ADASS XIV
ASP Conference Series, Vol. XXX, 2004
P. Shopbell, M. Britton and R. Ebert eds.

Caltrans Keeps the Spitzer Pipelines Moving

Wen Lee, Russ Laher, John Fowler, and Mehrdad Moshir

Spitzer Science Center, MS 314-6, California Institute of Technology, Pasadena, CA 91125[1]

Abstract. The computer pipelines used to process digital infrared astronomical images from NASA's Spitzer Space Telescope require various input calibration-data files for characterizing the attributes and behaviors of the onboard focal-plane-arrays and their detector pixels, such as operability, dark-current offset, linearity, non-uniformity, muxbleed, droop, and point-response functions. The telescope has three very different science instruments, each with three or four spectral-band-pass channels, depending on the instrument. Moreover, each instrument has various operating modes (e.g., full array or sub-array in one case) and parameters (e.g., integration time). Calibration data that depend on these considerations are needed by pipelines for generating both science products (production pipelines) and higher-level calibration products (calibration pipelines). The calibration files are created in various formats either "off-line" or by the aforementioned calibration pipelines, depending on the above configuration details. Also, the calibration files are generally applicable to a certain time period and therefore must be selected accordingly for a given raw input image to be correctly processed. All of this complexity in selecting and retrieving calibration files for pipeline processing is handled by a procedural software-program called "caltrans". This software, which is implemented in C and interacts with an Informix database, was developed at the Spitzer Science Center (SSC) and is now deployed in SSC daily operations. The software is rule-based, very flexible, and, for efficiency, capable of retrieving multiple calibration files with a single software-execution command.

1. Introduction

The provision of the correct calibrations for reducing the raw data measured by a scientific instrument can be a challenge because the properties of the instrument generally change over time and, therefore, so does the calibration data. Pipeline data-processing of digital infrared astronomical images from NASA's Spitzer Space Telescope is even more complex because there are three uniquely different, multi-channel science instruments onboard (IRAC, MIPS, and IRS), and there

[1]Author e-mail addresses, respectively: wplee@ipac.caltech.edu, laher@ipac.caltech.edu, jwf@ipac.caltech.edu, and mmm@ipac.caltech.edu.

are correspondingly many different kinds of calibrations required (e.g., Masci *et al.*, 2005). The Spitzer pipeline-processing is done in parallel on about 50 workstations, and multiple parallel processes running on each workstation can access common calibration files stored on a local hard disk. For this system, fast software for precise calibration-file retrieval, utilizing data management in a database, is essential for determining the best set of calibration files for a given process and then copying them to the target workstation.

This paper describes the calibration-file-retrieval software and associated database schema that we have developed and implemented for the Spitzer mission. The software is called caltrans, which is short for calibration transfer. It has been successfully used in operations at the Spitzer Science Center (SSC) since the telescope's launch in August 2003.

2. High-Level Requirements

A primary requirement levied on caltrans, perhaps the most important of all, is that the complications involved in retrieving calibration data should be decoupled from the design and development of the other pipeline software-programs. These programs were difficult enough to develop without the added complexity of every one of them having to have similar logic about which calibration files to use, where to get them, etc.

Other important requirements of caltrans include fast performance and minimal impact on the project's Informix database. Also, to ensure that a complete set of calibration files will always be retrieved by caltrans, a requirement was made that "fallback" calibration files will always be available as substitutes if caltrans database queries fail to find the requested calibration files.

3. Database Schema

Caltrans queries several database tables for calibration files (see Figure 1). The caltransControls table associates a unique database index (calId) with a given Spitzer instrument, channel number (each instrument has three or four channels, depending on the instrument), and the specific database-table names where meta-data about the corresponding calibration files are stored. Calibration files are classified as either "fallback" or "non-fallback" in the database.

In order to keep the number of caltrans database-tables manageable, there is a channel-number field (chnlnum) in the database tables for IRAC and IRS, which is permissible because of the parallel nature of their respective data channels. On the other hand, because the three MIPS channels are so different is was necessary to design separate database tables for them.

Meta-data about the *fallback* calibration files are stored in five database tables: iracFallback, irsFallback, mips1Fallback, mips2Fallback, and mips3Fallback. Each of these tables has a primary key (fbId) for fast access to the meta-data. Records in these database tables are populated manually via a script calling database stored-procedures.

Meta-data about the more generally used *non-fallback* calibration files, which are generated by automated calibration pipelines in Spitzer operations, are

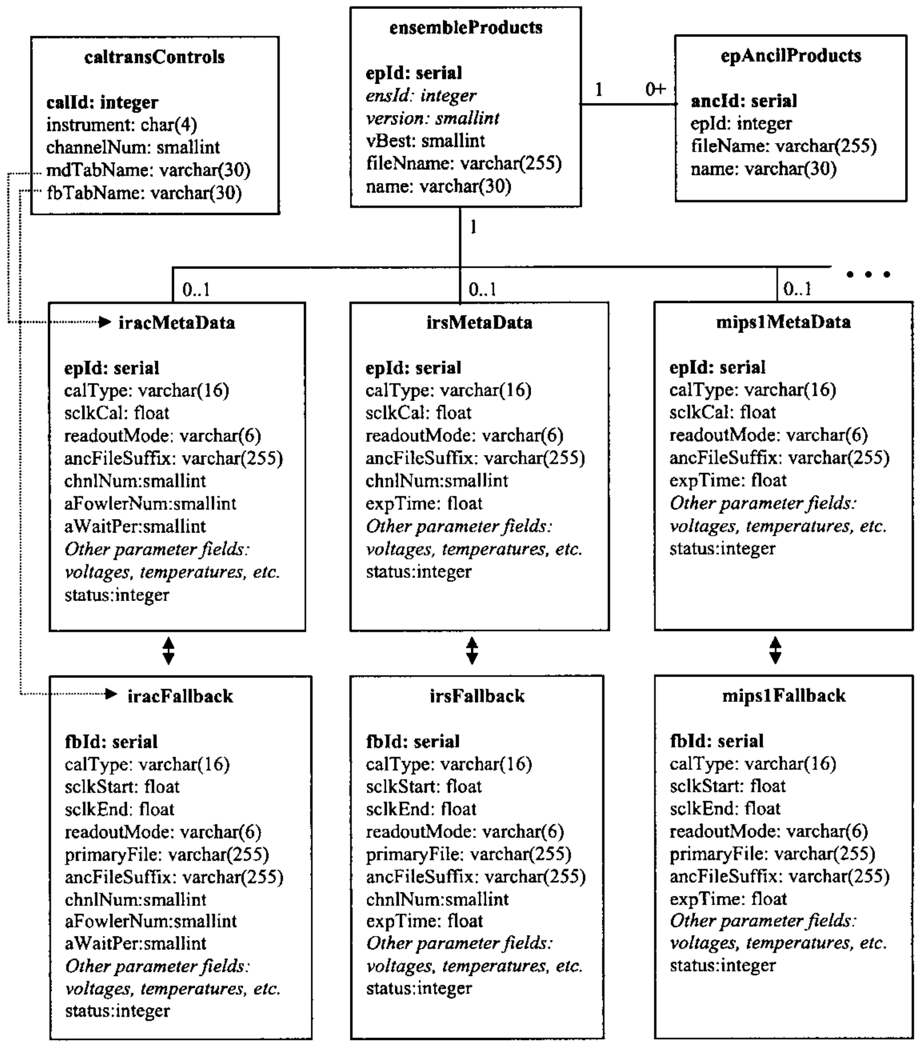

Figure 1. Schema for the database tables used by caltrans. The database tables not shown are mips2MetaData, mips3MetaData, mips2Fallback, and mips3Fallback.

also stored in five database tables, separately from the fallbacks: iracMetaData, irsMetaData, mips1MetaData, mips2MetaData, and mips3MetaData. Since these calibration files are registered by calibration pipelines in the ensembleProducts database-table, where each file is assigned a unique database ID called epId, the calibration non-fallback database-tables use epId as a primary key, which is distinctly different from the calibration fallback database-tables.

Records in the calibration fallback and non-fallback database-tables contain a number of useful fields for distinguishing a given calibration file from another. The caltype field allows various broad categories of calibration types to be defined, such as "labdark" or "skyflat". Each record is applicable to a primary calibration file (e.g., a dark-offset correction image) and zero or more ancillary calibration files (e.g., associated mask image, uncertainty image, etc.).

For the calibration fallback database-tables, the complete path/filename of the primary calibration file is stored in the primaryfile field and ancillary-file suffixes are stored in the ancFileSuffix field as a comma-separated list. It is assumed for fallbacks that the primary and ancillary files are stored together in the same directory and the ancillary filenames are basically derivable by inserting the ancillary-filename suffixes in the primary filename before the extension.

For the calibration non-fallback database-tables, the complete path/filename of the primary file is stored in the filename field of the ensembleProducts database-table, and so a table join is necessary in the caltrans database query. The ancillary filenames are derived as described above, and for calibration files that have been archived (and renamed to have archival filenames), the archival locations are looked up in the corresponding record in the epAncilProducts database-table (not shown).

A given calibration file is generally applicable to a certain range or point in time. The sclkStart and sclkEnd (sclk stands for Spitzer spacecraft time) fields in the calibration fallback database-tables allow a time range to be specified for the calibration file. The sclkCal field in the calibration non-fallback database tables is for storing the average time associated with the calibration file as computed by a calibration pipeline.

4. Software

Caltrans is a stand-alone, procedural software-program that was implemented in C for the fastest performance possible. Its architecture includes embedded SQL functionality for querying the calibration database-tables. The software runs under the Solaris operating system in Spitzer operations; it has also been shown to compile/build successfully under Linux.

Caltrans is normally executed as an early pipeline step. A single caltrans command is capable of retrieving all the calibration files that the pipeline may require. After a database connection has been established, multiple database queries are performed in order to find all the requested files, and then the database connection is closed. Making maximal use of the open database connection fulfills our requirement for minimally impacting the database.

Caltrans has available several rules for retrieving calibration files (Table 1 give the basic ones). A different rule can be applied for each individual calibration file in the set of files to be retrieved. Note that a variant of rule 400, which employs a random-walk method of interpolating uncertainties between sample times (see Moshir *et al.*, 2003), is currently under development.

Table 1. Basic calibration-file retrieval rules implemented in caltrans.

Rule	Description
0	Get fallback-only calibration file
100	Get nearest-in-time-before calibration file
200	Get nearest-in-time-after calibration file
300	Get nearest-in-time (before or after) calibration file
400	Create calibration file via linear interpolation in time (tabular data only)

Input configuration/control parameters are read from either a specially-formatted input file ("namelist"), the command line, or a combination of both. Command-line parameters always override any identical ones that may be in the namelist. There are options for tailoring the database query; e.g., chnlNum=0, ASLDTmpA=6.21±0.01 (IRS Short-Low-Detector Temperature A), etc.

The reference time of the retrieved calibration files must be input to caltrans, in units of Spitzer spacecraft clock, for querying the calibration database-tables. The reference time can either be read from the caltrans namelist, command-line, or header of a specified FITS-image (SCLK_OBS keyword).

Only those calibration files with the status field in the fallback and non-fallback database-tables set to values ranging from 1024 to 8191 are retrievable by caltrans; values outside of this range cause the calibration file to be effectively removed from the system. Additionally, only non-fallback calibration files with the vBest field in the ensembleProducts database-table set to a value greater than 0 are retrievable by caltrans.

A local-disk "cache" directory may be specified as caltrans input. Caltrans will check the cache to see if the calibration file already exists there from a prior execution, before retrieving the file from the networked file system. This reduces network traffic, which is important for multi-workstation parallel-computing.

Acknowledgments. This work was performed at the Spitzer Science Center as part of a mission/project managed by Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Masci, F., R. Laher, F. Fang, D. Makovoz, W. Lee, S. Stolovy, D. Padgett, & M. Moshir 2005, this volume

Mehrdad Moshir, John Fowler, & David Henderson 2003, in APS Conf. Ser., Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 181