents fluence, r represents a band-dependent rate, and the persistence current saturates at band-dependent fluence $F = F_{sat}$.

$$B_n = \begin{cases} 0 & F < 0 \\ r \times F & 0 \leq F \leq F_{sat} \\ r \times F_{sat} & F > F_{sat} \end{cases} \quad (12)$$

There are several instances represented by the flag layer of each image in which the fluence is improperly determined by the multiplication of the flux and the elapsed time from start to stop of the exposure. The flag layer is used to identify these cases and properly assign persistence current according to our interpretation of the specific flag combinations. If the flag combinations include overflow, as illustrated in Table 7, the persistence current is presumed to saturate, and the fluence is overwritten with F_{sat} for the purposes of predicting persistence current in subsequent images.

Table 7: Flag Conditions Affecting Predicted Persistence

OVERFLOW	SUR_ERROR	TRANSIENT	Interpretation	Overwrite F with F_{sat}
1	1	0	Extra High Flux	YES
1	0 or 1	1	Transient Caused Saturation	YES
0	1	1	Early Transient	NO
0	0	1	Late Transient	NO

There is an electronic condition in the detector known as blooming (related to the brighter-fatter effect) where the signal generated in a saturated pixel appears to spill over into a neighboring pixel, causing a change in slope. This causes transient flags to be raised in pixels surrounding bright sources. Pixels subject to this condition, denoted as early transients in Table 7, display saturating persistence current, though they do not exhibit the overflow flag. To avoid assigning excess persistence current to non-saturating cosmic rays with the same flag combination, the fluence of these pixels is not currently overwritten. Thus, their persistence current will be under-predicted, resulting in a ring of pixels around every saturating source that may be under-flagged for persistence in the subsequent two to nine exposures.

This module generates several diagnostic quantities to evaluate temporal trends in residual persistence. These include the total number of pixels flagged for persistence in each exposure frame; the cross-correlation between the current image and the previous exposure, computed both with and without application of the persistence mask; and the auto-correlation of the current exposure to assess residual artifacts originating from bright sources observed prior to the start of the integration.

3.2.4 Bad Pixel Flagging

This processing unit generates a pixel-level map for each detector array, identifying pixels that should be excluded from downstream photometric measurements. For each exposure, the module produces a lookup table flagging the small subset of instrument-frame pixels deemed unsuitable for use beyond Level 1 processing. Rather than employing a simple binary mask, the system encodes the reason for exclusion in a coded map. While this approach entails a modest increase in data volume, the added diagnostic detail enhances the utility of the product for future performance analysis and troubleshooting throughout the mission. The corresponding inputs and outputs of this processing step are summarized in Table 8.

Table 8: Inputs and Outputs of the Bad Pixel Correction Module

Input	Output
Level 1 detector images, Flag maps generated by onboard SUR processing, Persistent images from prior exposures, Map of permanently inoperative pixels	Bad pixel mask images

Input and Output Description The input to this module includes Level 1 photocurrent FITS images, each with dimensions of $2,040 \times 2,040$ pixels in units of e^-/s , and corresponding flag maps in FITS format generated by the onboard SUR system for each of the six detectors. These flag maps are telemetered along with the science data and share the same dimensionality as the science images. Additional inputs include a $2,040 \times 2,040$ FITS map of permanently non-functional pixels for each of the six flight arrays, based on Caltech/IPAC

pre-flight laboratory characterization, as well as flags set by the Correct Nonlinearity (Sect. 3.2.1), Update Persistence Mask (Sect. 3.2.3), Create Source Mask (Sect. 3.2.8) and Outlier Pixel Detection (Sect. 3.2.9) Level 2 modules.

The output comprises a per-exposure, per-detector bad pixel mask, delivered as a $2,040 \times 2,040$ FITS image in which each pixel is annotated with a symbolic flag indicating its classification. Each flag corresponds to a specific bit position, with standardized symbolic identifiers assigned to defined conditions. The full set of flag names is recorded in the header of the associated spectral image output file (see, e.g, Table A.20 in Sect. 2.1).

Image Pixel Flag Definitions Pixel flags are generated across multiple modules within the image-processing pipeline. For each spectral image released as part of the QR dataset, these flags are preserved and publicly distributed through IRSA. IRSA’s visualization interface supports the display of these flags as overlay layers, enabling users to assess pixel-level data quality interactively. Table 9 summarizes the flag names, their corresponding bit values, and associated descriptions.

Table 9: Image pixel flags used in spectral images at Level 2

Symbolic name	Bit #	Description	Note
TRANSIENT	0	Transient (such as cosmic ray hit) detected during SUR. <i>Note:</i> Transients occurring during the first ten seconds of the observation may not be flagged. This flag can also indicate pixels affected by charge spillover (bloom) from very bright neighboring sources. When saturation from a bright star causes charge to spill into adjacent pixels, the onset of such spillover can trigger TRANSIENT flagging by the on-board SUR algorithm.	Determined onboard (Level 0)
OVERFLOW	1	Overflow threshold reached during SUR. This threshold is set at approximately half the level of full-well saturation.	
SUR_ERROR	2	chksum_error bit from instrument data processing. In the absence of other flags, it indicates inconsistencies in the statistics maintained by the SUR algorithm onboard. When combined with the TRANSIENT or OVERFLOW flags, signals early transient (first half of exposure) or early overflow (first ~5 seconds) conditions.	
NONFUNC	6	Static, initially pre-flight assessment that a pixel is permanently unusable/dead. May be updated from time to time during the mission. This is unusable for any purpose.	Static
DICHROIC	7	Pixels in the “dark corners” of Bands 3 and 4, where the “smile” effect results in pixels with a bandpass incompatible with which side of the dichroic they are on. This flag indicates pixels with low expected throughput that are not usefully calibrateable or of value for photometry.	

continued ...

... continued

Symbolic name	Bit #	Description	Note
MISSING_DATA	9	Pixels for which the onboard data was lost before ingest by the SSDC, e.g., as corrupted packages on the downlink that were not retransmitted successfully, resulting that there is no data available for this pixel in this image.	Set by the Level 1 pipeline based on per-image data content and quality assessment.
HOT	10	Hot pixel – noisy or otherwise affected by spurious signal. Determined per-exposure; permanently hot pixels are flagged as NONFUNC. Not usable for science.	
COLD	11	Pixel that appear unresponsive to light or otherwise has an anomalously low signal in a particular exposure; permanently dead pixels are flagged as NONFUNC. Not usable for science.	
FULLSAMPLE	12	Pixel for which the full sample history is available, because of the rotating 8×32 pixel region for full readout. Not usable for science, purely informational.	
PHANMISS	14	Applied for pixels that can not be corrected due to missing phantom pixels.	
NONLINEAR	15	Pixel for which a reliable nonlinearity correction could not be determined. Due to poor photometric accuracy, this pixel should be excluded from most scientific analyses.	Set by the Level 2 pipeline based on per-image data content and quality assessment.
PERSIST	17	Pixel found to have been affected by persistent charge above the Level 2 threshold for this.	
OUTLIER	19	Pixel flagged by “Detect Outlier Pixels” module (Sect. 3.2.9).	
SOURCE	21	Pixel mapped to a known source.	

3.2.5 Dark Subtraction

Dark current (with units of e^-/s) is estimated simultaneously with the flat-field response derivation, as described in 4.2.1. This additive component is first subtracted from the $2,040 \times 2,040$ active pixel region of the Level 1 image prior to applying the absolute gain matrix that converts to physical units.

3.2.6 Absolute Gain Correction

This processing step multiplies the dark current-subtracted $2,040 \times 2,040$ Level 1 image by the absolute gain matrix, described in more detail in 4.1.1. The absolute gain matrix is a $2,040 \times 2,040$ array with units $(MJy/sr) / (e^-/s)$. It encodes pixel-to-pixel response variations within spectral channels as well as the absolute gain conversion for each spectral channel, such that multiplying it by the dark-subtracted Level 1 image brings the image to the physical unit of MJy/sr.

3.2.7 Astrometry

This processing unit refines the astrometric solution by determining the optimal translation and rotation required to align detected stars in each of the six exposures with their counterparts in the Reference Catalog. This correction ensures photometric accuracy at the sub-arcsecond level and supersedes the coarse astrometry computed during Level 1 pipeline processing. Leveraging multiple exposures, this processing unit can also derive an accurate field distortion model for each detector, which is built up and monitored throughout the mission. The corresponding inputs and outputs of this processing step are summarized in Table 10.

Table 10: Inputs and Outputs of the Fine Astrometry Module

Input	Output
Spectral images, Instrument distortion map,	WCS solution coefficients, Refined higher-order Simple Imaging Polynomial (SIP) polynomials,
Proper-motion-corrected Gaia reference catalog with predicted SPHEREx fluxes	Fine Astrometry flag (FINAST)

Input and Output Description The input to this module includes spectral FITS images from all six detectors, each consisting of a $2,040 \times 2,040$ pixel array. These images contain initial coarse astrometric solutions encoded as WCS coefficients in their headers, as described in Section 3.1, accurate to a few arc-minutes. The Gaia DR3 catalog [Vallenari et al., 2023] serves as the astrometric reference. To estimate SPHEREx fluxes, Gaia entries are matched to Wide-field Infrared Survey Explorer (WISE) and Two Micron All Sky Survey (2MASS) photometry. The catalog is pre-processed to flag high-confidence calibration stars – non-saturated, high S/N, and spatially isolated in SPHEREx bands. Proper motions provided by Gaia are used to propagate stellar positions to the observation epoch; sources with poorly constrained proper motions are excluded. Additionally, the module requires estimates of the spacecraft and Earth velocity vectors relative to the solar system barycenter to apply a precise differential aberration correction.

The output includes refined WCS coefficients and updated SIP distortion terms for each detector, which are written to the headers of the corresponding spectral image files (see Table A.20). In addition, an astrometry flag (FINAST) is added to the header, indicating a good (0) or poor (>0) astrometric solution.

Description of Algorithm The algorithms employed by this module are based on the `AstroPy`⁴ Python library, the `Astrometry.net`⁵ suite, the `scamp` software [Bertin, 2006], and the WISE astrometry pipeline⁶. The astrometric calibrators are selected to have (i) good Gaia astrometric solution [*astrometric_params_solved* ≥ 31 in the Gaia catalog Vallenari et al., 2023]; (ii) S/N > 10 but not saturated in bands Gaia *RP*, *izy*, *JHK_s*, and WISE *W1/W2*; and (iii) no sources within 2 SPHEREx pixels that are brighter than 10 % of the flux of the calibrator. Figure 15 illustrates the overall structure of the Fine Astrometry algorithm.

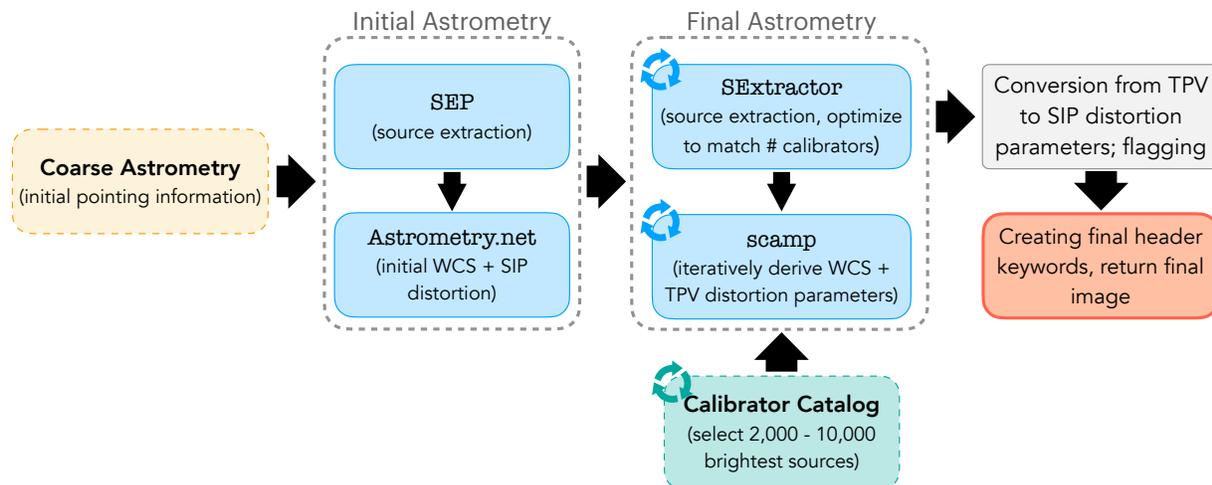


Figure 15: Schematic of the Fine Astrometry algorithm.

⁴<https://www.astropy.org/>
⁵<https://www.astrometry.net>
⁶http://wise2.ipac.caltech.edu/docs/doc_tree/sds/sds-PRex.pdf

The following summarizes the step-by-step implementation of the module:

1. The module initializes with a linear WCS solution (origin, translation, rotation) obtained from the Level 1 coarse astrometry estimation. This first-guess solution is derived from the planned pointing as described in Section 3.1 and is accurate to a few arc-minutes.
2. Using these initial pointing coordinates, we run the `solve-field` module from the `Astrometry.net` suite to derive an initial WCS solution including SIP distortion parameters. For the latter, we assume a 3rd order polynomial parameterization with a zero point $\text{CRPIX} = (1,020.5, 1,020.5)$ (center of the spectral image). We run `solve-field` on point sources extracted from the spectral image using the `sep`⁷ package and a threshold of 100σ . For the matching, we use the 4,100 and 5,200 index files⁸, which are based on Gaia DR3 and 2MASS stars. This second solution is accurate to a few arc-seconds.
3. In the next step, we extract the astrometric calibrators (see above) over the extent of the spectral image. From this calibrator selection, we select the N brightest (below saturation limit) calibrators and adjust the faint magnitude limit such that N is between 2,000 and 10,000. This adjustment is crucial to obtain a similar density of sources in crowded (e.g., galactic plane) vs. less crowded (e.g., extragalactic) fields.
4. We use `SExtractor` to extract sources on the spectral image (with astrometric solution from above applied). At first, a pre-defined detection threshold depending on the galactic latitude is used. This threshold is refined by running `SExtractor` iteratively until the number of extracted sources is within 20 % of the number of calibrators. This again accounts for different source densities at different positions on the sky.
5. Finally, the calibrator catalog and `SExtractor` catalog (both in *LDAC* format) are fed to `scamp`, which is used to derive the final WCS and distortion solution for a given detector image, which is accurate to sub-arcsecond level. We run `scamp` iteratively (re-extracting the sources by running `SExtractor` on the image with updated astrometry) until the RMS in RA and Declination is less than $1''$ or the maximal number of iterations is reached.
6. The distortion is parameterized as Tangent plane projection with parameter value (PV) polynomial distortion terms (TPV) polynomials [defined in sky coordinates; Shupe et al., 2012] by `scamp`. In a final step, we use the `wcs-sip2pv` module from the `Astrometry.net` package to convert the TPV representation to SIP, defined in pixel coordinates (thus fixed from exposure to exposure).
7. The resulting WCS solution, along with the most recent SIP distortion terms, is written to the header of the output spectral image file. Additionally, a flag (`FINAST`) is added to quantify the accuracy of the derived astrometry.

The performance of the astrometry is evaluated by computing a sigma-clipped RMS of the spatial difference between derived positions and the position of known Gaia and 2MASS stars. The flag `FINAST` (provided in the image header) gives an overall estimate of the robustness of the solution based on the output of the last (best) `scamp` iteration. The flag has bits 0 for good astrometry ($\text{RMS} < 1''$), 1 for poor astrometry (offsets in corners due to bad distortion fit), and >1 for failed astrometry.

3.2.8 Source Masks

This processing unit generates an image-plane mask for each spectral image to identify pixels with significant flux contributions from known sources. The module operates on spectral images with valid WCS information, assigning mask bits in a dedicated flag array at pixel locations corresponding to known source positions. The mask array is returned as output, while the original image data remains unaltered. The corresponding inputs and outputs of this processing step are summarized in Table 11.

Input and Output Description The inputs to this module include a spectral image with dimensions of $2,040 \times 2,040$ pixels, along with WCS information derived from the astrometric solution described in Section 3.2.7. It also requires an associated mini-catalog that includes the pixel coordinates and estimated

⁷<https://sep.readthedocs.io/en/stable/>

⁸<https://data.astrometry.net/>

Table 11: Inputs and Outputs of the Source Mask Module

Input	Output
Spectral images, Reference Catalog information, SUR flags propagated from the lower-level data	Binary array of mask flags

brightnesses of all sources expected to appear in the image or the surrounding frame. This catalog is generated from the Reference Catalog and the known Solar System object catalog. To properly account for sources that may fall partially off the detector edge or produce frame edge ghosts, the mini-catalog includes all relevant sources within 0.4° of the array boundaries. The reference catalog input used by this module includes all necessary flags that determine which sources should be masked. In addition, this module requires the SUR flags propagated from the lower-level data processing.

The output is a binary mask array with the same dimensions as the input image ($2,040 \times 2,040$ pixels), where each bit encodes the presence of a masked source. The information contained in this mask array is used to update the pixel-level flag map described in Section 3.2.4 and is stored as metadata in the headers of spectral images (see Table A.20).

Description of Algorithm The current implementation of the module generates masks for astronomical sources only. Future versions of the module will include other effects, such as electrical crosstalk, DBS ghosts, and frame edge ghosts.

The module uniquely masks unresolved sources, such as stars and compact galaxies. Resolved objects such as large nearby galaxies or planetary nebulae are not currently included in the masking. Masking is based on the expected PSF profile in each band and the source brightness.

The algorithm iterates over all cataloged sources requiring masking. These sources are flagged in the SPHEREx Reference Catalog with a `spherex_class` bit value ≤ 7 . For each source, a masking radius is computed that is a function of the estimated brightness of the source at the array central wavelength. Pixels whose centers fall within this radius from the subpixel source location are masked by assigning the value `SOURCE`.

The radius as a function of magnitude – originally approximated with simulated images – was re-derived based on on-sky data from IOC. Circles were drawn around a large number of sources in narrow bins in magnitude until the enclosed regions were visually consistent with encompassing the full source fluxes. Different image scalings were used to ensure the faint wings of the PSF were being masked for brighter sources. The measurements of radius as a function of magnitude were recorded, and a functional form was found to fit the measurements. The function is a two-exponential function model with five parameters total for each array:

$$r(m) = a \cdot \exp(b * m) + c \cdot \exp(d * m) + e \quad (13)$$

The best-fit parameters for each array are summarized in Table 12, while Table 13 summarizes the parameter values implemented in the source masking algorithm.

Table 12: Parameters fit to mask radius scaling in Eq. 13

Band	a	b	c	d	e
1	1549.4	-0.568	21.7	-0.114	-1.17
2	2687.3	-0.700	452.1	-0.003	-430.3
3	1129.4	-0.496	396.4	-0.0002	-392.3
4	2277.9	-0.575	1102.5	-0.0003	-1095.0
5	1072.98	-0.44	11491.65	$5e-5$	-11499.76
6	1291.4	-0.522	1579.8	-0.0002	-1571.3

Table 13: Implemented Algorithm Parameters

Parameter	Value
Masking radius as a function of source brightness	see Table 12
Crosstalk masking photocurrent threshold	$\geq 400 \text{ e}^-/\text{s}$
Crosstalk coupling amplitude threshold	$A_{ij} \geq 10^{-4}$
Crosstalk masking span	$\{-1, 0, +1\}$ pixels in fast read direction
DBS ghost masking source flux threshold	$F_{W1} \geq 2 \times 10^8 = \mu\text{Jy}$
Frame edge ghost source brightness threshold	$m_{AB} \leq 11$
Frame edge ghost mask length	$\left[11 \cdot (25,000 \cdot 10^{-\frac{m_{AB}}{2.5}}) + 24 \right]$ pixels
Frame edge ghost mask width	$2 \cdot \left[4 \cdot (25,000 \cdot 10^{-\frac{m_{AB}}{2.5}}) + 1 \right]$ pixels

3.2.9 Outlier Pixel Detection

This processing unit detects and flags transient events in each spectral image that may have bypassed onboard flagging procedures. It operates by applying a sigma-clipping algorithm to a median-filtered image with known sources masked. The output of this module contributes to updating the bad pixel inventory and is stored as metadata in the headers of spectral images (see Table A.20). The corresponding inputs and outputs of this processing step are summarized in Table 14.

Table 14: Inputs and Outputs of the Outlier Pixel Detection Module

Input	Output
Spectral images, SOURCE mask, Bad pixel mask, VARIANCE mask	New pixel mask images

Input and Output Description The inputs to this module consist of a spectral image with dimensions of $2,040 \times 2,040$ pixels, along with the associated SOURCE mask, bad pixel mask, and variance estimates.

For each exposure, the module generates a flag array matching the input image dimensions. This output array includes a new flag type (OUTLIERS), identifying statistically significant deviations from expected signal levels. Definitions of these flag values are listed in Table 9 in Sect. 3.2.4.

Description of Algorithm The algorithm begins by applying a median filter to the spectral image to suppress the overall mean signal and remove low spatial frequency variations. It then applies the SOURCE and bad pixel masks (Sect. 3.2.4) to exclude any pixels already known to be unreliable or associated with astrophysical sources. From the remaining unmasked pixels, the module identifies statistical outliers: any pixel whose signal exceeds n_σ times the standard deviation, as defined by the variance map. These outliers are flagged in a $2,040 \times 2,040$ binary mask.

3.2.10 Zodi Prediction

The Zodiacal light model is intended to provide an estimate of the diffuse component of the ZL in the SPHEREx LVF images. Specifically, the module looks up the ZL at each LVF pixel from an empirical ZL model, which encodes both the large-scale spatial and spectral gradients of the ZL. Importantly, the ZL model *is not* derived from the SPHEREx data itself. Instead, it is an empirical model that is based on the Kelsall et al. [1998] spatial model (updated with Planck data) and the spectral measurements compiled in Tsumura et al. [2013]. For a more detailed description of the model, see also Crill et al. [2025]. The ZL model will be provided as an additional $2,040 \times 2,040$ px layer for each LVF multi-extension image in units

of MJy/sr. Importantly, the ZL model *is not* subtracted from the LVF images. See also Section 2 for more information on the LVF image layers.

3.2.11 Level 2 DQA

A series of simple, automated “sanity” checks are executed to verify the quality of the data, following the L2 processing, and before the data can be ingested into the archive.

The following tests are performed on a per-image basis:

- **Nonlinearity:** No pixel in the image has a measured flux corrected by more than 50 %
- **Variance:** The measured per-pixel variance in the image is within the expected range
- **Persistence:** No more than 20 % of the pixels in the image are flagged with the PERSIST flag
- **Bad-Pixel Mask:** No more than 1 % of the pixels in the image are flagged;
- **Source Count:** The number of astronomical sources found in the image is within the expected range
- **Source Mask:** No more than of the pixels in the image are flagged as SOURCE
- **Outlier Pixels:** The number of pixels flagged as `textttOUTLIER` is within a factor of 2 of the total number of pixels already flagged for other reasons
- **LVF layers:** All required layers (data, flags, uncertainty, ZL, PSF cube) are present and correctly attached to the LVF file
- **Missing Value Pixels:** All NaN or Inf pixels are properly flagged
- **Distortion:** The image median of the pixel solid angles is within the expected range
- **Astrometry:** The image astrometry has been successfully computed (`FINAST = 0`).

If at least half of the eleven checks above are successful, the image is assigned an overall PASS grade. Overall, a spectral image is included in the archive at IRSA if the following three conditions are verified:

- `N_TRANSIENTS < 435,000`
- Not included in the defined boundaries for the South Atlantic Anomaly (SAA).
- `FINAST flag = 0` (Sec. 3.2.7)
- `L2DQA = PASS`

3.3 Photometric Measurements

The photometry module described here is used to derive photometric measurements used as part deriving the absolute gain matrix (Sect. 3.2.6). It will also be used as part of a user tool that will be provided later in summer 2025 to allow users to make photometric measurements of user-selected sources from the spectral images. Here we describe how those photometric measurements are made.

This processing unit performs photometric measurements of astrophysical sources by leveraging astrometric and source morphology information provided by the user, along with the pixel-level flag information propagated from preceding modules. The primary objective is to obtain optimally deblended photometry using prior knowledge. The corresponding inputs and outputs of this processing step are summarized in Table 15.

Input and Output Description The inputs to this module include spectral images, each containing a WCS solution capable of translating RA/DEC coordinates to pixel positions and to the associated wavelength. These images are dark-subtracted, flat-fielded, and brought to the physical unit of flux surface density (MJy/sr), enabling conversion of photometric measurements to absolute fluxes. The module works by finding the maximum likelihood solution for source fluxes by comparing the measured image with a generative model that predicts the scene given the instrument PSF and known source positions. While not required, intrinsic galaxy morphology can be supplied for the generative modeling. Additional inputs used by the photometry tool include the variance map (Sect. 3.2.2), the per-frame PSF model (Sect. 4.2.3), a pixel solid angle map

Table 15: Inputs and Outputs of the Forced Photometry Module

Input	Output
Spectral images, Source information, PSF model, Variance map, Pixel solid angle map, Flag map,	Photometry catalog

incorporating optical distortion information, and flag maps (Sects. 3.2.4 and 3.2.9). Eventually, a derived sub-pixel response map may be incorporated as well, but is not currently implemented.

The output of this module is a catalog of fluxes for all objects provided by the user. For each source, the catalog reports the source ID (as defined in the input catalog), measured flux and associated uncertainty, the corresponding wavelength, and the (i, j) pixel coordinates. It also includes fit quality flags and a goodness-of-fit metric derived from the residual between the observed and modeled image. Please note that the best performance will in general be obtained if all sources that may contribute substantial flux in a given scene are provided for the fit.

Description of Algorithm The algorithm is based on the **Tractor** software package⁹, which uses generative modeling of astronomical images to measure a maximum likelihood solution. The **Tractor** constructs a model of the image by combining this prior information with instrument characteristics and then optimizes the model parameters by fitting to the observed data. This forward modeling approach has been adopted in other tools, such as **xID** [Hurley et al., 2017] and **T-PHOT** [Merlin et al., 2016], which are tailored for crowded-field photometry with point sources and extended morphologies, respectively. The flexibility of the **Tractor** makes it well-suited for SPHEREx photometry.

Background subtraction Before forced photometry is performed, the spectral image must be background subtracted in the region containing the source. The default for the user tool is to perform local background subtraction. For this procedure, a box of size 15×15 pixels around the source is considered, and the background is inferred from the flux in the pixels not masked as containing sources. This local background is then removed from the cutout prior to photometry.

Forced photometry We use the **Tractor** in forced-photometry mode, applying it to sources with known positions and morphological models provided by the user. Source positions are not floated in the fit. The statistical foundations and modeling framework are described in Hogg et al. [2010], while the underlying optimization methods are based on the **emcee** affine-invariant ensemble sampler [Foreman-Mackey et al., 2013].

The following outlines the step-by-step implementation of the photometry module:

1. Load the input catalogs containing source positions, flags, and optionally, morphological parameters.
2. Convert celestial coordinates to detector pixel coordinates (i, j) .
3. Construct a cutout containing the source(s) of interest
4. Compute and remove a local background from a 15×15 region surrounding the cutout after masking known sources.
5. Retrieve the high-resolution, position-dependent PSF model for the corresponding detector region.
6. Use the **Tractor** to construct a generative model of the cutout, incorporating source positions, morphological models, and the local PSF.
7. Fit the generative model to the cutout data and associated error map by adjusting source fluxes.

⁹<http://thetractor.org>, <https://github.com/dstndstn/tractor>

⁹Computed using the **Background2D** class from **photutils**

8. Convert the fluxes reported by the `Tractor` to μJy using the local pixel solid angle, derived from the distortion map.
9. For each source in the central cutout, record the fitted flux, flux uncertainty, and corresponding wavelength.
10. Check for flagged pixels affecting each object and propagate the relevant flags to the output catalog. Apply additional photometry-specific flags indicating potential issues, such as strong non-linearity or saturation.
11. Compute a fit quality metric (as defined in Note #3) based on the residual between the model and data within a defined aperture around each source.

Fit Quality Metric A simple metric will be reported for each photometered source based on the residual between the model and the image in comparison with the per-pixel uncertainty. For each source, we define a `CirclePixelRegion`¹⁰ around its subpixel position of radius 2.5 SPHEREx pixels. For pixels k falling within this region, we compute:

$$\text{fit_quality} = \langle |p_k - m_k| / \sigma_k \rangle, \quad (14)$$

where p_k is the pixel value in the Level 3 image, m_k is the `Tractor` model value at pixel k , and σ_k is the uncertainty in pixel k from the variance map. The fit quality should tend to be close to 1 for good fits, and will be significantly larger if something went wrong.

¹⁰From the `AstroPy Regions` package.

4 Calibrations

4.1 Calibration Products

Alongside its science products, SPHEREx generates calibration products that are routinely updated through the data processing pipelines (Sect. 3). These calibration datasets are delivered to the IRSA at regular intervals throughout the mission to ensure continued public access and scientific usability. The data processing pipeline and calibrations used for the QR products are initial versions, and both will be updated for later data releases.

4.1.1 Absolute Gain Matrix

The absolute gain matrix is a calibration product generated by the SSDC Level 2 pipeline (Sect. 3.2.6). It characterizes pixel-to-pixel gain variations within a given spectral channel k as well as relative gain differences between spectral channels. This product is delivered in parallel with the spectral images (Sect. 2.1).

Data Format Each matrix (one per detector) encodes the absolute gain G_{ij} for each pixel (i, j) , calibrated in units of $(\text{MJy/sr}) / (e^-/\text{s})$. It is distributed as a FITS image file with dimensions of $2,040 \times 2,040$ pixels. This file serves as the input to the Absolute Gain Correction module of the Level 2 pipeline (see Fig. 5).

Image Header Description The header of an absolute gain matrix file encodes key metadata, including the array dimensions and the physical units of the pixel values. A representative header structure is shown in Table B.1.

Data Product Size Each absolute gain matrix file is approximately 16 MB in size.

4.1.2 Exposure-Averaged Point Spread Functions (PSFs)

The average PSF is a calibration product generated by the SSDC Level 2 pipeline (Sect. 4.2.3). It provides “super-resolution” PSF estimates on an 11×11 grid across each detector image. 121 zones are chosen to track the spatial and wavelength-dependent variations of the PSF across the detectors. The PSFs are constructed by first stacking cutouts of individual stars in each of the zones. Several exposures are used to obtain $>1,000$ stars per zone and per detector. The stacking is done at $10\times$ oversampled pixel scale as the PSF is significantly undersampled at SPHEREx resolution. In a last step, the resulting super-resolution stack is iteratively deconvolved by the pixel grid function, which takes into account the spread of the flux over the super-resolved pixels (Section 4.2.3). The average PSF product (including 121 zones for all six detectors) is released alongside the spectral images (Section 2.1).

Data Format Each PSF is a two-dimensional array with a size of 101×101 pixels at an oversampling rate of 10 (i.e., 10 oversampled pixels equals one SPHEREx pixel, or corresponds to a pixel scale of $0.615''/\text{px}$). There are 121 PSFs in total, corresponding to the 121 zones (11×11 across image) on a detector. The PSFs are provided as a FITS data cube with dimensions $121 \times 101 \times 101$. Each PSF is normalized by its integral and therefore is unitless.

Image Header Description The header of an average PSF file encodes key metadata, including astrometric reference parameters relative to the PSF image center as well as the center positions (in image coordinates) and widths of the oversampled zones across the detector. A representative header structure is shown in Table C.1; some values are representative of the specific product illustrated.

Data Product Size Each average PSF image file is approximately 6 MB in size.

4.1.3 Dark Current

The dark current maps are calibration products generated by the SSDC Level 2 pipeline (Sect. 3.2.5).

Data Format Each dark current map (one per detector) encodes the dark current for each pixel (i, j) in units of e^-/s . They are distributed as FITS image files with dimensions of $2,040 \times 2,040$ pixels. These files serve as the input to the Dark Subtraction module of the Level 2 pipeline (see Fig. 5).

Data Product Size Each absolute gain matrix file is approximately 16 MB in size.

4.2 Calibration Procedures

4.2.1 Photometric/Absolute Gain Calibration Derivation

The absolute gain matrix is derived by combining a flat field / dark measurement with measured gain factors from primary calibrator stars.

Derivation of flat-field / dark current The flat field response of each detector is derived by fitting the slope of each pixel value in e^-/s against a reference flux level measured from the median brightness in the pixel's spectral channel in each image. We define 510 nominal spectral channels per detector for this procedure, such that the flux of the background in each channel is expected not to significantly vary in the spectral dimension. Because the detectors are always oriented the same way relative to the Sun, we need to fit an additional tilt to the flux in each spectral channel to account for solar elongation, i.e., enhanced brightness of the ZL background on one side of the detector.

Each pixel value is then compared with the global flux prediction at its location. For each image, we only consider pixels expected not to have a significant source flux contribution. The slope of the relation over many images for each pixel represents the relative pixel response, or flat field, within the spectral channel, while the y-intercept of the line fit gives the dark current in e^-/s . We find that $\sim 1,000$ images are sufficient to yield a robust dark current estimate and flat field response. Images for this procedure are selected to avoid ones close to the galactic plane ($b < 20^\circ$) or ones that have high intrinsic variation. The derived flat field responses per detector are shown in Figure 16.

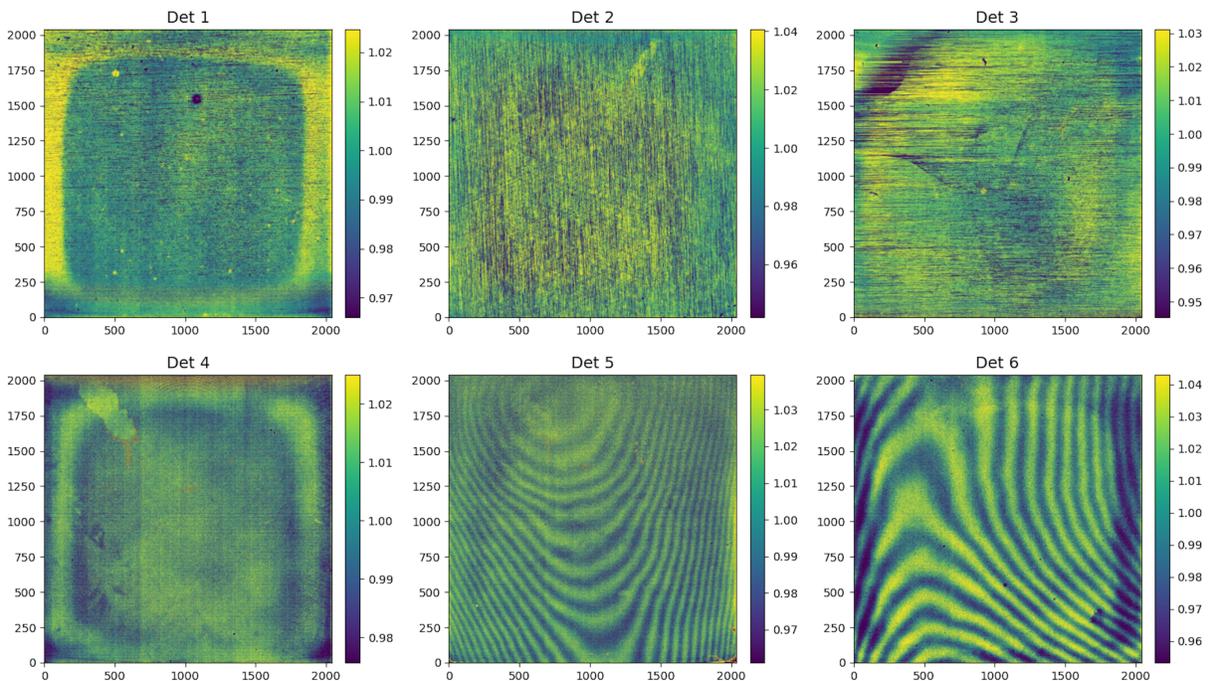


Figure 16: The relative flat field response derived for the six SPHEREx detectors.

Absolute gain measurement The absolute gain is determined by examining the photometry of 17 pre-defined photometric calibration stars as they are imaged across each detector. The photometry is converted back to e^-/s using the previously applied gain matrix, and compared with the expected flux from the intrinsic star spectrum integrated against the SPHEREx bandpass at the location where it was measured. Each measurement yields a single gain term for a particular spectral channel. The 17 spectrophotometric standard stars were selected from the CALSPEC database [Bohlin et al., 2014, 2017, 2022] to meet our requirements on brightness, summarized in Table 16. The absolute calibration of the QR data release is based on nine of these calibration sources.

Table 16: Stellar magnitude criteria used to select calibrator stars for SPHEREx.

Band	Wavelength Range (μm)	Saturation [†] Limits	High-SNR Threshold (50σ)
1	0.75–1.11	11.5	14.74
2	1.11–1.64	11.5	14.79
3	1.64–2.42	11.3	14.61
4	2.42–3.82	11.71	14.85
5	3.82–4.42	10.19	13.46
6	4.42–5.00	9.75	13.17

Notes: [†]: Magnitudes at which nonlinearity effects are expected to manifest based on a pre-launch performance model. All magnitudes are in AB.

The set of stars used for absolute calibration is summarized in Table 17. Once a sufficient number of measurements of these stars have been acquired, it allows us to average the gains together to build a 1D gain function, i.e., the gain as a function of the spectral dimension on each detector. This gain function is then multiplied by the inverse of the derived flat-field response, yielding the absolute gain matrix. For the initial determination of the 1D gain function, approximately 2,700 individual measurements of primary calibrator stars were used.

For all standard stars included in the calibration, instrumental fluxes and the associated uncertainties

$$I_{ijk}, \sigma_{ijk},$$

at pixel p_{ijk} are regularly collected. In this formulation, i is the band index, which ranges from 1 to 6, j is the integer observation index, which increments every time a new observation occurs, and k is an index indicating the calibrator.

The SPHEREx spectral throughput function $t(p_{ijk}, \lambda)$ and the wavelength of peak response $\lambda_{\text{mean}}(p_{ijk})$ for every pixel p_{ijk} is known exactly due to the prelaunch measurements carried out by Hui et al. (2025, in preparation). For that reason, the high-resolution spectra $S_k(\lambda)$ of the calibrators can be convolved with the SPHEREx bandpass measured for the specific pixel p_{ijk} at which they were detected, to estimate the ‘true’ calibrator fluxes and mean wavelengths:

$$s_{ijk} = \frac{\int t(p_{ijk}, \lambda) S_k(\lambda) d\lambda}{\int t(p_{ijk}, \lambda) d\lambda}$$

$$\lambda_{\text{obs};ijk} = \lambda_{\text{mean}}(p_{ijk})$$

A color correction is estimated and applied for each observation of each calibrator:

$$k_{\text{zodi};ijk} = \frac{\int t(p_{ijk}, \lambda) z(\lambda) d\lambda}{\int t(p_{ijk}, \lambda) d\lambda}$$

The precomputed absolute gain matrix $G_{\text{abs}}(p_{ijk})$ (Ashby et al. 2025, in preparation) is accessed, as is a precomputed lookup table containing the solid angle associated with each measurement $\Omega_{ijk} = \Omega_{\text{pix}}(p_{ijk})$. Likewise, a reference pixel size Ω_{ref} , nominally the solid angle of the central pixel of detector

two, is retrieved, and then the previous calibration step is inverted to convert the flux back from Jy into the original engineering units

$$i_{ijk} = I_{ijk}/G_{abs;i}(p_{ijk})\Omega_{ref}/\Omega_{ijk}$$

Finally, a statistical weight W_{ijk} is computed:

$$W_{ijk} = 1/\sigma_{ijk}^2$$

This is repeated for the very large numbers of calibrator observations spanning the entire SPHEREx spectral bandpass with measurements of the ratios of the raw and predicted calibrator fluxes. These are collectively fitted with a 1D function to create the calibration function for the primary data as a function of wavelength. Combining these 1D calibration factors with the measured flat field response yields the absolute gain matrices used to convert the images from e^-/s to MJy/sr. The initial gain matrices are shown in Figure 17.

Calibration uncertainty The initial uncertainty in the calibration has been determined using repeat observations of primary calibrator stars. The calibration appears reliable at the $\sim 10\%$ level currently. It will be improved and validated further as data is accumulated on primary and secondary calibration stars.

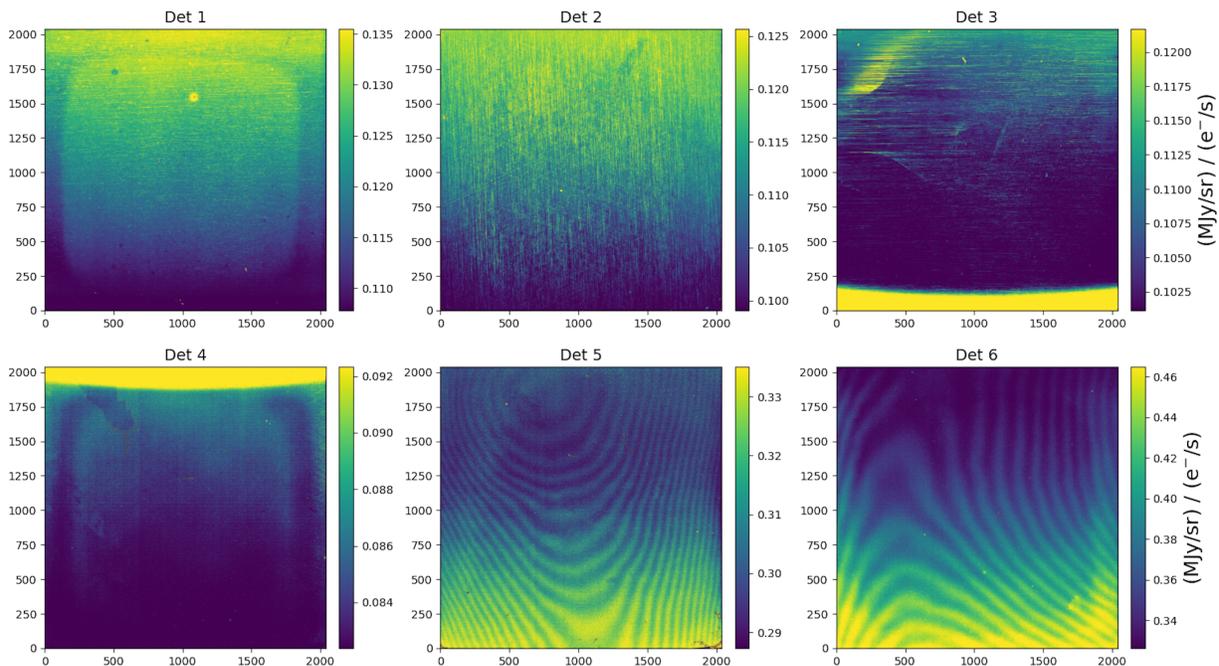


Figure 17: The absolute gain matrices used to convert SPHEREx images from engineering to physical units.

4.2.2 Astrometric Calibration Derivation

The initial pointing (coarse astrometry) for each exposure of each detector is derived from the planned pointing and rotation angle (Sect. 3.1). The initial pointing accuracy is between $2'$ and $10'$. From this, the final WCS linear astrometry solution, including the SIP distortion coefficients resulting in sub-arcsecond accuracy, is derived during the Fine Astrometry procedure (Sect. 3.2.7). In brief, during this two-step approach, an initial WCS+SIP solution is found using the `astrometry.net` package, which is then refined by iteratively running the `scamp` software. As astrometric calibrators, proper motion-corrected Gaia stars are used. The derivation of an astrometric solution is challenging as SPHEREx observes regions of different star densities (e.g., compare the galactic bulge to an extragalactic field where the stellar density differs by Catech/IPAC

Table 17: SPHEREx Photometric Standard Stars.

Star	Type	Band-Averaged Magnitudes (AB)						$\sigma(\text{sys})$
		Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	
BD+29 2091 [‡]	F5	9.94	9.97	10.32	11.04	11.76	12.03	2–4%
1802271 ^{*‡}	A2V	12.39	12.84	13.43	14.22	14.94	15.22	2–4%
Feige110 [†]	sdO	12.70	13.51	14.33	14.90	2–4%
AGK+81 266 [†]	sdO	12.84	13.68	14.50	15.08	2–4%
1732526 [‡]	A4V	12.84	13.23	13.87	14.59	15.31	15.59	2–4%
GRW+70 5824 ^{†‡}	DA2.4	13.46	14.23	14.97	1%
P330E [*]	G2V	12.66	12.70	13.02	13.78	14.49	14.79	1%
WD2341+322	DA3.8	13.51	14.10	1%
P177D	G0V	13.12	13.15	13.49	14.25	14.95	15.25	2–4%
The three primary <i>HST</i> calibrators. They are all white dwarfs, and very blue.								
G191B2B	DA.8	12.62	13.51	14.29	14.89	~ 16.0	~ 16.2	1%
GD71	DA1.5	13.84	14.70	15.48	16.05	~ 17.2	~ 17.5	1%
GD153	DA1.2	14.17	15.05	15.82	16.41	~ 17.5	~ 17.8	1%
The <i>JWST</i> Cycle 1 calibrator targets that lie in the NEP EBL field.								
1743045 ^{*‡}	A8III	13.63	13.91	14.47	15.20	15.93	16.20	2–4%
1757132 ^{*‡}	A3V	11.94	12.26	12.82	13.55	14.27	14.55	2–4%
1805292 ^{*‡}	A4V	12.61	13.00	13.63	14.36	15.09	15.36	2–4%
1808347 ^{*‡}	A3V	12.29	12.62	13.20	13.93	14.66	14.93	2–4%
1812095 ^{*‡}	A5V	11.96	12.32	12.92	13.65	14.37	14.65	2–4%
TOTAL STARS:		16	15	12	8	1	1	
TOTAL STARS w/ NEP:		16	15	12	9	6	6	

Notes: All suitable stars from the *HST*/CALSPEC database known to satisfy the brightness and saturation requirements in at least one SPHEREx band. **Bold** indicates either S/N < 50 per observation, or that the star is so bright that it intrudes into the nonlinear regime (BD+29 2091). The models and spectra for all stars listed have been inspected; No other validation has yet been done. In-Band average magnitudes computed only using SPHEREx observations with S/N > 50. † : Model not spanning all six SPHEREx Bands. *: Observed with NIRSpec in *JWST* Cycle 1 as part of PID 1678 or as part of the *JWST* instrumental cross-calibration program. ‡: Used for absolute calibration of the QR data.

several 100–1,000 \times). An iterative approach to select the calibrators and extract sources on the SPHEREx images optimizes this approach, resulting in a very low fraction of failures (<3 % currently, but this number will be improved in the future). A significant fraction of failures is due to excessive numbers of pixels flagged as transients. Furthermore, the large distortion corrections (up to 15 SPHEREx pixels corresponding to $\sim 90''$ in the image corners) require an iterative approach to the final astrometric solution.

The final astrometric accuracy is tested and quantified by re-extracting bright stars on the images and matching them to the Gaia and 2MASS catalogs. Here we focus on exposures that are not in the SAA and have less than 300,000 pixels flagged as transients (L2 N_TRANSIENT flag). The left panel of Figure 18 shows the resulting difference in RA and DEC between extracted star positions and Gaia/2MASS stars for detector 3. We find RMS values of $<0.3''$. No significant changes in the residual are observed towards the corners (where larger distortion is expected), suggesting a robust astrometric solution. The right panel of Figure 18 shows the average distortion for each detector in pixels (1 px $\sim 6.15''$) derived from several hundred exposures. The maximum distortion in the corners of the exposures is ~ 13 px or $\sim 80''$.

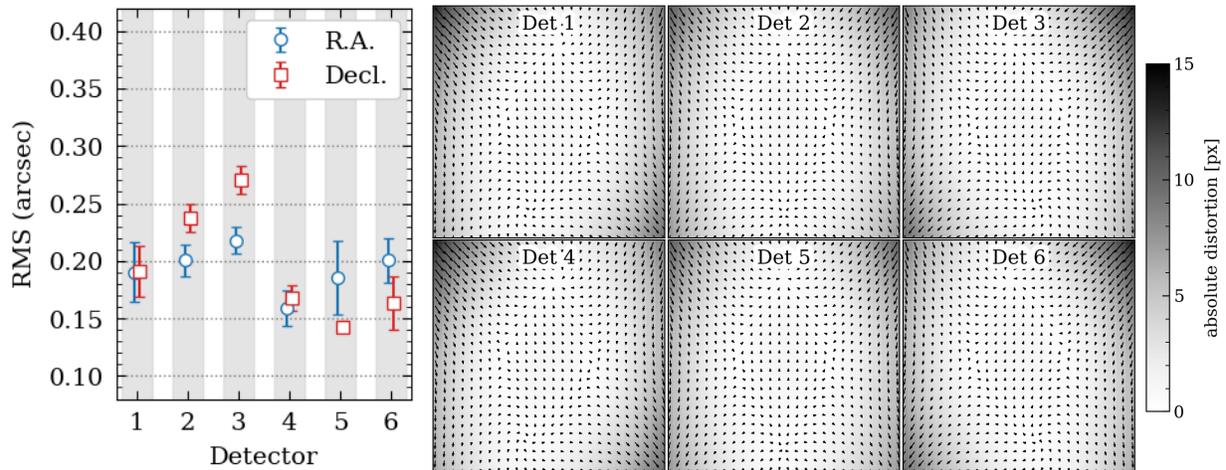


Figure 18: Left: SPHEREx astrometric performance measured after the Level 2 astrometric correction module, showing positional uncertainties (RMS, in arcsec) in Right Ascension (RA, blue circles) and Declination (Decl, red squares) for each detector. The performance has been measured by matching extracted stellar sources to the Gaia DR3 catalog. No statistical difference in performance is observed between the center and the corners of the detector, suggesting a well-fitted distortion solution. Right: Average distortion (in pixels) and its direction obtained across each SPHEREx detector by the Fine Astrometry module.

4.2.3 PSF Measurement

This processing unit generates a finely sampled grid of position – and thus wavelength-dependent PSF measurements across each detector array by averaging data from several exposures (see left panel of Fig. 19). These PSF products constitute one of the principal deliverables of the QR (Sect. 4.1.2). The resulting PSFs are *effective* in nature, capturing not only the intrinsic optical response but also the average blurring effects introduced by telescope jitter and pointing instability accumulated over multiple observations. The corresponding inputs and outputs of this processing step are summarized in Table 18.

Table 18: Inputs and Outputs of the PSF Measurement Module

Input	Output
Spectral images, Fine astrometry information, Gaia DR3 catalog	Effective PSFs

Input and Output File Description The inputs to this module include fully corrected and calibrated individual LVF exposures in units of MJy/sr, formatted as $2,040 \times 2,040$ detector pixel arrays. Accurate astrometric solutions from the Fine Astrometry processing step (Sect. 3.2.7) are crucial to identify calibrators (point sources) for this processing step. The point sources for measuring the PSF are taken from the same calibrator catalog as for deriving the astrometric solution – including isolated, high-S/N stars from Gaia DR3 [Vallenari et al., 2023] with robust proper motion determinations – as described in Section 3.2.7. In addition, the flag layer that is part of the images is used to reject calibrator stars that have flagged pixels nearby (see below).

The output is a library of effective PSFs – representing the convolution of the optical PSF with average pointing jitter – deconvolved with the detector pixel response for each detector. Each PSF is normalized to unit integral and centered using the light distribution’s zeroth moment. The PSFs are sampled on a 10×10 oversampled pixel grid with respect to the native SPHEREx pixel size, yielding 101×101 pixel images (i.e., a pixel scale of $0.615''/\text{px}$). For each detector, we produce a library of 121 PSFs. Each of them captures the

spatial and wavelength dependence across the detector in 11×11 zones. These libraries are stored as one of the layers in the corresponding spectral images (Sect. 2.1).

Description of Algorithm The main algorithm is based on Symons et al. [2021]. The right panel of Figure 19 shows a schematic illustration of the PSF construction workflow. In the following, we briefly outline the main steps.

1. **Identification and binning of point sources:** Point sources from the calibrator catalog are identified across N_{exp} exposures (typically $N_{\text{exp}} \sim 1,000$) of a given detector. To account for spatial and wavelength-dependent variations in the PSF per detector, the sources are grouped into 11×11 zones across the image. Different zoning (also differing per axis) can be used depending on the PSF variations across the image. Each zone has a width and a center (x, y) pixel coordinate, which is recorded in the header of the final data product (see also Table C.1).
2. **Construction of exposure-averaged effective PSFs:** In each zone \mathcal{Z} (defined in pixel space), point sources from the calibrator catalog are selected and cut out from N_{exp} exposures (labeled by k). These cutouts are stacked on a $10 \times$ oversampled pixel grid (see below) to obtain a stacked high-resolution exposure-averaged PSF $P_{\mathcal{Z}}^{\text{S}}$. Importantly, for each zone, $P_{\mathcal{Z}}^{\text{S}}$ is a convolution of the true optical PSF ($P_{\mathcal{Z}}^{\text{opt}}$) with the pixel response function (P^{pix}) and the exposure-averaged jitter ($\langle P_k^{\text{jit}} \rangle_k$) caused by spacecraft motion:

$$P_{\mathcal{Z}}^{\text{S}} = P_{\mathcal{Z}}^{\text{opt}} * P^{\text{pix}} * \langle P_k^{\text{jit}} \rangle_k .$$

Note that the pixel response function is assumed to be a constant, and the jitter is assumed to be constant per exposure (assuming the focal plane is a rigid body). The effective exposure-averaged PSF is then obtained by deconvolving the stack with the pixel response function:

$$P_{\mathcal{Z}}^{\text{eff}} = P_{\mathcal{Z}}^{\text{S}} * (P^{\text{pix}})^{-1} .$$

As described below, we do not correct for possible differences in the jitter between different exposures, as the current in-flight data shows that the jitter is negligible.

3. **Normalization and centering:** Each produced PSF is normalized to its sum and centered (e.g., based on the zeroth moment of light).
4. **PSF library generation:** The final effective PSFs are compiled into a library, which is a 3-dimensional array of size $121 \times 101 \times 101$.

Two key algorithmic operations are the generation of oversampled (super-resolution) stacks from point sources and the deconvolution of these stacks with the pixel response function. In the following, we describe in more detail each of those two algorithmic components.

- **Creation of Stacked “super-resolution” PSF:** The stacking method for the SPHEREx PSF reconstruction takes advantage of the fact that an “ideal” gridded image of the astronomical sky can be represented as:

$$\mathcal{F}^j = \mathcal{N}^j + \sum_{\alpha} S_{\alpha} (N_{\alpha}^j - \mu_{\alpha}), \quad (15)$$

where \mathcal{F}^j and \mathcal{N}^j is the brightness and the noise, respectively on an image in a given pixel j , S_{α} is the list of the fluxes of all sources of emission, N_{α}^j is the number of sources falling into pixel j from list α , and μ_{α} is the average number of sources in list α [Marsden et al., 2009, Viero et al., 2013]. Instruments with non-zero extended PSFs contribute flux into more than one pixel. We define $P^j(x_{\alpha}, y_{\alpha})$ to be the beam-convolved and mean subtracted value of the PSF at pixel j , centered at some source position (x_{α}, y_{α}) , and then write Equation 15 as:

$$\mathcal{F}^j = \mathcal{N}^j + \sum_{\alpha} S_{\alpha} P^j(x_{\alpha}, y_{\alpha}) . \quad (16)$$

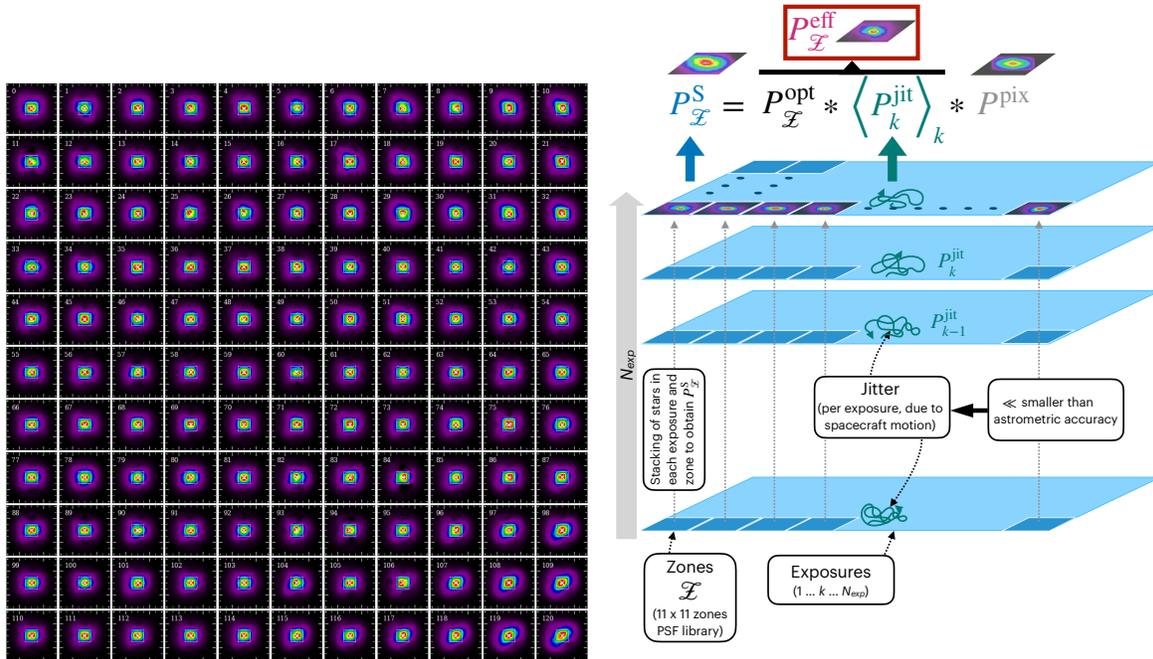


Figure 19: Left: Example 11×11 PSF layer for detector 3. The PSFs are $10\times$ oversampled with respect to the native SPHEREx pixel scale. The white square denotes a SPHEREx pixel. Notice the elongated shape in the lower right corner due to spatial variations on the detector. Right: Schematic illustration of the PSF construction workflow, highlighting the contributing components and variables.

This expression accounts for the fact that all sources can contribute to the flux in pixel j (weighted by the PSF’s spatial profile), since the PSF spreads flux to neighboring pixels.

Our goal is to estimate the shape of the PSF P from the measured sky image \mathcal{F} . Since P has the same amplitude for each source α , we can invert Equation 16 and solve:

$$P = \sum_{\alpha} \frac{\mathcal{F}(x_{\alpha}, y_{\alpha})}{S_{\alpha}}, \quad (17)$$

where $\mathcal{F}(x_{\alpha}, y_{\alpha})$ is the image centered on the source position (x_{α}, y_{α}) , and we assume that the noise obeys $\langle \mathcal{N} \rangle = 0$ over the sum. Furthermore, as with simple stacking, we require the source positions to be uncorrelated so that contributions from sources in the image but not being stacked on do not add coherently [Marsden et al., 2009].

The super-resolution recovery comes about in this formalism because of the nature of pixels, which have the property of averaging photons. That is, (in a purely image processing sense) one can always create larger pixels which contain more photons, but spread over a larger area, in such a way that the measured surface brightness is conserved. The corollary for smaller pixels from larger also holds. The cost of this regridding operation is changes in the noise: regridding larger pixels to smaller increases the noise per pixel, while regridding smaller pixels to larger decreases the noise in the larger pixel.

This averaging property of pixels allows us to write the relation:

$$P^j = \frac{1}{\Omega} \sum_i \tilde{P}^i, \quad (18)$$

where \tilde{P}^i is the image of the PSF sampled on a finer grid and Ω is the scale factor between the areas of pixel i and j (in our case, for an oversampling of $10\times$, it is $\Omega = 10^2$). Since the pixel size does not

appear in Equation 17, we can write:

$$\tilde{P} = \sum_{\alpha} \frac{\tilde{\mathcal{F}}(x_{\alpha}, y_{\alpha})}{S_{\alpha}}, \quad (19)$$

where \tilde{M} is a version of the image on the finer grid, and is related to \mathcal{F} through $\mathcal{F}^j = \frac{1}{\Omega} \sum_i^{\Omega} \tilde{\mathcal{F}}^i$. The cost of this super-resolution stacking is that the per-pixel noise is increased over the native pixel resolution. This can easily be addressed by making the list α large to compensate, though for a fixed list size, the noise in smaller pixels will always be larger than in larger pixels.

In reality, the SPHEREx images are first regridded to a super-resolution fine grid (i.e., 10 times smaller pixels). The stars are then cut out at the fine grid pixels closest to their actual position. Note that the centers of the stars are not necessarily centered at the fine grid pixels, as the pixel centers are floating-point numbers and not integers. The cutouts are then stacked to obtain a super-resolution stack, which is called P^S (or $P_{\mathcal{Z}}^S$ for a given zone, see above).

- **PSF deconvolution of pixel response function:** The stacked PSF P^S (or $P_{\mathcal{Z}}^S$ per zone) is a convolution of the optical (P^{opt}), jitter (P^{jit}), and pixel response function (P^{pix}). The latter defines the distribution of flux of a point source on the pixels of the fine grid and is a known function. In order to reconstruct the effective PSF (i.e., $P^{\text{opt}} * P^{\text{jit}}$), we must deconvolve the stacked PSF with the pixel response function. To do this, we implemented an advanced form of the Richardson-Lucy deconvolution (Richardson-Lucy Deconvolution (RLD)) algorithm [Richardson, 1972, Lucy, 1974] as described in Biggs and Andrews [1997] (“accelerated damped RLD”). The algorithm is identical to the *Matlab* function `deconvlucy()` and faster than the conventional RLD algorithms. A similar approach has been used for super-resolution image reconstruction for blurred or degraded images [Park et al., 2003]. With the known P^{pix} defined as $P^{\text{pix}}(x, y) = [r - (x - x_0)][r - (y - y_0)]$ (a matrix where each pixel contains the percentage of light detected from a point source on the $r \times$ oversampled pixels, here $r = 10$, and (x_0, y_0) is the center of the stacking area), the deconvolution algorithm proceeds as follows:

$$P_{i+1}^{\text{RLD}} = P_i^{\text{RLD}} \left(\frac{P^S}{P_i^{\text{RLD}} * P^{\text{pix}}} * (P^{\text{pix}})^{\top} \right), \quad (20)$$

where the initial $P_{i=0}^{\text{RLD}}$ is simply P^S , and i increases on every iteration for N total iterations. $(P^{\text{pix}})^{\top}$ is the transposed matrix for P^{pix} . The final result is $P^{\text{RLD}} \equiv P^{\text{eff}}$, the deconvolution of P^S .

Due to the settling time of the spacecraft, we expect additional blurring due to jitter. The effective PSF includes the jitter component averaged over the N_{exp} exposures over which the stack (P^S) has been created. The jitter has a directional high-frequency (approx. Gaussian) and low-frequency (random walk) component due to the settling after a slew and vibrations. In principle, the jitter can be different (in amplitude and possibly also direction) from exposure to exposure. In-flight tests, however, show that the settling time is very rapid (<1 s) and the amplitude of the jitter is significantly below the astrometric accuracy ($0.2 - 0.3''$), thus negligible. We therefore do not try to deconvolve the resulting $P_{\mathcal{Z}}^{\text{eff}}$ by the jitter. However, we reserve this possibility for the future in case the jitter becomes more significant at later times of the mission.

4.3 Additional Products and Inputs

This section describes other calibration products or inputs to the pipeline that are available for users to download at IRSA.

4.3.1 Spectral WCS

The spectral WCS files contain the central wavelengths and bandwidths for each pixel, as extensions in a separate FITS file for each spectral band. The central wavelengths are provided in the `CWAVE` extension, and are set to the point where the cumulative transmission curve reaches 50 % of its total area (i.e., each pixel’s transmission curve’s median value). The bandwidths are provided in the `CBAND` extension, and are computed as `CWAVE` divided by the detector’s measured spectral resolution.

The per-pixel values in these extensions are also used to create a compact, piecewise-bilinear model of the spatial variation of the effective wavelength and bandwidth for each detector. This model is encoded using the FITS lookup-table “spectral WCS” standard as a `WAVE-TAB` WCS. This WCS is included in every spectral image, in order to allow an approximate readout of the central wavelength and bandwidth to be packaged with the images without the data volume for the full per-pixel data. The lookup table for the WCS is provided as a binary-table extension, `WCS-WAVE`, in the format mandated by the standard. The WCS and its lookup table are also included, for reference, in the spectral WCS files, to facilitate comparisons. (The bandwidth coordinate in the WCS was inappropriately named “Bandpass” in the WCS headers; this will be fixed in a future reprocessing.) Table 19 shows a representative header from the spectral WCS file for Detector 1.

Table 19: Spectral WCS FITS Header Keywords (Example)

Header	Value	Units	Description
<code>XTENSION</code>	<code>'IMAGE '</code>	–	Image extension
<code>BITPIX</code>	<code>–32</code>	–	Array data type
<code>NAXIS</code>	<code>2</code>	–	Number of array dimensions
<code>NAXIS1</code>	<code>2040</code>	–	–
<code>NAXIS2</code>	<code>2040</code>	–	–
<code>PCOUNT</code>	<code>0</code>	–	Number of parameters
<code>GCOUNT</code>	<code>1</code>	–	Number of groups
<code>ARRAY</code>	<code>1</code>	–	Detector array
<code>BUNIT</code>	<code>'um '</code>	–	Unit
<code>WCSAXES</code>	<code>2</code>	–	–
<code>WCSNAME</code>	<code>'Approximated per-pixel wave-length and bandpass'</code>	–	–
<code>CTYPE1</code>	<code>'WAVE-TAB'</code>	–	–
<code>CTYPE2</code>	<code>'WAVE-TAB'</code>	–	–
<code>CNAME1</code>	<code>'Wavelength'</code>	–	–
<code>CNAME2</code>	<code>'Bandpass'</code>	–	–
<code>CUNIT1</code>	<code>'um '</code>	–	–
<code>CUNIT2</code>	<code>'um '</code>	–	–
<code>CRPIX1</code>	<code>1</code>	–	–
<code>CRPIX2</code>	<code>1</code>	–	–
<code>CRVAL1</code>	<code>1</code>	–	–
<code>CRVAL2</code>	<code>1</code>	–	–
<code>CDELTA1</code>	<code>1</code>	–	–
<code>CDELTA2</code>	<code>1</code>	–	–
<code>PC1_1</code>	<code>1</code>	–	–
<code>PC1_2</code>	<code>0</code>	–	–
<code>PC2_1</code>	<code>0</code>	–	–
<code>PC2_2</code>	<code>1</code>	–	–
<code>PS1_0</code>	<code>'WCS-WAVE'</code>	–	Lookup table extension name
<code>PS2_0</code>	<code>'WCS-WAVE'</code>	–	Lookup table extension name
<code>PS1_1</code>	<code>'VALUES '</code>	–	Column name for the coordinate values array
<code>PS2_1</code>	<code>'VALUES '</code>	–	Column name for the coordinate values array
<code>PS1_2</code>	<code>'X '</code>	–	Column name for the indexing vector
<code>PS2_2</code>	<code>'Y '</code>	–	Column name for the indexing vector
<code>PS1_3</code>	<code>1</code>	–	Axis number in coordinate values array
<code>PS2_3</code>	<code>2</code>	–	Axis number in coordinate values array
<code>EXTNAME</code>	<code>'CWAVE '</code>	–	Extension name

4.3.2 Additional Products

Dichroic impacted pixels This bit mask indicates pixels where the flux attenuation due to the dichroic filter is 50 % or higher and corresponds to the `DICHOIC` flag set in Section 3.2.4. The pixel value is 0 for unaffected pixels and 1 for impacted pixels. Note that only Bands 3 and 4 have any non-zero values. These values are supplied by the SPHEREx Instrument Team.

- **Data Format:** This product is distributed as one FITS image file with dimensions of $2,040 \times 2,040$ pixels per band. Each file name starts with `dichroic_Dn` where n is the band number.

Nonfunctional pixels This bit mask indicates pixels known to be permanently non-functioning and corresponds to the `NONFUNC` flag set in Section 3.2.4. The pixel value is 0 for unaffected pixels and 1 for impacted pixels. These values are supplied by the SPHEREx Instrument Team.

- **Data Format:** This product is distributed as one FITS image file with dimensions of $2,040 \times 2,040$ pixels per band. Each file name starts with `nonfunc_Dn` where n is the band number.

Non-linearity parameters This file contains the input non-linearity calibration parameters described in Section 3.2.1. The `Q_n1` HDU contains the values used in the equations in Section 3.2.1. The other HDUs represent coefficients that are currently not used in the pipeline but may be in a future version.

- **Data Format:** This product is distributed as one FITS image file with dimensions of $2,040 \times 2,040$ pixels per band. Each file name starts with `nonlinear_pars_Dn` where n is the band number.

Electronic Gain Factors As described in Section 3.1, the Level 1 data processing includes converting the raw pixel data from ADU to the engineering units of electrons per second. These values are per amplifier channel and are provided by the SPHEREx Instrument Team.

- **Data Format:** These gains are provided in a single YAML file, which includes the provenance information for the detectors and a list of 32 gain values per detector. The file name starts with `gain_factors`.

Read noise parameters This file contains the per-pixel electronic noise used in the variance map calculation (Sect. 3.2.2). The per-pixel values are given in units of electrons and are provided by the SPHEREx Instrument Team.

- **Data Format:** This product is distributed as a single FITS image file with dimensions of $2,040 \times 2,040$ pixels per band. Each file name starts with `readnoise_pars_Dn`, where n is the band number. The file contains two image extensions:
 - `READOUT-1`: Per-pixel electronic noise for a single readout ($N = 1$).
 - `READOUT-2`: Per-pixel electronic noise for $N = 71$ total reads.

References

- Farah Alibay, Oleg V Sindiy, PA Trisha Jansma, Charles M Reynerson, Eric B Rice, Jennifer Rocca, Sara Susca, Stephen C Unwin, Rachel L Akeson, Shannon E Mihaly, et al. Spherex preliminary mission overview. In *2023 IEEE Aerospace Conference*, pages 1–18. IEEE, 2023.
- E Bertin. Automatic astrometric and photometric calibration with scamp. In *Astronomical Data Analysis Software and Systems XV*, volume 351, page 112, 2006.
- David S. C. Biggs and Mark Andrews. Acceleration of iterative image restoration algorithms. *ApOpt*, 36(8): 1766–1775, March 1997. doi: 10.1364/AO.36.001766.
- Ralph C. Bohlin, Karl D. Gordon, and P. E. Tremblay. Techniques and Review of Absolute Flux Calibration from the Ultraviolet to the Mid-Infrared. *PASP*, 126(942):711, August 2014. doi: 10.1086/677655.
- Ralph C. Bohlin, Szabolcs Mészáros, Scott W. Fleming, Karl D. Gordon, Anton M. Koekemoer, and József Kovács. A New Stellar Atmosphere Grid and Comparisons with HST/STIS CALSPEC Flux Distributions. *AJ*, 153(5):234, May 2017. doi: 10.3847/1538-3881/aa6ba9.
- Ralph C. Bohlin, Jessica E. Krick, Karl D. Gordon, and Ivan Hubeny. How Do Spitzer IRAC Fluxes Compare to HST CALSPEC? *AJ*, 164(1):10, July 2022. doi: 10.3847/1538-3881/ac6fe1.
- G. Brammer, N. Pirzkal, P. McCullough, and J. MacKenty. Time-varying Excess Earth-glow Backgrounds in the WFC3/IR Channel. Instrument Science Report WFC3 2014-03, 14 pages, April 2014.
- Brendan P Crill, Michael Werner, Rachel Akeson, Matthew Ashby, Lindsey Bleem, James J Bock, Sean Bryan, Jill Burnham, Joyce Byunh, Tzu-Ching Chang, et al. Spherex: Nasa’s near-infrared spectrophotometric all-sky survey. In *Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave*, volume 11443, pages 61–77. SPIE, 2020.
- Brendan P. Crill, Yoonsoo P. Bach, Sean A. Bryan, Jean Choppin de Janvry, Ari J. Cukierman, C. Darren Dowell, Spencer W. Everett, Candice Fazar, Tatiana Goldina, Zhaoyu Huai, Howard Hui, Woong-Seob Jeong, Jae Hwan Kang, Phillip M. Korngut, Jae Joon Lee, Daniel C. Masters, Chi H. Nguyen, Jeonghyun Pyo, Teresa Symons, Yujin Yang, Michael Zemcov, Rachel Akeson, Matthew L. N. Ashby, James J. Bock, Tzu-Ching Chang, Yun-Ting Cheng, Yi-Kuan Chang, Asantha Cooray, Olivier Doré, Andreas L. Faisst, Richard M. Feder, and Michael W. Werner. The SPHEREx Sky Simulator: Science Data Modeling for the First All-Sky Near-Infrared Spectral Survey. *arXiv e-prints*, art. arXiv:2505.24856, May 2025.
- Olivier Doré, Jamie Bock, Matthew Ashby, Peter Capak, Asantha Cooray, Roland de Putter, Tim Eifler, Nicolas Flagey, Yan Gong, Salman Habib, Katrin Heitmann, Chris Hirata, Woong-Seob Jeong, Raj Katti, Phil Korngut, Elisabeth Krause, Dae-Hee Lee, Daniel Masters, Phil Maukopf, Gary Melnick, Bertrand Mennesson, Hien Nguyen, Karin Öberg, Anthony Pullen, Alvise Raccanelli, Roger Smith, Yong-Seon Song, Volker Tolls, Steve Unwin, Tejaswi Venumadhav, Marco Viero, Mike Werner, and Mike Zemcov. Cosmology with the SPHEREX All-Sky Spectral Survey. *arXiv e-prints*, art. arXiv:1412.4872, December 2014. doi: 10.48550/arXiv.1412.4872.
- Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, Mar 2013. doi: 10.1086/670067.
- David W. Hogg, Jo Bovy, and Dustin Lang. Data analysis recipes: Fitting a model to data. *arXiv e-prints*, art. arXiv:1008.4686, Aug 2010.
- P. D. Hurley, S. Oliver, M. Betancourt, C. Clarke, W. I. Cowley, S. Duivenvoorden, D. Farrah, M. Griffin, C. Lacey, E. Le Floch, A. Papadopoulos, M. Sargent, J. M. Scudder, M. Vaccari, I. Valtchanov, and L. Wang. HELP: XID+, the probabilistic de-blender for Herschel SPIRE maps. *MNRAS*, 464(1):885–896, Jan 2017. doi: 10.1093/mnras/stw2375.

- Željko Ivezić, Vedrana Ivezić, Joachim Moeyens, Carey M. Lisse, Schelte J. Bus, Lynne Jones, Brendan P. Crill, Olivier Doré, and Joshua P. Emery. Simulated SPHEREx spectra of asteroids and their implications for asteroid size and reflectance estimation. *Icarus*, 371:114696, January 2022. doi: 10.1016/j.icarus.2021.114696.
- T. Kelsall, J. L. Weiland, B. A. Franz, W. T. Reach, R. G. Arendt, E. Dwek, H. T. Freudenreich, M. G. Hauser, S. H. Moseley, N. P. Odegard, R. F. Silverberg, and E. L. Wright. The COBE Diffuse Infrared Background Experiment Search for the Cosmic Infrared Background. II. Model of the Interplanetary Dust Cloud. *ApJ*, 508(1):44–73, Nov 1998. doi: 10.1086/306380.
- Leon B Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, Vol. 79, p. 745 (1974), 79:745, 1974.
- Gaelen Marsden, Peter AR Ade, James J Bock, Edward L Chapin, Mark J Devlin, Simon R Dicker, Matthew Griffin, Joshua O Gundersen, Mark Halpern, Peter C Hargrave, et al. Blast: resolving the cosmic submillimeter background. *The Astrophysical Journal*, 707(2):1729, 2009.
- E. Merlin, N. Bourne, M. Castellano, H. C. Ferguson, T. Wang, S. Derriere, J. S. Dunlop, D. Elbaz, and A. Fontana. T-PHOT version 2.0: Improved algorithms for background subtraction, local convolution, kernel registration, and new options. *A&A*, 595:A97, Nov 2016. doi: 10.1051/0004-6361/201628751.
- Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.
- William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the optical society of America*, 62(1):55–59, 1972.
- Massimo Robberto. Analysis of the sampling schemes for wfc3-ir. *WFC3 ISR*, 12, 2007.
- David L. Shupe, Russ R. Laher, Lisa Storrie-Lombardi, Jason Surace, Carl Grillmair, David Levitan, and Branimir Sesar. More flexibility in representing geometric distortion in astronomical images. In Nicole M. Radziwill and Gianluca Chiozzi, editors, *Software and Cyberinfrastructure for Astronomy II*, volume 8451 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 84511M, September 2012. doi: 10.1117/12.925460.
- Roger M Smith, Maximilian Zavodny, Gustavo Rahmer, and Marco Bonati. A theory for image persistence in hgcdte photodiodes. In *High energy, optical, and infrared detectors for astronomy III*, volume 7021, pages 180–191. SPIE, 2008.
- Teresa Symons, Michael Zemcov, James Bock, Yun-Ting Cheng, Brendan Crill, Christopher Hirata, and Stephanie Venuto. Superresolution Reconstruction of Severely Undersampled Point-spread Functions Using Point-source Stacking and Deconvolution. *ApJS*, 252(2):24, February 2021. doi: 10.3847/1538-4365/abcaa5.
- Kohji Tsumura, Toshio Matsumoto, Shuji Matsuura, Jeonghyun Pyo, Itsuki Sakon, and Takehiko Wada. Low-Resolution Spectrum of the Zodiacal Light with the AKARI InfraRed Camera. *PASJ*, 65:119, Dec 2013. doi: 10.1093/pasj/65.6.119.
- Antonella Vallenari, Anthony GA Brown, Timo Prusti, Jos HJ De Bruijne, F Arenou, Carine Babusiaux, Michael Biermann, Orlagh L Creevey, Christine Ducourant, Dafydd Wyn Evans, et al. Gaia data release 3-summary of the content and survey properties. *Astronomy & Astrophysics*, 674:A1, 2023.
- RA Viereck, Edmond Murad, B David Green, P Joshi, CP Pike, R Hieb, and G Harbaugh. Origin of the shuttle glow. *Nature*, 354(6348):48–50, 1991.
- MP Viero, L Moncelsi, RF Quadri, V Arumugam, RJ Assef, M Béthermin, J Bock, C Bridge, CM Casey, A Conley, et al. Hermes: the contribution to the cosmic infrared background from galaxies selected by mass and redshift. *The Astrophysical Journal*, 779(1):32, 2013.

Martin C. Weisskopf, Paolo Soffitta, Luca Baldini, Brian D. Ramsey, Stephen L. O'Dell, Roger W. Romani, Giorgio Matt, William D. Deinger, Wayne H. Baumgartner, Ronaldo Bellazzini, Enrico Costa, Jeffery J. Kolodziejczak, Luca Latronico, Herman L. Marshall, Fabio Muleri, Stephen D. Bongiorno, Allyn Tennant, Niccolo Bucciantini, Michal Dovciak, Frédéric Marin, Alan Marscher, Juri Poutanen, Pat Slane, Roberto Turolla, William Kalinowski, Alessandro Di Marco, Sergio Fabiani, Massimo Minuti, Fabio La Monaca, Michele Pinchera, John Rankin, Carmelo Sgrò, Alessio Trois, Fei Xie, Cheryl Alexander, D. Zachery Allen, Fabrizio Amici, Jason Andersen, Angelo Antonelli, Spencer Antoniak, Primo Attiná, Mattia Barbanera, Matteo Bachetti, Randy M. Baggett, Jeff Bladt, Alessandro Brez, Raffaella Bonino, Christopher Boree, Fabio Borotto, Shawn Breeding, Daniele Brienza, H. Kyle Bygott, Ciro Caporale, Claudia Cardelli, Rita Carpentiero, Simone Castellano, Marco Castronuovo, Luca Cavalli, Elisabetta Cavazzuti, Marco Ceccanti, Mauro Centrone, Saverio Citraro, Fabio D'Amico, Elisa D'Alba, Laura Di Gesu, Ettore Del Monte, Kurtis L. Dietz, Niccolò Di Lalla, Giuseppe Di Persio, David Dolan, Immacolata Donnarumma, Yuri Evangelista, Kevin Ferrant, Riccardo Ferrazzoli, MacKenzie Ferrie, Joseph Footdale, Brent Forsyth, Michelle Foster, Benjamin Garelick, Shuichi Gunji, Eli Gurnee, Michael Head, Grant Hibbard, Samantha Johnson, Erik Kelly, Kiranmayee Kilaru, Carlo Lefevre, Shelley Le Roy, Pasqualino Loffredo, Paolo Lorenzi, Leonardo Lucchesi, Tyler Maddox, Guido Magazzu, Simone Maldera, Alberto Manfreda, Elio Mangraviti, Marco Marengo, Alessandra Marrocchesi, Francesco Massaro, David Mauger, Jeffery McCracken, Michael McEachen, Rondal Mize, Paolo Mereu, Scott Mitchell, Ikuyuki Mitsuishi, Alfredo Morbidini, Federico Mosti, Hikmat Nasimi, Barbara Negri, Michela Negro, Toan Nguyen, Isaac Nitschke, Alessio Nuti, Mitch Onizuka, Chiara Oppedisano, Leonardo Orsini, Darren Osborne, Richard Pacheco, Alessandro Paggi, Will Painter, Steven D. Pavelitz, Christina Pentz, Raffaele Piazzolla, Matteo Perri, Melissa Pesce-Rollins, Colin Peterson, Maura Pilia, Alessandro Profeti, Simonetta Puccetti, Jaganathan Ranganathan, Ajay Ratheesh, Lee Reedy, Noah Root, Alda Rubini, Stephanie Ruswick, Javier Sanchez, Paolo Sarra, Francesco Santoli, Emanuele Scalise, Andrea Sciortino, Christopher Schroeder, Tim Seek, Kalie Sodian, Gloria Spandre, Chet O. Speegle, Toru Tamagawa, Marcello Tardiola, Antonino Tobia, Nicholas E. Thomas, Robert Valerie, Marco Vimercati, Amy L. Walden, Bruce Weddendorf, Jeffrey Wedmore, David Welch, Davide Zanetti, and Francesco Zanetti. Imaging X-ray Polarimetry Explorer: prelaunch. *Journal of Astronomical Telescopes, Instruments, and Systems*, 8(2):026002, 2022. doi: 10.1117/1.JATIS.8.2.026002. URL <https://doi.org/10.1117/1.JATIS.8.2.026002>.

Gregory R Zengilowski, Craig W McMurtry, Judith L Pipher, Nicholas S Reilly, William J Forrest, Mario S Cabrera, Meghan L Dorn, Andre F Wong, and AK Mainzer. Signal nonlinearity measurements and corrections in mwir and lwir hgcdte h2rg arrays for neo surveyor. In *X-Ray, Optical, and Infrared Detectors for Astronomy IX*, volume 11454, pages 616–633. SPIE, 2020.

A Spectral Image Header Example

Table A.20: Spectral Image FITS Header Keywords (Example)

Header	Value	Units	Description
XTENSION	'IMAGE '	–	Image extension
BITPIX	–32	–	Array data type
NAXIS	2	–	Number of array dimensions
NAXIS1	2040	–	–
NAXIS2	2040	–	–
PCOUNT	0	–	Number of parameters
GCOUNT	1	–	Number of groups
EXTNAME†	'IMAGE '	–	Extension name
EXTTYPE†	'IMAGE '	–	–
BUNIT	'MJy / sr'	–	–
DETECTOR	1	–	1-3: SWIR, 4-6: MWIR
OBSID	'2025W19_2B_0022_1'	–	Observation ID
EXPIDN	202519202002211	–	Integer Exposure ID
DATE	'2025-06-12T03:31:58.332'	–	HDU creation time (UTC)
TSAMP	1.5349	s	Interval between pixel readouts (frame)
KOVR_A	25000	adu	Overflow detection limit, active pixels
KOVR_R	25000	adu	Overflow detection limit, ref pixels
KOVR_P	25000	adu	Overflow detection limit, phantom pixels
OVRDLY	5	–	Threshold for overflow error flag
SURLIM	77	–	Number of frames detector accumulated charge
SURDLY	3	–	Frame count when SUR sums accumulation begins
TRANDLY	7	–	Frame count when transient detection begins
TCUT	38	–	Threshold for transient error flag
KEN_A	20000	–	Transient white noise threshold, active pixels
KEN_R	10000	–	Transient white noise threshold, ref pixels
KEN_P	500	–	Transient white noise threshold, phantom pixels
KSN_A	100	–	Transient shot noise threshold, active pixels
KSN_R	0	–	Transient shot noise threshold, ref pixels
KSN_P	0	–	Transient shot noise threshold, phantom pixels
NSKIP	32	–	Number of rows skipped after a full row is read
NVISITROW	4	–	Number of visits to each reference row
NVISITCOL	1	–	Number of visits to each reference col
MPAN	4	–	Number of phantom columns reads
NCHANNEL	32	–	Number of readout channels
WCHANNEL	64	–	Number of columns in each readout channel
JREF	8	–	Number of physical reference rows on H2RG
KREF	8	–	Number of physical reference cols on H2RG
JACTIVE	2040	–	Number of light-sensitive pixel rows
KACTIVE	2040	–	Number of light-sensitive pixel cols
SPS_TARGET	126055	–	Target ID assigned by Survey Planning

continued ...



... continued

Header	Value	Units	Description
SPS_FLAGS	'all_sky'	–	Flags defined by Survey Planning
OBS_IN_SBAND	'False '	–	Observed during S-band transmission
NON_SURVEY	'False '	–	Observation is not part of a science survey
SGT_LAT_MIDPT	–27.270774	deg	Ground track latitude at midpoint
SGT_LON_MIDPT	40.175792	deg	Ground track longitude at midpoint
SGT_MIN_GCD2SAA	20.158484	deg	Closest ground track dist to SAA
SPS_ELONG	164.000000	deg	Ecliptic longitude of planned telescope bore-sight
SPS_ELAT	–37.107076	deg	Ecliptic latitude of planned telescope bore-sight
SPS_EPA	180.000000	deg	PA of FP y axis at planned boresight, E of ecl N
L1DQAFLG	'Pass '	–	L1 DQA outcome
RADESYS	'ICRS '	–	Equatorial coordinate system
L2_N_NONLINEAR	0	–	Number of flagged nonlinear pixels
PSTHRESH	0.1	electron/s	Persistence threshold
L2_N_PERSIST	3026	–	Number of flagged persist pixels
DETCOORD	'sky '	–	MWIR pixels flipped on X to match SWIR view
L2_N_HOT	3	–	Number of flagged hot pixels
L2_N_COLD	1932	–	Number of flagged cold pixels
OMEGA_MEDIAN	37.86041790648254	arcsec**2	Solid angle pixel map median
OMEGA_STDDEV	0.20122470784388144	arcsec**2	Solid angle pixel map std
FINAST	0	–	Good fine-astrometric solution
WCSAXES	2	–	Number of coordinate axes
CRPIX1	1020.5	–	Pixel coordinate of reference point
CRPIX2	1020.5	–	Pixel coordinate of reference point
PC1_1	0.00152922340081	–	Coordinate transformation matrix element
PC1_2	–0.000822133914188	–	Coordinate transformation matrix element
PC2_1	–0.000765144685491	–	Coordinate transformation matrix element
PC2_2	–0.0014975756327	–	Coordinate transformation matrix element
CDEL1	1.0	deg	Coordinate increment at reference point
CDEL2	1.0	deg	Coordinate increment at reference point
CUNIT1	'deg'	–	Units of coordinate increment and value
CUNIT2	'deg'	–	Units of coordinate increment and value
CTYPE1	'RA---TAN-SIP'	–	TAN (gnomonic) projection + SIP distortions
CTYPE2	'DEC---TAN-SIP'	–	TAN (gnomonic) projection + SIP distortions
CRVAL1	146.217746981	deg	Coordinate value at reference point
CRVAL2	–26.057757842	deg	Coordinate value at reference point
LONPOLE	180.0	deg	Native longitude of celestial pole
LATPOLE	–26.057757842	deg	Native latitude of celestial pole
PXSRCMSK	42.91	–	Percentage of pixels with SOURCE flag
L2_N_SOURCE	1785594	–	Number of source flagged pixels
L2_N_OUTLIER	80786	–	Number of pixels flagged outlier
PSF_FWHM	5.029344163983229	arcsec	PSF FWHM Median
L2DQAFLG	'Pass '	–	L2 DQA outcome
L2_N_TRANSIENT	25803	–	Number of flagged transient pixels
L2_N_OVERFLOW	485	–	Number of flagged overflow pixels
L2_N_SUR_ERROR	17206	–	Number of flagged SUR error pixels

continued ...



... continued

Header	Value	Units	Description
L2 N_NONFUNC	1898	–	Number of flagged nonfunctional pixels
L2 N_DICHROIC	0	–	Number of flagged pixels due to low efficiency
L2 N_MISSING	0	–	Number of flagged pixels with missing data
L2 N_FULLSAMPLE	240	–	Number of flagged full-sample pixels
L2 N_PHANMISS	0	–	Number of flagged phanmiss pixels
A_ORDER	3	–	SIP polynomial order, axis 0, detector to sky
A_0_0	0.495747204533	–	SIP distortion coefficient
A_0_1	–1.41949696287E–05	–	SIP distortion coefficient
A_0_2	–1.40383661259E–06	–	SIP distortion coefficient
A_0_3	1.25399352333E–12	–	SIP distortion coefficient
A_1_0	0.000262045718522	–	SIP distortion coefficient
A_1_1	4.17604879086E–06	–	SIP distortion coefficient
A_1_2	6.46177906164E–10	–	SIP distortion coefficient
A_2_0	–1.60749850764E–06	–	SIP distortion coefficient
A_2_1	–1.22417054076E–10	–	SIP distortion coefficient
A_3_0	5.35256062728E–10	–	SIP distortion coefficient
B_ORDER	3	–	SIP polynomial order, axis 1, detector to sky
B_0_0	–1.22400910022	–	SIP distortion coefficient
B_0_1	0.000166554236478	–	SIP distortion coefficient
B_0_2	1.98774036244E–06	–	SIP distortion coefficient
B_0_3	4.43054642974E–10	–	SIP distortion coefficient
B_1_0	–0.00025387592197	–	SIP distortion coefficient
B_1_1	–6.80692276129E–07	–	SIP distortion coefficient
B_1_2	–5.91623937458E–11	–	SIP distortion coefficient
B_2_0	6.66644884707E–06	–	SIP distortion coefficient
B_2_1	6.24971161669E–10	–	SIP distortion coefficient
B_3_0	–5.14517694777E–11	–	SIP distortion coefficient
AP_ORDER	3	–	SIP polynomial order, axis 0, sky to detector
AP_0_0	–0.493651052992	–	SIP distortion coefficient
AP_0_1	1.98740552872E–05	–	SIP distortion coefficient
AP_0_2	1.39790401184E–06	–	SIP distortion coefficient
AP_0_3	–1.28417225913E–11	–	SIP distortion coefficient
AP_1_0	–0.000269625447233	–	SIP distortion coefficient
AP_1_1	–4.15417192622E–06	–	SIP distortion coefficient
AP_1_2	–6.11194622365E–10	–	SIP distortion coefficient
AP_2_0	1.59790347715E–06	–	SIP distortion coefficient
AP_2_1	7.97708902644E–11	–	SIP distortion coefficient
AP_3_0	–4.99904813857E–10	–	SIP distortion coefficient
BP_ORDER	3	–	SIP polynomial order, axis 1, sky to detector
BP_0_0	1.21866453149	–	SIP distortion coefficient
BP_0_1	–0.00017257194721	–	SIP distortion coefficient
BP_0_2	–1.97444600352E–06	–	SIP distortion coefficient
BP_0_3	–4.32385560254E–10	–	SIP distortion coefficient
BP_1_0	0.000261407790443	–	SIP distortion coefficient
BP_1_1	6.72548555973E–07	–	SIP distortion coefficient
BP_1_2	3.26470826544E–11	–	SIP distortion coefficient
BP_2_0	–6.64226909824E–06	–	SIP distortion coefficient
BP_2_1	–5.38798703265E–10	–	SIP distortion coefficient
BP_3_0	2.49348855333E–11	–	SIP distortion coefficient
WCSAXESA	2	–	Number of coordinate axes

continued ...



... continued

Header	Value	Units	Description
CRPIX1A	1.0	–	Pixel coordinate of reference point
CRPIX2A	1.0	–	Pixel coordinate of reference point
CDELTA1A	1.0	pixel	Coordinate increment at reference point
CDELTA2A	1.0	pixel	Coordinate increment at reference point
CUNIT1A	‘pixel’	–	Units of coordinate increment and value
CUNIT2A	‘pixel’	–	Units of coordinate increment and value
CTYPE1A	‘LINEAR’	–	Coordinate type code
CTYPE2A	‘LINEAR’	–	Coordinate type code
CRVAL1A	0.0	pixel	Coordinate value at reference point
CRVAL2A	0.0	pixel	Coordinate value at reference point
LATPOLEA	90.0	deg	Native latitude of celestial pole
WCSNAMEA	‘0-based active pixel coordinates’	–	Coordinate system title
TIMESYS	‘UTC’	–	Time scale
MJDREF	0.0	d	Modified Julian Date (MJD) of fiducial time
DATE-OBS	‘2025-05-08T15:00:24.955’	–	ISO-8601 time of observation
MJD-OBS	60803.625288831	d	MJD of observation
DATE-BEG	‘2025-05-08T15:00:24.955’	–	ISO-8601 time at start of observation
MJD-BEG	60803.625288831	d	MJD at start of observation
DATE-AVG	‘2025-05-08T15:01:22.514’	–	ISO-8601 time at midpoint of observation
MJD-AVG	60803.62595502	d	MJD at midpoint of observation
DATE-END	‘2025-05-08T15:02:20.073’	–	ISO-8601 time at end of observation
MJD-END	60803.626621209	d	MJD at end of observation
XPOSURE	113.5826	s	Exposure (integration) time
TELAPSE	118.1873	s	Elapsed time (start to stop)
WCSAXESW	2	–	Two output axes: wavelength and bandpass
WCSNAMEW	‘Approximated per-pixel wavelength and bandpass’	–	WCS label
CNAME1W	‘Wavelength’	–	First output axis is central wavelength
CNAME2W	‘Bandpass’	–	Second output axis is bandpass
CRPIX1W	1	–	Lookup table is in raw FITS coordinates
CRPIX2W	1	–	Lookup table is in raw FITS coordinates
CDELTA1W	1	pixel	Lookup table is in raw FITS coordinates
CDELTA2W	1	pixel	Lookup table is in raw FITS coordinates
CUNIT1W	‘um ’	–	Wavelength is reported in microns
CUNIT2W	‘um ’	–	Bandpass is reported in microns
CTYPE1W	‘WAVE-TAB’	–	Bilinear lookup table, FITS Paper 3, Sect 6.1
CTYPE2W	‘WAVE-TAB’	–	Bilinear lookup table, FITS Paper 3, Sect 6.1
CRVAL1W	1	pixel	Lookup table is in raw FITS coordinates
CRVAL2W	1	pixel	Lookup table is in raw FITS coordinates
PV1_3W	1	–	Index of wavelength in VALUES array
PV2_3W	2	–	Index of bandpass in VALUES array
PS1_0W	‘WCS-WAVE’	–	Extension with wavelength lookup table
PS2_0W	‘WCS-WAVE’	–	Extension with bandpass lookup table
PS1_1W	‘VALUES ’	–	Column with values for wavelength
PS2_1W	‘VALUES ’	–	Column with values for bandpass
PS1_2W	‘X ’	–	Column with control points for detector in X

continued ...



... continued

Header	Value	Units	Description
PS2_2W	'Y'	-	Column with control points for detector in Y

Notes: †: The image extension is identified by `EXTNAME = 'IMAGE'` and `EXTTYPE = 'IMAGE'`, while the corresponding flags extension uses `EXTNAME = 'FLAGS'` and `EXTTYPE = 'MASK'`. The designation `EXTTYPE = 'MASK'` indicates that the values within the extension represent bit masks, where each bit encodes a distinct flag condition.

B Absolute Gain Matrix Header Example

Table B.1: Absolute Gain Matrix Header Keywords (Example)

Header	Value	Units	Description
XTENSION	'IMAGE '	-	Image extension
BITPIX	-32	-	Array data type
NAXIS	2	-	Number of array dimensions
NAXIS1	2040	-	Length of axis 1
NAXIS2	2040	-	Length of axis 2
PCOUNT	0	-	Number of parameters
GCOUNT	1	-	Number of groups
BUNIT	'MJy s / (electron sr)'	-	Unit of pixel value
EXTNAME	'IMAGE '	-	Extension name

C Average PSF Header Example

Table C.1: Average PSF Image FITS Header Keywords (Example)

Header	Value	Units	Description
XTENSION	'IMAGE '	–	Image extension
BITPIX	–32	–	Array data type
NAXIS	3	–	Number of array dimensions
NAXIS1	101	–	Length of axis 1
NAXIS2	101	–	Length of axis 2
NAXIS3	121	–	Length of axis 3
PCOUNT	0	–	Number of parameters
GCOUNT	1	–	Number of groups
WCSAXES	2	–	Number of coordinated axes
CRPIX1	51.0	–	Pixel coordinate of reference point
CRPIX2	51.0	–	Pixel coordinate of reference point
CDELTA1	0.615	arcsec	Coordinate increment at reference point
CDELTA2	0.615	arcsec	Coordinate increment at reference point
CUNIT1	'arcsec'	–	Units of coordinate increment and value
CUNIT2	'arcsec'	–	Units of coordinate increment and value
CTYPE1	'XANG'	–	Coordinate type code
CTYPE2	'YANG'	–	Coordinate type code
CRVAL1	0.0	arcsec	Coordinate value at reference point
CRVAL2	0.0	arcsec	Coordinate value at reference point
CNAME1	'Relative offset in detector X from center of PSF, in arcseconds'	–	–
CNAME2	'Relative offset in detector Y from center of PSF, in arcseconds'	–	–
MJDREF	0.0	d	MJD of fiducial time
OVERSAMP	10	–	–
For integer $n \in [1, 11]$,			
XCTR_a	93.22727273	–	Center of x zone (0, $n-1$), where $a = n$
YCTR_b	93.22727273	–	Center of y zone ($n-1$, 0), where $b = 1 + 11*(n-1)$
XWID_a	186.45454546 [†]	–	Width of x zone (0, $n-1$)
YWID_a	186.45454546 [†]	–	Width of y zone (0, $n-1$)
XCTR_c	278.68181818	–	Center of x zone (1, $n-1$), where $c = a + 11$
YCTR_d	278.68181818	–	Center of y zone ($n-1$, 1), where $d = 2 + 11*(n-1)$
XWID_c	186.45454546 [†]	–	Width of x zone (1, $n-1$)
YWID_c	186.45454546 [†]	–	Width of y zone (1, $n-1$)
XCTR_e	464.13636364	–	Center of x zone (2, $n-1$), where $e = c + 11$
YCTR_f	464.13636364	–	Center of y zone ($n-1$, 2), where $f = 3 + 11*(n-1)$
XWID_e	186.45454546 [†]	–	Width of x zone (2, $n-1$)
YWID_e	186.45454546 [†]	–	Width of y zone (2, $n-1$)
XCTR_g	649.59090909	–	Center of x zone (3, $n-1$), where $g = e + 11$
YCTR_h	649.59090909	–	Center of y zone ($n-1$, 3), where $h = 4 + 11*(n-1)$
XWID_g	186.45454546 [†]	–	Width of x zone (3, $n-1$)
YWID_g	186.45454546 [†]	–	Width of y zone (3, $n-1$)
XCTR_i	835.04545455	–	Center of x zone (4, $n-1$), where $i = g + 11$
YCTR_j	835.04545455	–	Center of y zone ($n-1$, 4), where $j = 5 + 11*(n-1)$
XWID_i	186.45454546 [†]	–	Width of x zone (4, $n-1$)

continued ...



... continued

Header	Value	Units	Description
YWID_i	186.45454546 [†]	–	Width of y zone (4, n-1)
XCTR_k	1020.5	–	Center of x zone (5, n-1), where $k = i + 11$
YCTR_l	1020.5	–	Center of y zone (n-1, 5), where $l = 6 + 11*(n-1)$
XWID_k	186.45454546 [†]	–	Width of x zone (5, n-1)
YWID_k	186.45454546 [†]	–	Width of y zone (5, n-1)
XCTR_m	1205.95454545	–	Center of x zone (6, n-1), where $m = k + 11$
YCTR_o	1205.95454545	–	Center of y zone (n-1, 6), where $o = 7 + 11*(n-1)$
XWID_m	186.45454546 [†]	–	Width of x zone (6, n-1)
YWID_m	186.45454546 [†]	–	Width of y zone (6, n-1)
XCTR_p	1391.40909091	–	Center of x zone (7, n-1), where $p = m + 11$
YCTR_q	1391.40909091	–	Center of y zone (n-1, 7), where $q = 8 + 11*(n-1)$
XWID_p	186.45454546 [†]	–	Width of x zone (7, n-1)
YWID_p	186.45454546 [†]	–	Width of y zone (7, n-1)
XCTR_r	1576.86363636	–	Center of x zone (8, n-1), where $r = p + 11$
YCTR_s	1576.86363636	–	Center of y zone (n-1, 8), where $s = 9 + 11*(n-1)$
XWID_r	186.45454546 [†]	–	Width of x zone (8, n-1)
YWID_r	186.45454546 [†]	–	Width of y zone (8, n-1)
XCTR_t	1762.31818182	–	Center of x zone (9, n-1), where $t = r + 11$
YCTR_u	1762.31818182	–	Center of y zone (n-1, 9), where $u = 10 + 11*(n-1)$
XWID_t	186.45454546 [†]	–	Width of x zone (9, n-1)
YWID_t	186.45454546 [†]	–	Width of y zone (9, n-1)
XCTR_v	1947.77272727	–	Center of x zone (10, n-1), where $v = t + 11$
YCTR_w	1947.77272727	–	Center of y zone (n-1, 10), where $w = 11 + 11*(n-1)$
XWID_v	186.45454546 [†]	–	Width of x zone (10, n-1)
YWID_v	186.45454546 [†]	–	Width of y zone (10, n-1)
BUNIT	‘	–	Unit of pixel value
EXTNAME	‘PSF-DATA-CUBE’	–	Extension name

Notes: [†]: In practice, slight variations may occur due to rounding, but the value remains consistent to six decimal places (186.454545).



D List of Acronyms

ADCS Attitude Determination and Control Subsystem	5
ADU Analog Digital Units	20
ASIC Application-Specific Integrated Circuit	7
BAE BAE Space and Mission Systems	9
Caltech California Institute of Technology	5
CDH Command and Data Handling	5
CDS Correlated Double Sampling	22
DAPHNE Data Acquisition Processing & Handling Network Environment	20
DBS Dichroic Beam Splitter	14
DQA Data Quality Assurance	1
EPDS Electrical Power Distribution Subsystem	5
EBL Extragalactic Background Light	4
ESO European Southern Observatory	13
FITS Flexible Image Transport System	13
FoV Field of View	5
FPA Focal Plane Assemblies	5
FSW Flight Software	5
GDS Ground Data System	20
HDU Header and Data Unit	11
ICE Instrument Control Electronics	5
IOC In-Orbit Checkout	4
IRSA NASA/IPAC Infrared Science Archive	9
IXPE Imaging X-ray Polarimetry Explorer	4
JPL Jet Propulsion Laboratory	9
KASI Korea Astronomy and Space Science Institute	9
LEO low Earth orbit	4
LVF Linear Variable Filter	5
MEF Multi-Extension FITS	11
MIDEX Medium-class Explorer	4
MJD Modified Julian Date	52
MLTAN Mean Local Time of Ascending Node	5
MOS Mission Operations System	21
MWIR Mid-wave Infrared	7
NIR Near-infrared	4
NL Non-Linearity	21
NSN Near Space Network	9
PGSO Planning Generation and Simulation Operations	10
PA Position Angle	8
PI Principal Investigator	9
PSF Point Spread Function	2
QR Quick Release	11
RLD Richardson-Lucy Deconvolution	43
SAA South Atlantic Anomaly	32
SIP Simple Imaging Polynomial	28
SPS Survey Planning Software	7



SSDC SPHEREx Science Data Center	9
SSIRU Scalable Space Inertial Reference Unit	5
SST SPHEREx Science Team	9
S/N Signal-to-Noise	7
SUR Sample-Up-the-Ramp	6
SWIR Short-wave Infrared	7
TMA Three Mirror Anastigmat	5
TPV Tangent plane projection with parameter value (PV) polynomial distortion terms	29
2MASS Two Micron All Sky Survey	28
WCS World Coordinate System	13
WISE Wide-field Infrared Survey Explorer	28
ZL Zodiacal, or Zodi Light	16