Astronomical Data Analysis Software and Systems XII
ASP Conference Series, Vol. 295, 2003
H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds.

The Automated Data Processing Pipeline for SIRTF IRS

Fan Fang, Jing Li, Bob Narron, Clare Waterson, Iffat Khan, Wen P. Lee, John Fowler, Russ Laher, & Mehrdad Moshir

SIRTF Science Center, Caltech, Pasadena, CA 91125

Abstract. We present the design, structure, and implementation of the automated data processing pipelines for the Infrared Spectrograph onboard Space Infrared Telescope Facility. This includes science data reduction pipelines that generate Basic Calibrated Data and enhanced science products, and calibration pipelines generating calibration data that allows reduction of the science data.

1. Introduction

The Infrared Spectrograph (IRS) will be one of the three instruments onboard the NASA mission Space Infrared Telescope Facility (SIRTF). Four instrument modules of IRS are built to observe the mid-infrared (5 to 40 microns) spectra of astronomical sources in four overlapping wavelength channels with low- and medium-resolution dispersion optics and As:Si and As:Sb BIB detectors. The IRS data processing pipelines have been developed at the SIRTF Science Center (SSC) to reduce IRS data. The IRS Science and Coadd pipelines remove a combination of detector electronic and optical artifacts and generate high signal-to-noise ratio (S/N) 2-dimensional science images from raw data. Calibration pipelines reduce data taken for specific calibration purposes and generate calibration images to enable science data processing. The Pointing Transfer pipeline interacts with a pointing server to provide pointing data for each Data Collection Event (DCE). These pipelines work together to provide the Basic Calibrated Data (BCD). The Post-BCD pipeline enhances data products by extracting 1-dimensional spectra from 2-dimensional BCDs.

2. IRS Pipeline Architecture

The IRS data processing involves several sub-systems interacting with each other. Figure 1 illustrates the main components involved in the process. When a DCE is received from the Flight Operations System (FOS), it is ingested and records are made in the Science Operations Database (SODB) and data archive. The PrepareDCE subsystem retrieves the records and establishes processing directories. The appropriate processing pipeline is activated based on information in SODB. Each IRS pipeline communicates with the SODB, retrieving the calibration and controlled data files appropriate for the DCE during processing. Upon completion of reducing the data, the pipeline loads product and Quality Assurance (QA) data into the SODB. Data products are physically stored in local disk and an area called Sandbox, before being moved to the archive.

Pipelines have established priorities and are executed in order. The calibration pipelines are to be executed first on calibration-specific data since they are tasked to provide calibration files for science data reduction. Among the calibration pipelines the reference dark current calibration and linearity model are needed before calculating the flatfield or efficiency frame for spectra, and therefore have higher priorities. Before a high-quality calibration product is generated, each calibration DCE is reduced by a calibration pre-processing pipeline. When a given number of DCEs are pre-processed an ensemble processing pipeline is triggered and multiple pre-processing products are combined and reduced together to yield a calibration product. Similarly the science pipeline is to be executed first on individual DCEs before multiple products are coadded to yield a 2-dimensional BCD, which Post-BCD pipeline picks up and extracts the 1-dimensional spectra. The Automated Processing Executive for SIRTF (APES) controls the orderly pipeline execution.

3. The IRS Processing Pipelines

The IRS processing pipelines consist of stand-alone modules, each is communicated via wrapper scripts written in Perl. Each module performs a specific task, such as removing a detector electronic artifact, as well as calculating and propagating the uncertainty and updating pixel status mask files. The electronic and optical artifacts that the pipeline modules handle include baseline and dark current removal, analog-to-digital de-saturation, droop and row-droop effects removal, non-linearity correction, radhit detection, jail-bar pattern removal, stray-light or order cross-talk removal, and efficiency frame removal or flatfielding. Since the IRS science data is taken in a sample-up-the-ramp mode which creates a data cube, a slope image is calculated to reflect the total integration. Multiple such slope images of a given sky location and free of instrument artifacts are coadded to produce a high S/N BCD.

Further reduction of the BCD follows the curvature of the spectrum in both pixel and wavelength space in the 2-dimensional BCD image. The spectrum is Nyquist-sampled based on resolution of the instrument. The profile in the cross-dispersion direction is examined and an average profile generated for each spectral order. The spectrum extraction is done for each order based on the average profile and a wavelength-dependent extraction window. The extracted spectra of different orders are then stitched together using a set of tuning parameters, which are calibrated from reducing the IRS spectra of known astronomical sources. A 1-dimensional Post-BCD product is generated and archived.

Figure 2 shows a few examples of IRS BCD and Post-BCD pipelines. The nature of spectroscopy poses significant challenges on data reduction. A number of IRS calibrations, such as ensemble flatfield, wavelength, fringes, etc., are not suited for automated processing and have to be done offline. When automated pipelines finish the task, IRS scientists step in and close the loop of completing the calibration. Interactive tools are being developed to facilitate this task.

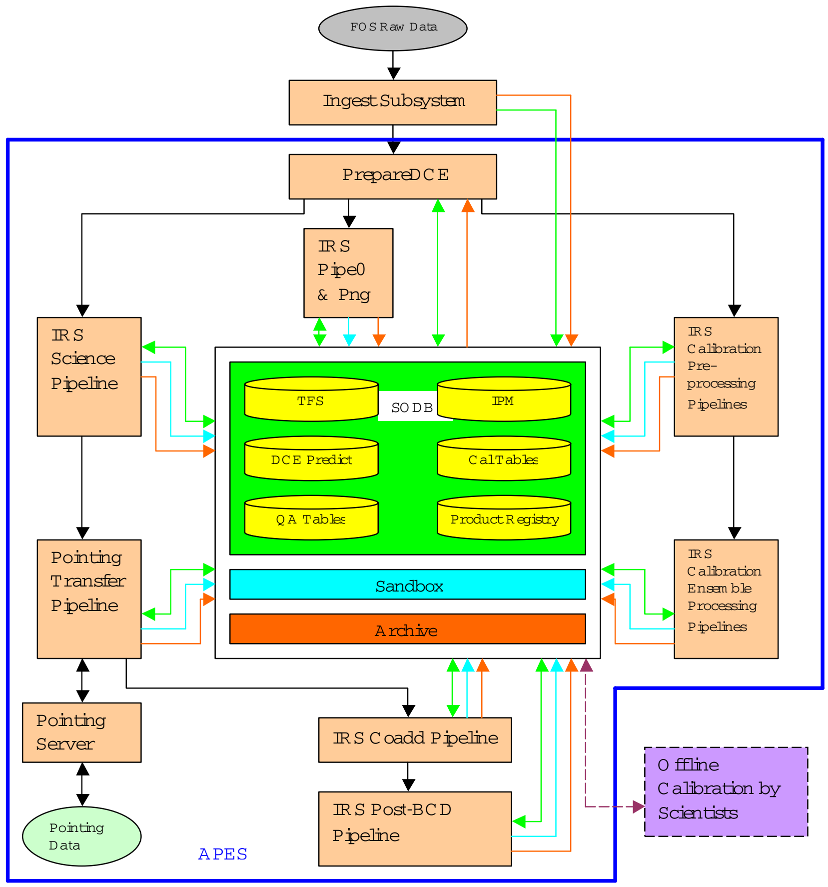

Figure 1. The architecture of the automated IRS data processing system. Several subsystems are involved in reducing an IRS DCE. The SODB plays a central role, from which each of the subsystem retrieves and loads information. The IRS pipelines use TFS to obtain version-controlled input configuration files, and call a Caltrans process to retrieve the best calibration files based on calibration table records. The QA subsystem, as a pipeline stage, loads selected ancillary data into the SODB. The Instrument Performance Monitoring (IPM) subsystem loads house-keeping data into the SODB. Both are used for monitoring purposes. The IRS pipeline0 & Png does minimal processing for DCEs, mostly engineering data. The automated pipeline processing is controlled and managed by APES.

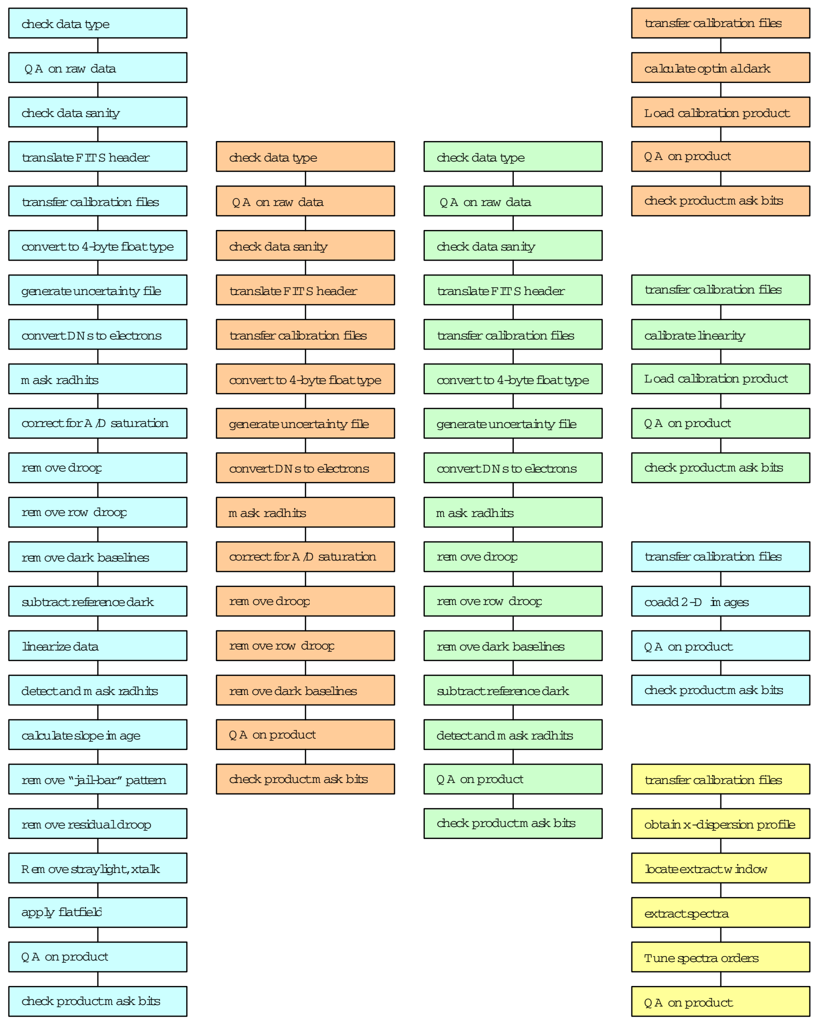

Figure 2. Examples of IRS BCD and Post-BCD processing pipelines. Shown here are IRS Science and Coadd (light blue), IRS Darkcal Pre- and Ensemble processing (light brown), IRS Lincal Pre- and Ensemble processing (light green), and IRS Post-BCD Science pipeline (yellow). Each pipeline processes from top down. IRS pipelines not shown here include Flatcal Pre-processing, Photocal Pre-processing, Pointing and Pipe0 & Png pipelines.